

Hard Evidence on Soft Skills: The GED and the Problem of Soft Skills in America (Forthcoming, University of Chicago Press, 2011)

James J. Heckman

University of Chicago

Cowles Foundation, Yale University

American Bar Foundation

John Eric Humphries

University of Chicago

Nicholas Mader

University of Chicago

**What Have We Learned? Where Are We Headed?
Testing, Determinants of School Choice, Experiments**

Santiago, Chile

November 22, 2010

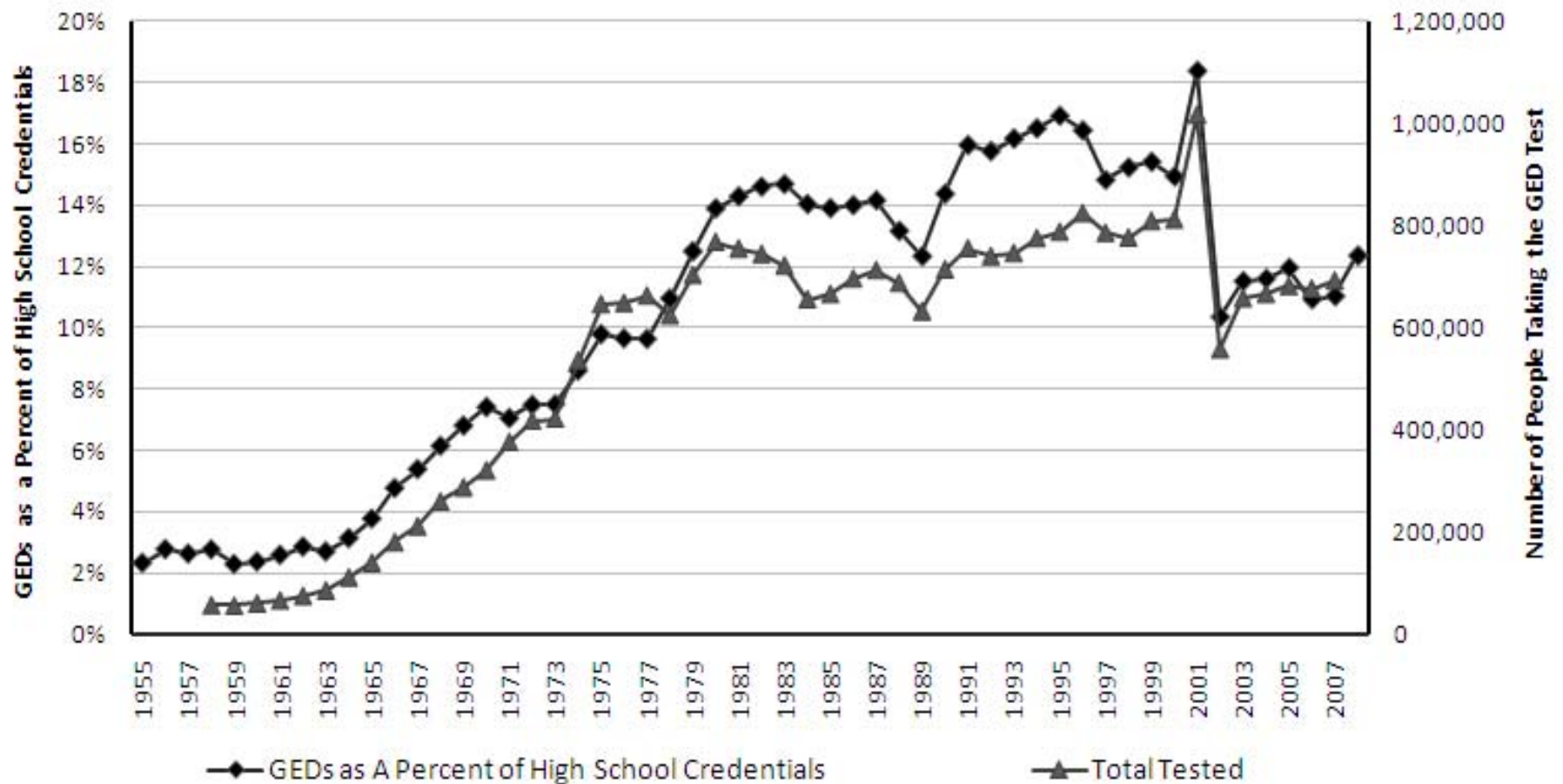
- This talk is about the achievement tests that are widely used to screen, sort and monitor the success of schools and society.
- PISA tests; No Child Left Behind; Iowa Tests; NAEP tests; and the GED (General Educational Development); LSAT; SAT.
- This lecture is about what these tests miss and why what they miss is important to know.
- They miss noncognitive (personality) traits.

- These are sometimes called “soft skills” or “character” traits.
- Perseverance, conscientiousness, motivation, willful planning.
- They are predictive of a range of important behaviors.
- Ignoring personality and character is a dangerous thing to do and can lead to costly mistakes in assessing and addressing social problems.

- I show this with the GED, an achievement test that high school dropouts can take to certify that they are the equivalents (in cognition) of high school graduates.
- It is a quantitatively important program in the U.S. and is also used in Canada.
- A version of it was adopted in Brazil to solve its problem of having a poorly educated population.

- The GED testing program is active in proselytizing its product.

Figure 1: Growth in the GED - Percent of High School Credentials and Number of Takers



- Lessons from a study of the GED exam program and the use of related exams are broader.
- They apply to many countries.

What Are These Lessons?

- More than academic achievement is required for success in life.
- Personality—“character”—can be measured.
- Personality can be fostered.
- Interventions that promote “character” are unexplored and potentially powerful avenues for social policy.

What Are These Lessons?

- There are other lessons as well.
- Movements for “accountability” in education often create perverse incentives.
- Tests and test certification can create and conceal problems.
- Uncritical reliance on tests as measures of the outcomes of schools and social programs is a dangerous policy.

The Power of Personality

- The wide array of outcomes causally affected by soft skills is remarkable: smoking, health, teen pregnancy, high school graduation, wages, criminal activity, employment, and welfare dependency to name only a few.
- In many cases, soft skills play a greater role in determining outcomes than do cognitive skills.

Character Can Be Measured

- If soft skills are so important, why have they been ignored in public policy discussions for so long?
- Many people view these skills as “fuzzy” concepts that have only tenuous effects on behavior and that may be more the consequence than the cause of behavior.
- Another reason is that, unlike cognitive skills, soft skills are thought to be difficult to measure such as an IQ or SAT score.
- That is wrong, as I show in this lecture.

Introduction:

Hard Evidence on Soft Skills

- It is a truism that many different skills are important for success in life.
- Achievement in different fields requires different bundles of talent at different levels.

Introduction:

Hard Evidence on Soft Skills

- Thomas Edison: an exceedingly clever inventor.
- He was also hard driving and persistent.
- He is celebrated both for his deep insights and for his willingness to perform endless experiments before he discovered the idea that became his next invention.
- His self-description: “genius 1% inspiration, 99% perspiration.”
- Woody Allen: “80% of success is showing up.”

Introduction:

Hard Evidence on Soft Skills

- In almost every task in life, more than just the raw ability to solve abstract problems is required for success, although the proportions of cognition and character required vary across tasks.
- Self-control, self-discipline, and motivation are required to foster and apply talent.

Introduction:

Hard Evidence on Soft Skills

- Despite the powerful intuitive force of the idea that many different abilities are required to succeed in most tasks in life, soft skills are ignored in most popular and academic discussions of skill and skill formation of the success and failure of schools.
- The emphasis in most public policy discussions is almost exclusively on cognitive skills—intelligence and acquired knowledge as measured by IQ and achievement tests. Everything else is given a back seat.

Introduction:

Hard Evidence on Soft Skills

- Schools are now expected to teach cognition and not character.
- School effectiveness is more often measured by test scores.
- Character education is thought to be the province of the family.
- Yet the family is under attack in many countries around the world.

Introduction:

Hard Evidence on Soft Skills

- More single-parent families that provide fewer resources for their children.
- Social policy has to look beyond the one-dimensional focus on cognition and tests of cognition that document current thinking.

I. Origins of Testing

- The modern accountability movement in education arose in U.S. educational reform movements.
- Horace Mann (1840s) introduced the standardized test.
- The instrument was crude.
- But Mann saw its limitations, even if the instrument were perfected.
- Mann viewed a primary function of schools as teaching morality and character.

Mann:

“Hence to value schools, by length instead of quality, is a matchless absurdity. Arithmetic, grammar, and the other rudiments, as they are called, comprise but a small part of the teachings in a school. The rudiments of feeling are taught not less than the rudiments of thinking. The sentiments and passions get more lessons than the intellect. Though their open recitations may be less, their secret rehearsals are more.”

-Horace Mann (1867, p. 420)

- Mann's ideas for standardized testing were not pursued on a widespread scale for another 50—70 years (exception, New York Regents in 1870s).
- 19th Century used input-based measures (e.g., standardized curriculum) to evaluate schools.
- There was often harsh discipline.
- Also schooling targeted elites: college and even high school only for the select.
- Progressives (John Dewey and others) sought to free up the curriculum (early 20th Century).
- Enroll a wider swath of society into schools.

IQ as a Filter

- The Progressives needed a device to filter and track students.
- IQ tests—an early 20th Century invention that played this role.
- First IQ test was designed to screen out misfits in school.
- Role of test broadened to sort students within schools—this produced tracking.
- But the creators of IQ tests realized their limitations.

Binet:

“[Success in school] . . . admits of other things than intelligence; to succeed in his studies, one must have qualities which depend on attention, will, and character; for example a certain docility, a regularity of habits, and especially continuity of effort. A child, even if intelligent, will learn little in class if he never listens, if he spends his time in playing tricks, in giggling, in playing truant.”

-Binet (1916, p. 254)

- At about the same time that Binet was writing, Charles Spearman, best known for his work on “g” –a unitary factor that is claimed to capture the structure of intelligence-along with his student, Edward Webb, undertook studies of “character” because of “the urgency of its practical application to all the business of life” (Webb 1915, p. 1).
- Spearman and Webb concluded that many positive aspects of character shared a relation to what modern personality psychologists term “Conscientiousness.”

Arthur Jensen, the intellectual heir of Spearman and ardent proponent of the power of g, a measure of intelligence or problem-solving ability, in explaining success in life, writes:

“What are the chief personality traits which, interacting with g, relate to individual differences in achievement and vocational success? The most universal personality trait is conscientiousness, that is, being responsible, dependable, caring, organized and persistent.”

-Jensen (1998, p. 575)

Taylorism: “Scientific Management”

- A late 19th and early 20th century obsession.
- Formalized the American passion for efficiency and productivity.
- Taylorism was highly influential—scientific management, measurement, and accountability.
- But how to apply it to schools?

Bobbitt, a Taylorist University of Chicago educator in the early 20th Century:

“Education is a shaping process as much as the manufacture of steel rails; the personality is to be shaped and fashioned into desirable forms. It is a shaping of more delicate matters, more immaterial things, certainly; yet a shaping process none the less.”

-Bobbitt (1913, pp. 12-13)

- Bobbitt lacked the tools to measure the output of schools, but like Mann he viewed personality as an output.
- IQ was (and is) held to measure a fixed trait.
- The achievement test was created in the wake of the IQ test---a way to implement Taylorism in the schools.
- To capture the knowledge acquired in schools

General Knowledge

- Achievement tests were created to measure “general knowledge.”
- Ralph Tyler (Chicago) and Edward Lindquist (Iowa) invented the achievement test as a way to measure “general knowledge” and developed the technology to implement the achievement test.
- Designed to capture important life skills; not specific knowledge of a course.
- Iowa tests; ACT; GED; No Child Left Behind; NAEP.
- All of the originators of these tests understood what they missed.

Tyler favored other mechanisms of assessment to capture these outputs of schools:

“We lean heavily on written examinations, on a few types of objective tests, and on the subjective impressions of teachers. Many other appraisal devices could be used, such as records of activities in which pupils participate, questionnaires, check lists, anecdotal records and observational records, interviews, reports made by parents, products made by the pupils, and records made by instruments (motion pictures, eye-movement records, sound recordings, and the like).”

-Tyler (1940, p. 27)

The GED

- GED—an achievement test created to certify “the general knowledge” of what soldiers had learned in WWII.
- Soldiers had been in US Army in World War II—2-3—years formed discipline and selected on traits that were required to serve in military: obedience, self-control, etc.
- Thus it was implicit at the time the GED was created that certifiers possessed “character.”
- They also acquired knowledge through course work (Armed Forces Institute) and through life experiences.
- GED test was later (1950s) applied to civilian populations as a way to solve the dropout problem.

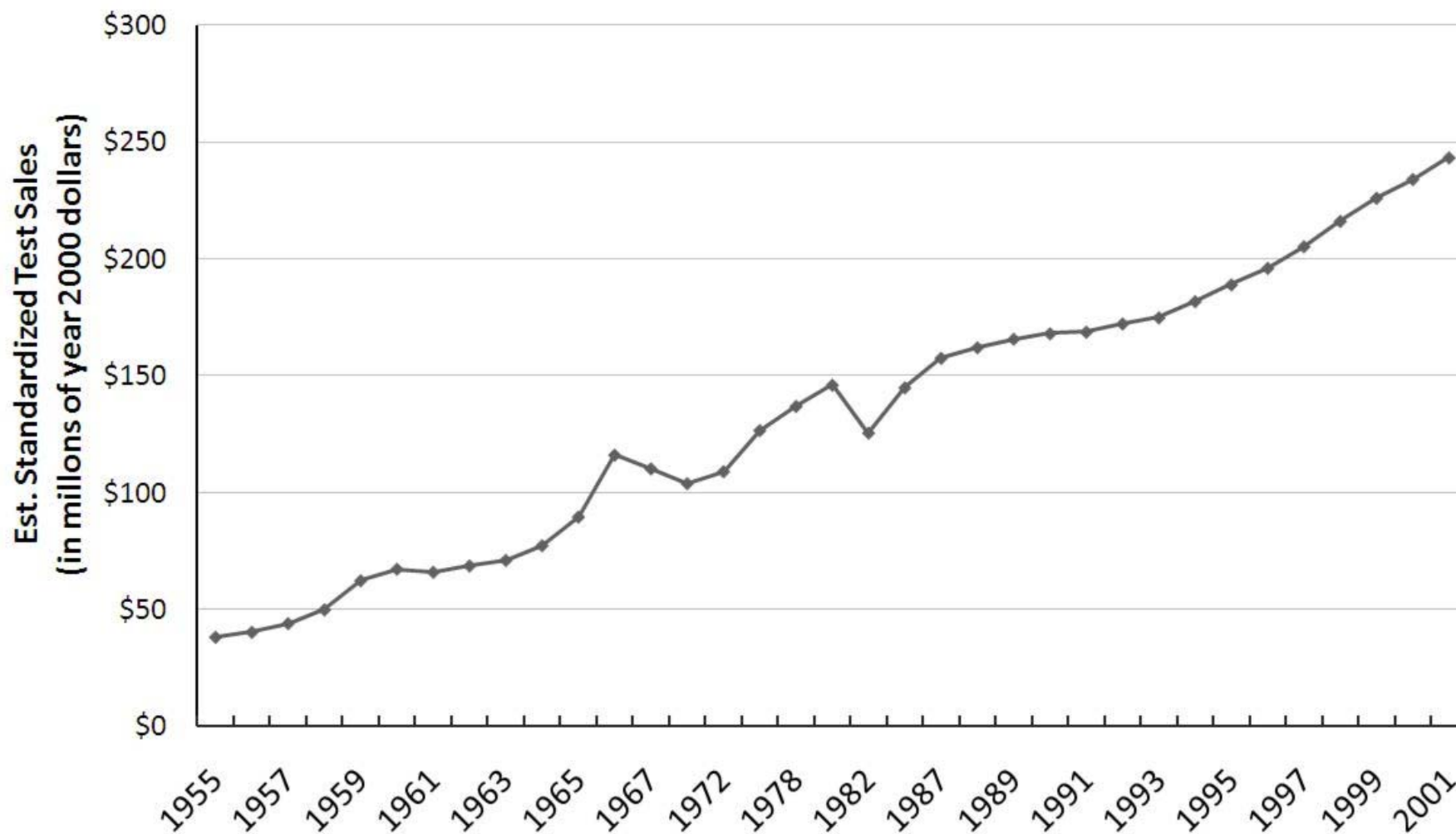
The GED

- Forces pushing toward widespread acceptance of tests.
 - Desire for egalitarianism—meritocracy .
 - SAT and Conant: target bright kids; break good old boy network
 - Creation of a meritocracy.

The GED

- Accountability movement in U.S. policy in the Kennedy-Johnson administration.
 - a. McNamara and the “Whiz Kids” revived Taylorism: Apply economic principles to social programs and produce a social version of a profit-loss statement.
 - Extreme case: body counts in Vietnam.
 - b. Great Society and Evaluation; Focused on Educational Policy; Accountability Returned (Henry Aaron: *Politics and the Professors: The Great Society in Perspective*)
 - c. Achievement and IQ tests scores in social programs were favored as good measures of success.
 - d. GED was a measure of output—producing high school graduates
 - e. Nation at Risk: American Education Failing: a renewed emphasis on testing
 - f. Accountability Goes on Steroids: ultimate expression is the No Child Left Behind movement in the 2000s.

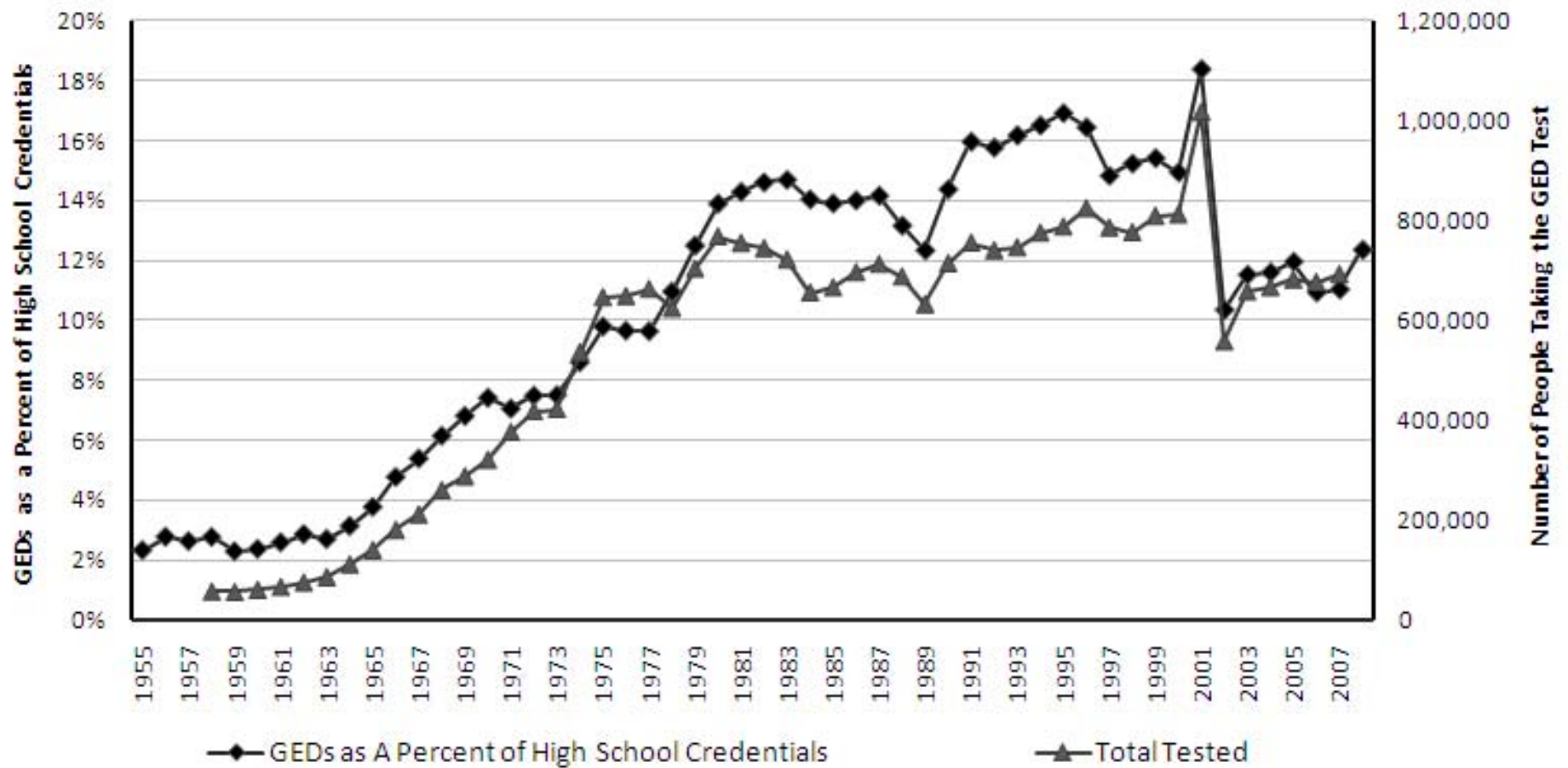
Figure 2: Sales of Standardized Tests



Sources: Collins and Schick (1970), Simora and Schick (1981), Simora and Harris (1991), Barr and Simora (1992), Bogart (1996, 1999, 2002).

- GED in particular took off

Figure 3: Growth in the GED - Percent of High School Credentials and Number of Takers



II. Using the GED as a Prism on the Skills that Matter and on American Society

- What is the GED?
- Look at GEDs (at the level of 9th grade to 10th grade).

Questions from GED Test

Figure 4: GED Sample Questions

Mathematics

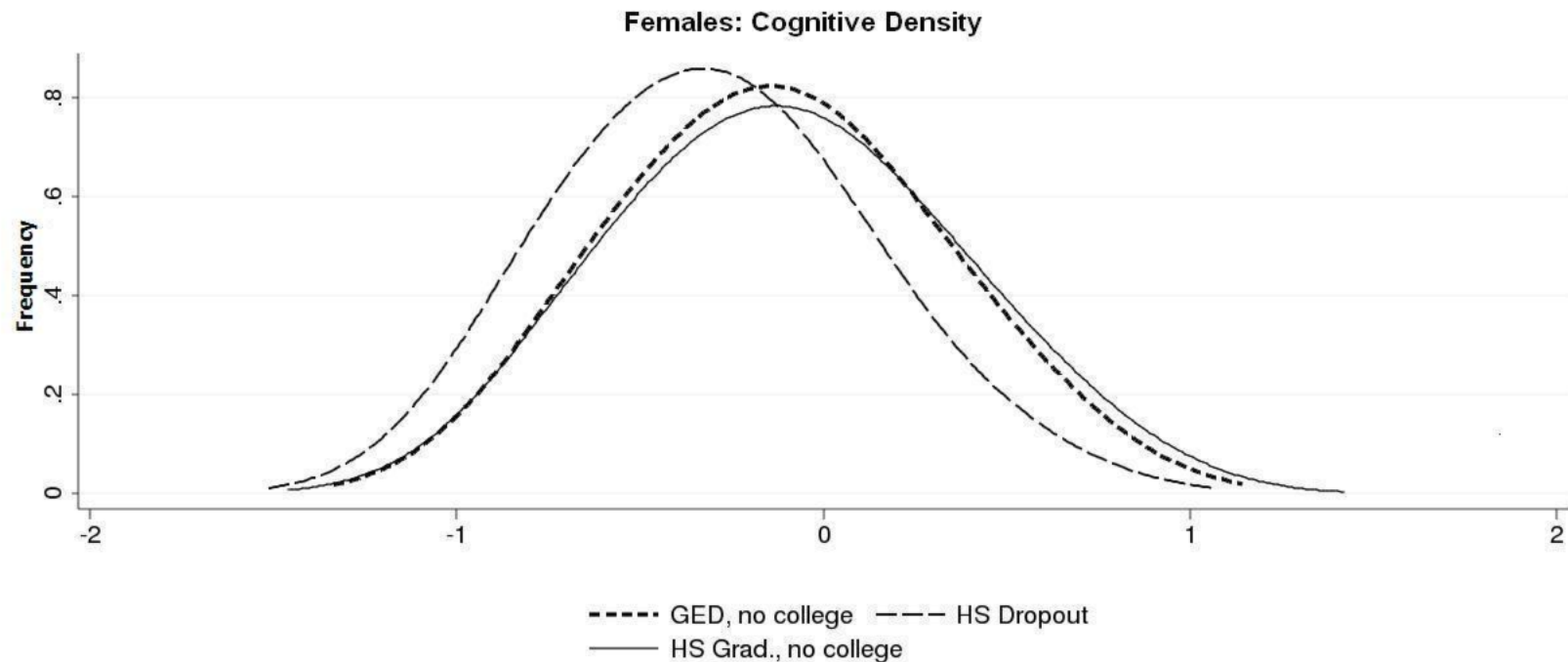
Easy Sample Question	<p>If $8x + 16 = 32$, what is x?</p> <p>A) 8 B) 2 C) 4 D) 3 E) 7</p>
Difficult Sample Question	<p>Alex has a job working for Adam's Apple Orchard. Two hundred new apple trees just arrived, which Mr. Adams would like Alex to plant. Alex can plant an average of 15 trees per workday. At this rate, approximately how many workdays will it take Alex to plant the 200 trees?</p> <p>A) between 7 and 9 B) between 9 and 11 C) between 11 and 13 D) between 13 and 15 E) between 15 and 17</p>

What are the GEDs?

- In terms of family background, they are intermediate between high school grads and dropouts.
- They are as smart as ordinary high school graduates who do not go on to college.

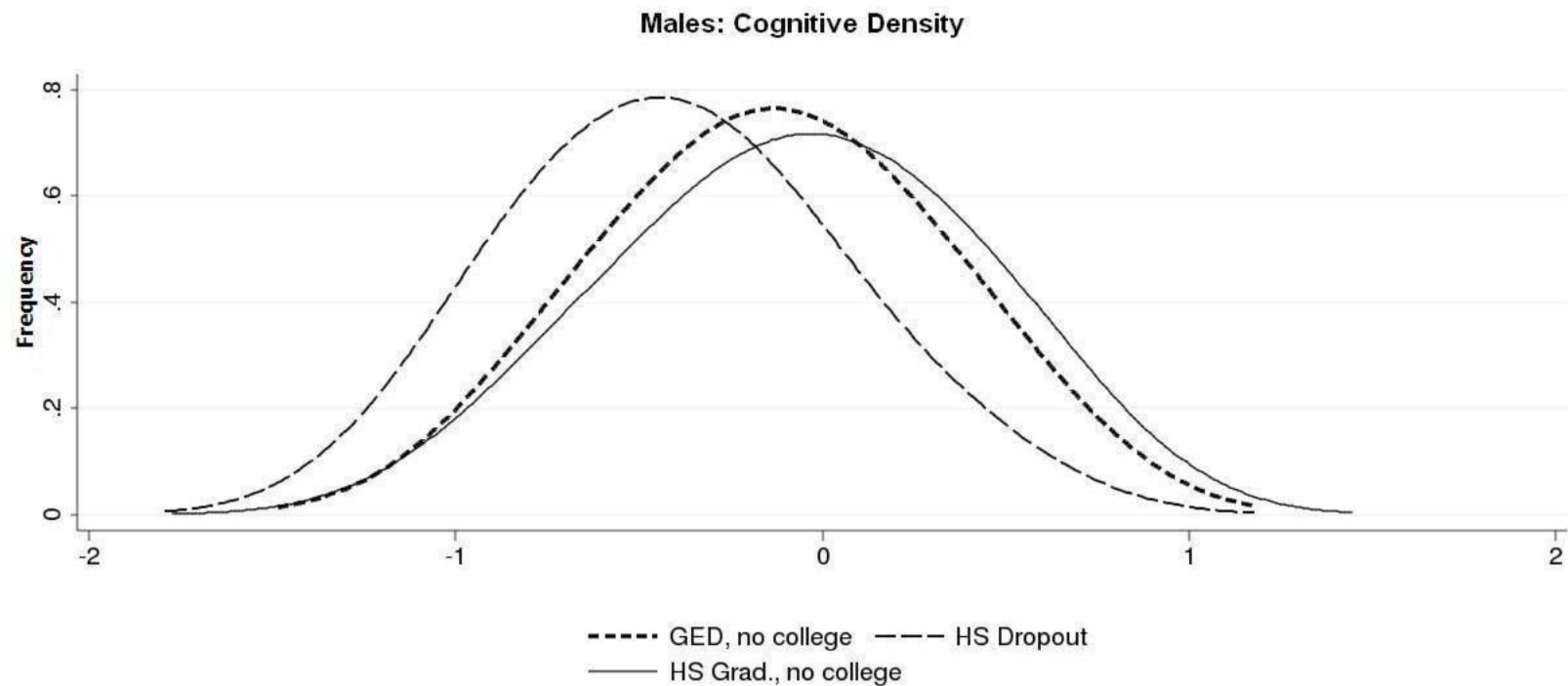
Cognitive Skills: Comparable for GEDs and High School Grads

Figure 5: Cognitive ability by educational status (no college sample, all ethnic groups)



Source: Heckman, Humphries, Urzua, and Veramendi (2010)

Figure 5: Cognitive ability by educational status (no college sample, all ethnic groups)

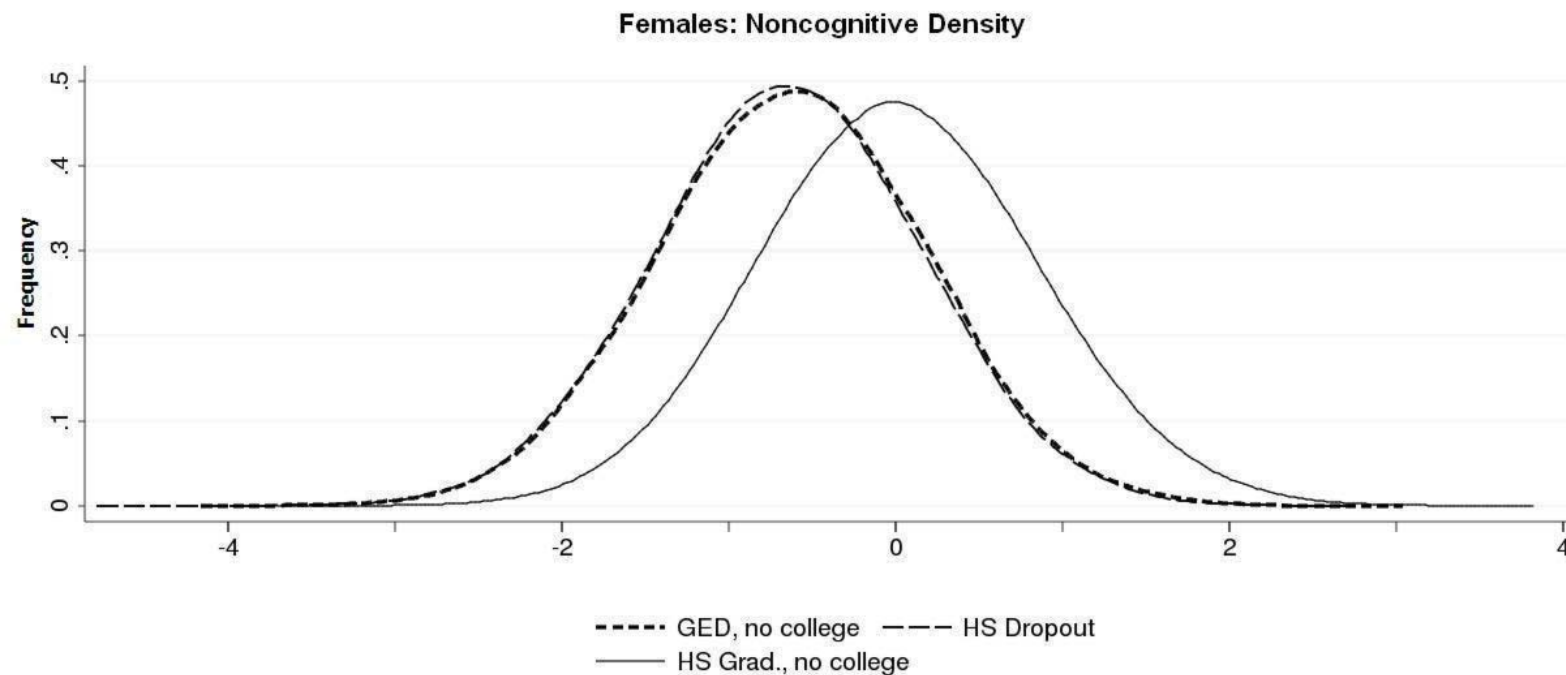


Source: Heckman, Humphries, Urzua, and Veramendi (2010)

- They lack noncognitive skills.

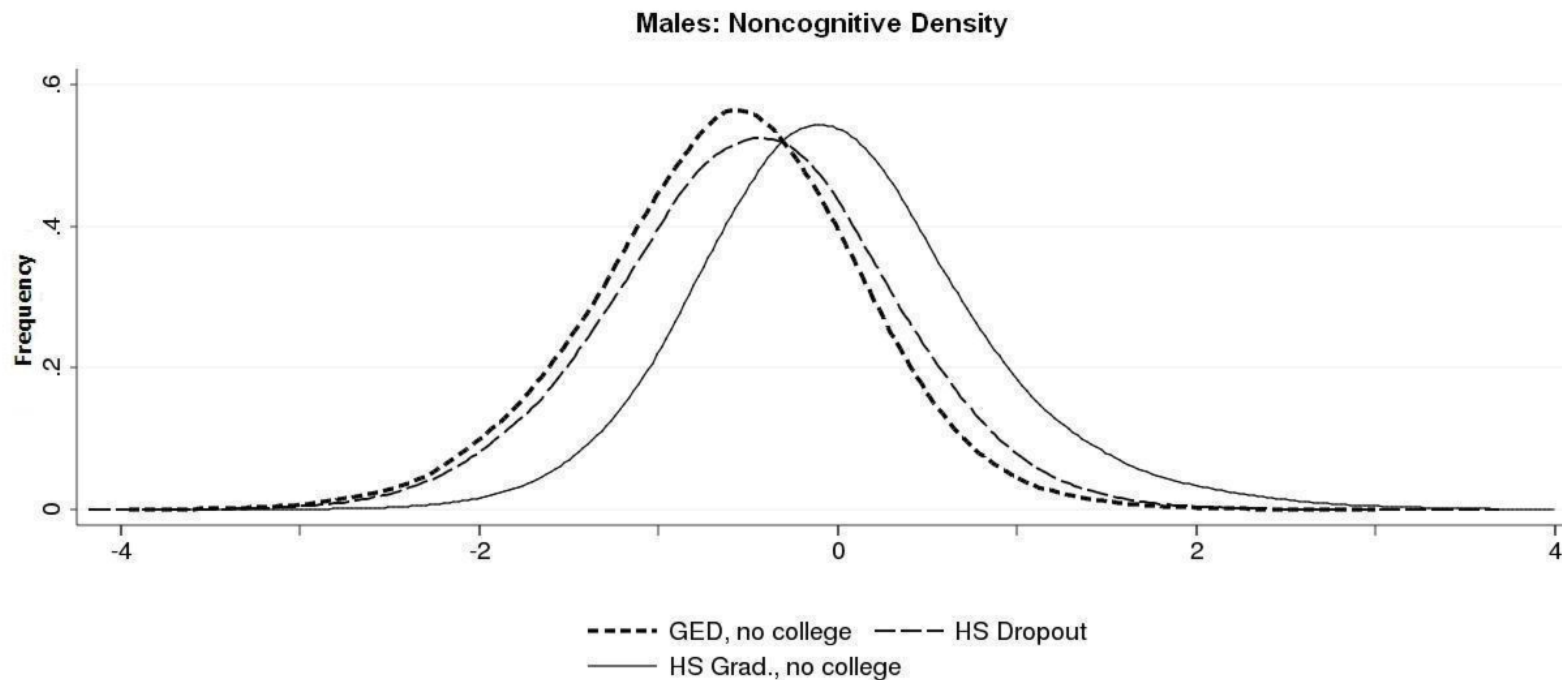
Noncognitive Skills: GEDs Resemble Dropouts

Figure 6: Noncognitive ability by educational status (no college sample, all ethnic groups)



Source: Heckman, Humphries, Urzua, and Veramendi (2010)

Figure 6: Noncognitive ability by educational status (no college sample, all ethnic groups)



Source: Heckman, Humphries, Urzua, and Veramendi (2010)

- Look at their performance in labor markets.
- Wages relative to dropouts
 1. Unadjusted (by cognitive ability): they have some advantage, but that advantage is there before they get the GED.
 2. Adjusted by ability: the effects vanish.

Figure 7: Economic gaps relative to dropouts: GEDs and high school graduates

Male Economic Gaps Relative to Dropouts: GEDs and High School Graduates

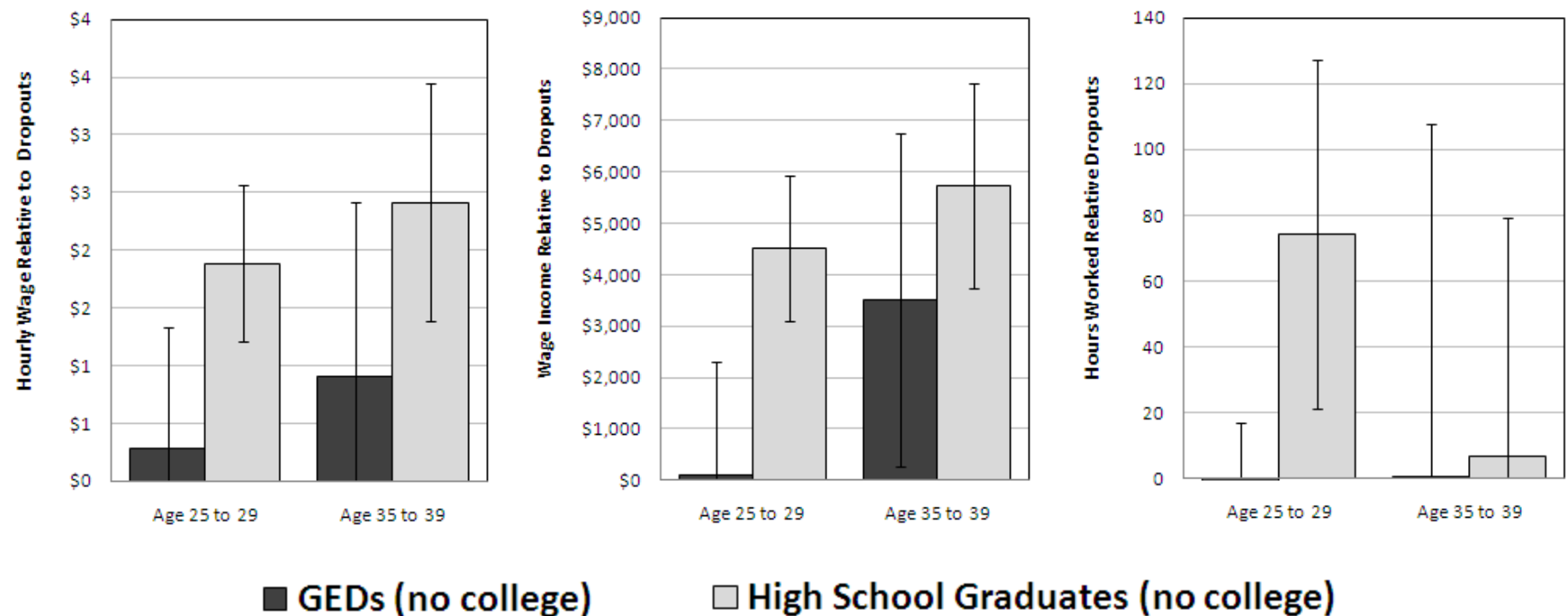
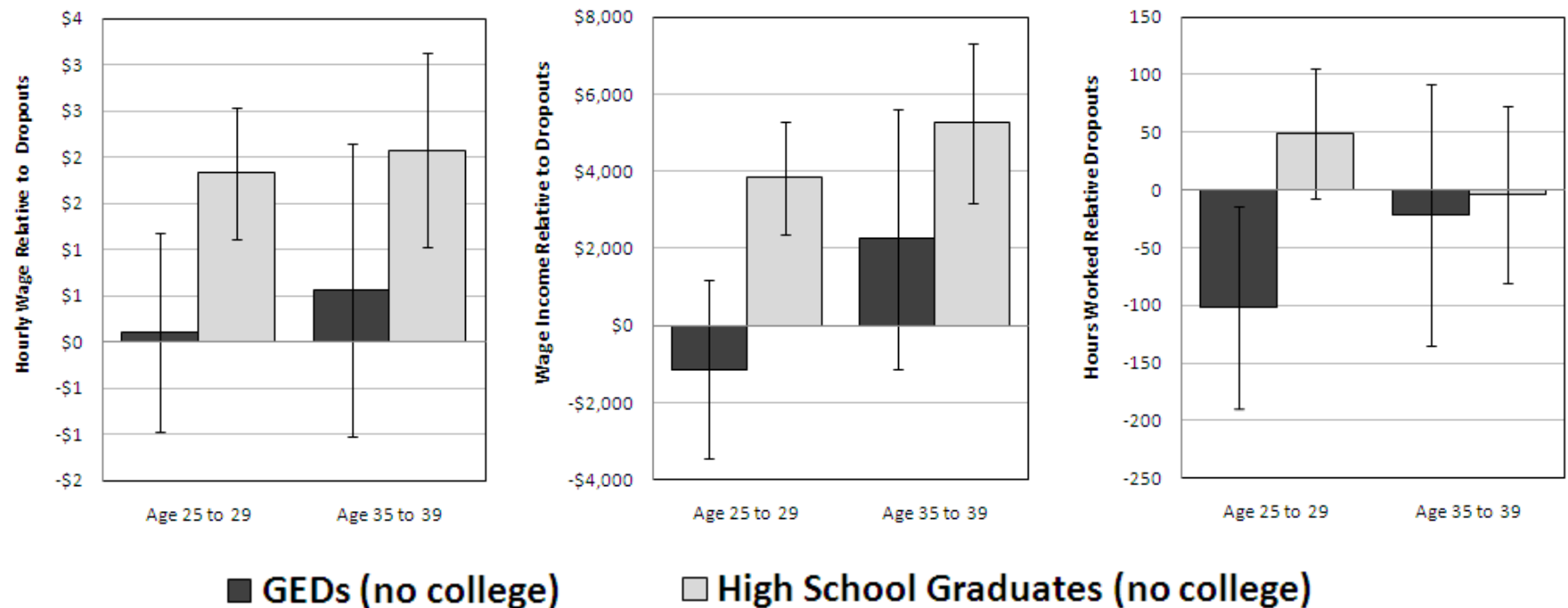


Figure 8: Ability-adjusted economic gaps relative to dropouts: GEDs and high school graduates

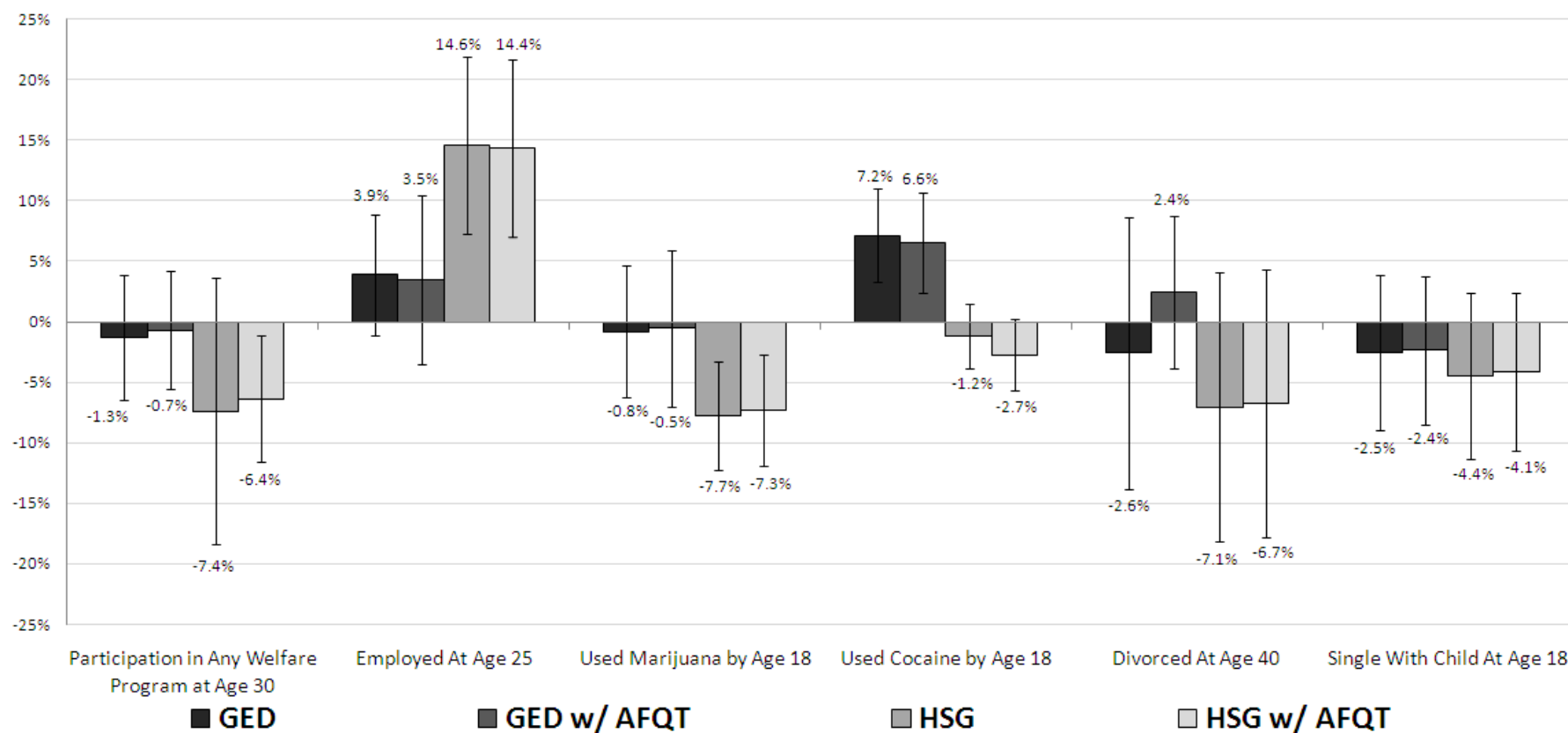
Male ability-adjusted economic gaps relative to dropouts: GEDs and high school graduates



The Social Performance of the GEDs

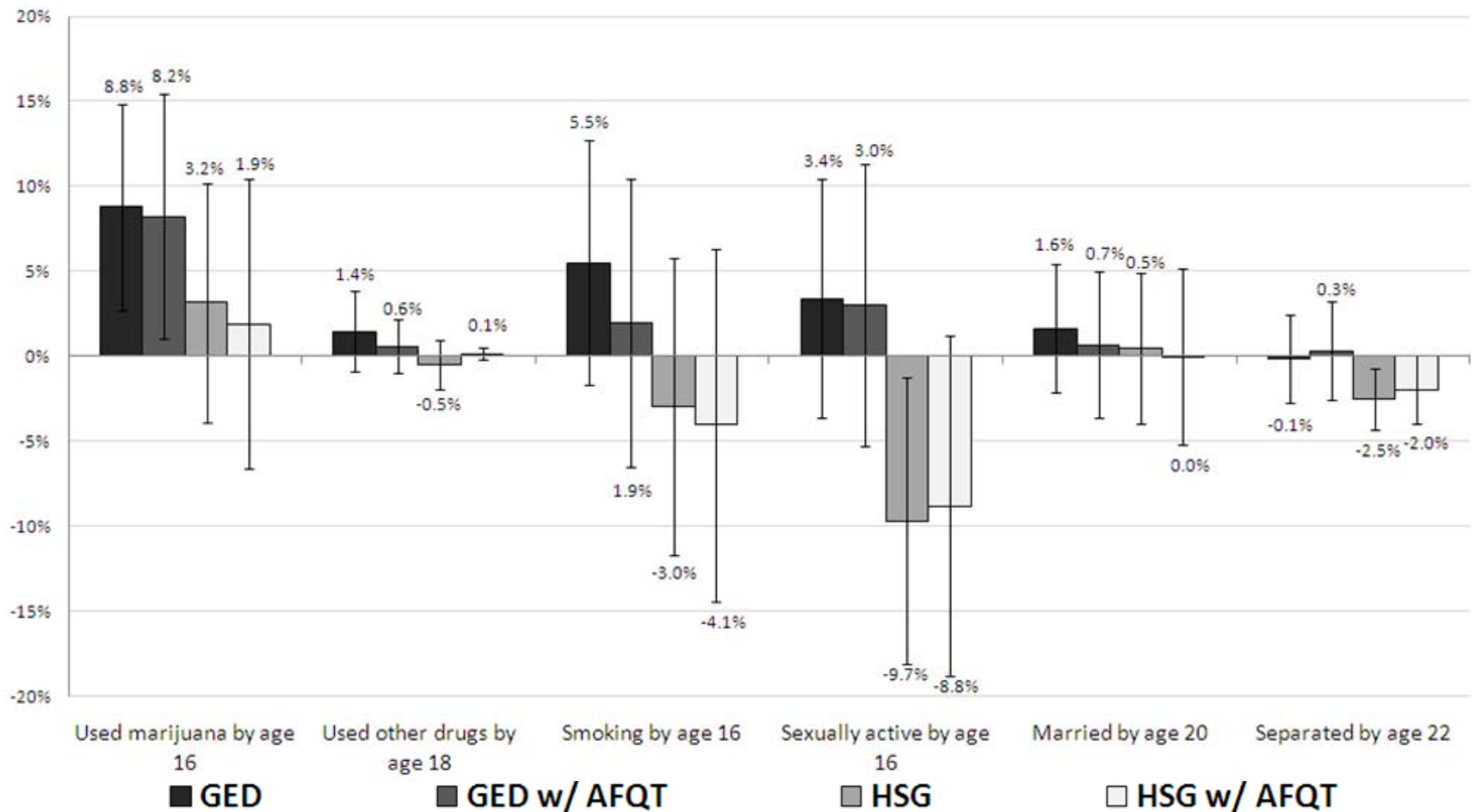
- Their behaviors are different.
- Often worse than dropouts.
- They drop out of everything they start (school; the army; marriage).
- More likely to engage in risky behaviors.

Figure 9: Gaps in the Probability of Various Social Outcomes Compared to High School Dropouts with and without controlling for scholastic ability (NLSY79). All demographic groups pooled unless otherwise noted.



Source: Heckman, Humphries, Urzua, and Veramendi (2010)

Figure 10: Gaps in the Probability of Various Social Outcomes Compared to High School Dropouts with and without controlling for scholastic ability (NLSY97). All demographic groups pooled unless otherwise noted.



Source: Heckman, Humphries, Urzua, and Veramendi (2010)

Signaling: GED may signal ability and this signal may be of value in the labor market

- Can examine this possibility by looking at before-after wages.
- No evidence of any gain from wages before the test is taken.

Native vs. Foreign Born: Another Signal?

- Clark and Jaeger (2006) argue that the GED might provide a signal of ability that is more familiar to employers than educational credentials earned outside the country, or may signal language ability and cultural assimilation.
- Clark and Jaeger use earnings data in the CPS and find that only foreign-born GEDs with no domestic credentials have a statistically significantly higher wage than native-born dropouts.

Native vs. Foreign Born: Another Signal?

- Heckman and LaFontaine (2006) examine the Clark and Jaeger (2006) analysis and find that their results are produced by data artifacts and limitations.
- No earnings premium to GED receipt for any group once selection and cognitive ability are accounted for (Heckman and LaFontaine, 2006).

Validities of Achievement Tests

- How were these widely used tests validated by their creators?
- Look at Predictive Validities (“effect sizes” or “correlations”).
- What do these tests predict? How strong is the predictive power of IQ, grades, and achievement tests?
- Contrast between What Tests Are Designed to Measure Vs. What They Actually Measure.
- Many tests of cognition are only validated on schooling tasks.

General Finding

1. The more complex the task, the greater the predictive power of cognition (“g” theory).
2. For less complex tasks, the greater the predictive power of personality. Soft skills are highly predictive in blue collar jobs.
3. But even in complex tasks, soft skills are highly predictive.

Table 1: Predictive Validities of Various Tests of Fluid and Crystallized Intelligence as Established by the Test Makers

Cognitive Achievement and IQ Tests				
Test	Domain over which it is validated	Estimated Validities	Source	Notes
SAT	First year college GPA	0.35 to 0.53	Validity of the SAT for Predicting First-Year College Grade Point Average	
ACT	Grades in early years of college	0.42	ACT Technical Manual	
Stanford-Binet	Correlations with other intelligence tests	0.77 to 0.87 with WISC-R	Rothlisburg (1987); Greene, Sapp, Chissom (1990)	
WISC (Wechsler Intelligence Scale for Children)	Correlations with academic achievement	WISC: 0.443 to 0.751 with WRAT tests, 0.482 to 0.788 with 1st grade grades, 0.462 to 0.794 with 2nd grade grades; WISC-R: 0.346 to 0.760 with WRAT tests, 0.358 to 0.537 with 1st grade grades, 0.420 to 0.721 with 2nd grade grades	Hartlage and Steele (1977)	WRAT = Wide Range Achievement Test; Ranges are given because correlations vary by academic subject
WAIS (Wechsler Adult Intelligence Scale)	Correlations with other intelligence tests, achievement tests, and outcomes	0.67 (median) with verbal tests, 0.61 (median) with nonverbal tests, 0.69 with education attained, 0.32 with employability of mentally challenged, 0.38 to 0.43 with college grades, 0.62 with high school grades, 0.14 with nursing grades	Feingold (1982)	

Table 1: Predictive Validities of Various Tests of Fluid and Crystallized Intelligence as Established by the Test Makers (Cont.)

Raven's Standard Progressive Matrices	Correlations with other intelligence tests	0.74 to 0.84 with WAIS-R	O'Leary, Rusch, Guastello (1991)	
GATB (General Aptitude Test Battery)	Supervisor rating performance in training programs and in job performance	0.23 to 0.65	Hunter (1986)	Large range due to variety of jobs
ASVAB (Armed Services Vocational Aptitude Battery)	Performance in military training programs and military attrition rates	0.37 to 0.78 for training (mean=0.56); - 0.15 for attrition	Schmidt (1988) for performance in training programs; Sticht et al (1982) for attrition rates	Large range in training correlations due to a variety of jobs
GED (General Educational Development)	Test difficulty is normed against graduating HS seniors. Test scores of high school seniors and grades of high school seniors	0.33 to 0.49 for HS Senior GPA	Technical Manual: 2002 Series GED Tests	
DAT (Differential Aptitude Tests)	Correlations with academic achievement	0.13 to 0.62 for college GPA	Omizo (1980)	Large range is due to varying validity of eight subtests of DAT
WIAT (Wechsler Individual Achievement Test)	Correlation with other achievement tests; teacher ratings of student achievement	0.80 with grade 4 CAT/2, 0.69 with grade 5 CAT/2, 0.83 with grade 6 CAT/2; 0.67 with teacher ratings	Michalko and Saklofske (1999)	CAT=California Achievement Test

Validities in Labor Market and Social Outcomes

Table 2: Validities in Labor Market Outcomes from the National Longitudinal Survey of Youth, 1979: Our Study

NLSY79 Correlation Table (tests and school performance)						
Outcomes	Males			Females		
	IQ	GPA (10th grade)	AFQT	IQ	GPA (10th grade)	AFQT
Hourly Wage Age 25	0.17***	0.17***	0.22***	0.14***	0.20***	0.25***
Hours Worked Age 25	0.08***	0.02	0.08***	0.14***	0.19***	0.28***
Wage Income Age 25	0.19***	0.17***	0.25***	0.21***	0.25***	0.36***
Weeks Worked Age 25	0.08***	0.04**	0.09***	0.16***	0.20***	0.30***
Weeks Unemployed Age 25	-0.14***	-0.11***	-0.18***	-0.12***	-0.11***	-0.12***
Weeks Out of Labor Force Age 25	-0.02	0.03	0.02	-0.11***	-0.15***	-0.26***
Total Jobs by Age 25	0.04	-0.08***	-0.04***	0.16***	0.03*	0.19***
Num. of Spouses/Partners by Age 25	-0.06**	-0.08***	-0.06***	0	-0.06***	-0.02
Any Welfare Age 25	-0.09***	-0.12***	-0.16***	-0.19***	-0.21***	-0.36***
Hourly Wage Age 35	0.03	0.05***	0.05***	0.11***	0.10***	0.13***
Hours Worked Age 35	0.10***	0.12***	0.21***	0.02	0.10***	0.17***
Wage Income Age 35	0.21***	0.21***	0.26***	0.08***	0.15***	0.19***
Weeks Worked Age 35	0.10***	0.15***	0.23***	0.11***	0.13***	0.23***
Weeks Unemployed Age 35	-0.10***	-0.11***	-0.15***	-0.17***	-0.11***	-0.14***
Weeks Out of Labor Force Age 35	-0.09**	-0.14***	-0.22***	-0.04	-0.11***	-0.18***
Total Jobs by Age 35	-0.02	-0.13***	-0.06***	0.09***	-0.02	0.18***
Num. of Spouses/Partners by Age 35	-0.05*	-0.10***	-0.05***	0.04	-0.05***	-0.01
Any Welfare Age 35	-0.09***	-0.11***	-0.23***	-0.20***	-0.23***	-0.36***

What About the Validity of Measures of Personality?

- Are they any better or worse than cognitive measures?
- Longitudinal and Latitudinal of Personality is Big Five: A lot of the study of validity has been done for this group.

Table 3: The Big Five Domains and their Facets

Factor	Facets	Definition of Factor	ACL ^a Marker Items for Factor
I. Openness to Experience (Intellect)	Fantasy, Aesthetics, Feelings, Actions, Ideas, Values	The degree to which a person needs intellectual stimulation, change, and variety.	Commonplace, Narrow-interest, Simple- vs. Wide-interest, Imaginative, Intelligent
II. Conscientiousness	Competence, Order, Dutifulness, Achievement striving, Self-discipline, Deliberation	The degree to which a person is willing to comply with conventional rules, norms, and standards.	Careless, Disorderly, Frivolous vs. Organized, Thorough, Precise

Table 3: The Big Five Domains and their Facets (Cont.)

III. Extraversion	Warmth, Gregariousness, Assertiveness, Activity, Excitement seeking, Positive emotions	The degree to which a person needs attention and social interaction.	Quiet, Reserved, Shy vs. Talkative, Assertive, Active
IV. Agreeableness	Trust, Straight-forwardness, Altruism, Compliance, Modesty, Tender-mindedness	The degree to which a person needs pleasant and harmonious relations with others.	Fault-finding, Cold, Unfriendly vs. Sympathetic, Kind, Friendly
V. Neuroticism (Emotional Stability)	Anxiety, Angry hostility, Depression, Self-consciousness, Impulsiveness, Vulnerability	The degree to which a person experiences the world as threatening and beyond his/her control.	Tense, Anxious, Nervous vs. Stable, Calm, Contented

Personality Test Validations as Conceived of By Their Creators

Table 4: Predictive Validities of Various Personality Tests. Personality Evaluations

Test	Domain of Validation	Estimated Validities	Source	Notes
Hogan Personality Inventory	Correlations with delinquency criterion; Factor correlations with outcomes	0.00 to 0.67 with School Success, 0.68 to 0.73 with Avoids Trouble, 0.22 to 0.33 with Non-experience Seeking, -0.44 to 0.01 with Enjoys Crowds, -0.42 to 0.09 with Exhibitionist, 0.25 to 0.43 with Easy to Live With, 0.36 to 0.44 with Good Sense of Attachment, 0.10 to 0.43 with Not Depressed, 0.26 to 0.54 with No Guilt; Delinquency factor correlates: 0.91 with chargeable accidents, 0.80 with warning letters, 0.44 with suspensions; Absenteeism factor correlates: 0.62 with grievances, 0.61 with absences, 0.55 with medical absences, 0.44 with workers compensation claims; Negative Sanctions factor correlates: 0.68 with suspension letters, 0.67 with discharges; No Fault factor correlates: 0.71 with nonchargeable accidents; Supervisor's Ratings factor: 0.60 with supervisor's ratings, -0.38 with health history	Hogan & Hogan (1989)	
Myers-Briggs Type Indicator	Correlations with other personality tests; agreement between reported personality type and best-fit personality type	Correlation with Big Five based on Adjective Check List: -0.70 (E-I to Extraversion), 0.44 (S-N to Openness), 0.47 (T-F to Agreeableness), -0.54 (J-P to Conscientiousness); 72.9% report same four preferences as best-fit type, 18.2% report same three out of four preferences as best-fit type	Schaubhut, Herk, Thompson (2009)	
NEO PI-R (Revised NEO Personality Inventory)	Correlations with other personality tests	Correlation with Positive Presentation Management Scale: -0.60 (N), 0.48 (E), 0.04 (O), 0.25 (A), 0.41 (C); correlations with Negative Presentation Management Scale: 0.39 (N), -0.46 (E), -0.31 (O), -0.38 (A), -0.54 (C); correlations with Big Five Index: 0.76 (E), 0.66 (A), 0.70 (C), -0.66 (N), 0.68 (O); correlations with Ten Item Personality Inventory: 0.65 (E), 0.59 (A), 0.68 (C), -0.66 (N), 0.56 (O)	Yang, Bagby, Ryder (2000); Gosling, Rentfrow, Swann (2003)	N=neuroticism, E=extraversion, O=openness, A=agreeableness, C=conscientiousness

Table 4: Predictive Validities of Various Personality Tests, Personality Evaluations (Cont.)

Test	Domain of Validation	Estimated Validities	Source	Notes
NEO-FFI (NEO Five Factor Inventory)	Correlations with other personality tests	0.73 overall with BFI (Big-Five Index)	Gosling, Rentfrow, Swann (2003)	Note: This is a shorter version of the NEO PI-R
Rotter Locus of Control	Correlation with high school GPA	Correlation with high school GPA is 0.09 in suburban schools, 0.26 in inner-city schools	Stipek & Weisz (1981)	
Rosenberg Self-Esteem Scale	Correlations with other self-esteem scales	0.73 to 0.80 with Single Item Self-Esteem Scale; 0.15 to 0.76 with Harter's Self-Perception Profile for Adolescents	Robins, Hendin, Trzeniewski (2001); Hagborg (1993)	Correlations with Harter's done on an item by item basis
Short GRIT Scale	Item-level correlations with outcomes	0.03 to 0.13 for West Point 2008 Retention, 0.00 to 0.11 for West Point 2010 Retention, -0.05 to 0.17 for Spelling Bee success, 0.03 to 0.32 for Ivy League GPA	Duckworth & Quinn (2009)	Large ranges due to variety of items

Table 5: The Relative Predictive Power of Conscientiousness and SAT Scores for College GPA

Source	Sample	Timing of Measurement and Outcome	Controls	Metric	Results	
Conard [2005]	University students in the US (N=186)	College GPA and SAT were both self-reported during college. Personality was measured in college.	Class Attendance	Standardized Regression Coefficient (β)	SAT Total	0.27
					Conscientiousness	0.30
Nofle and Robins [2007]	University students in the US (N=10,497)	College GPA and SAT were both self-reported during college. Personality was measured in college.	Gender, Other Big Five Traits	Standardized Regression Coefficient (β)	SAT Verbal	0.19
					SAT Math	0.16
					Conscientiousness	0.24
Nofle and Robins [2007]	University students in the US (N=507)	College GPA and SAT were both self-reported during college (a). Personality was measured in college.	Gender, Other Big Five Traits	Standardized Regression Coefficient (β)	SAT Verbal	0.13
					SAT Math	0.28
					Conscientiousness	0.18
Nofle and Robins [2007]	University students in the US (N=470)	College GPA and SAT were both self-reported during college. Personality was measured in college.	Gender, Other Big Five Traits	Standardized Regression Coefficient (β)	SAT Verbal	0.18
					SAT Math	0.25
					Conscientiousness	0.22
Wolfe and Johnson[1995]	University students in the US (N=201)	GPA and SAT were provided by the Colleges's Record Office. Personality was measured in college.	High School GPA	Standardized Regression Coefficient (β)	SAT Total	0.23
					Conscientiousness	0.31

Notes: (a) Self-reported SAT scores and those obtained from college records were highly correlated ($r=0.92$). Self-reported GPA and that obtained from college records were highly correlated ($r=0.89$).

Figure 11: Association of the Big Five and intelligence with years of schooling

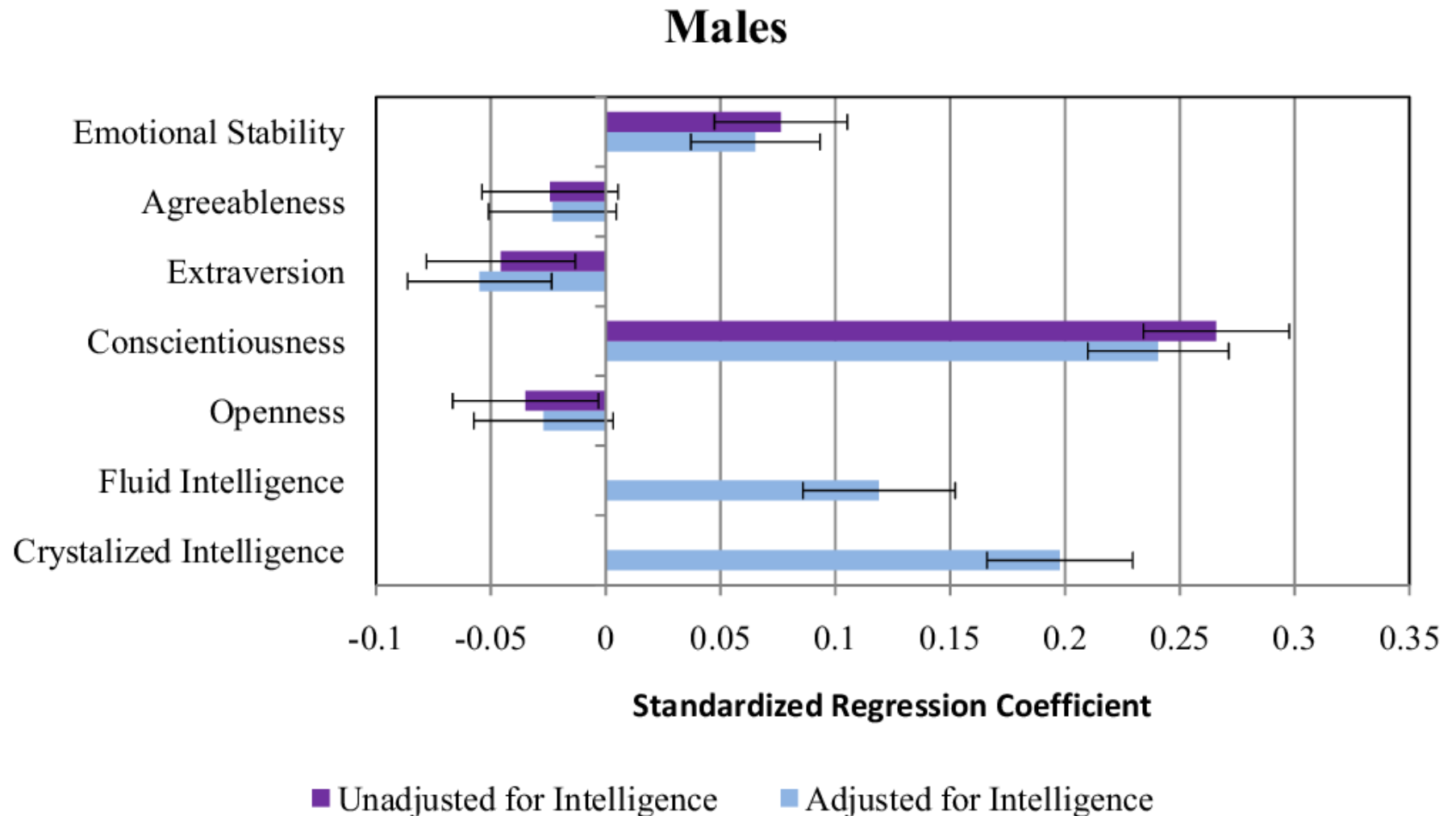


Figure 12: Correlations of The Big Five and Intelligence with Course Grades

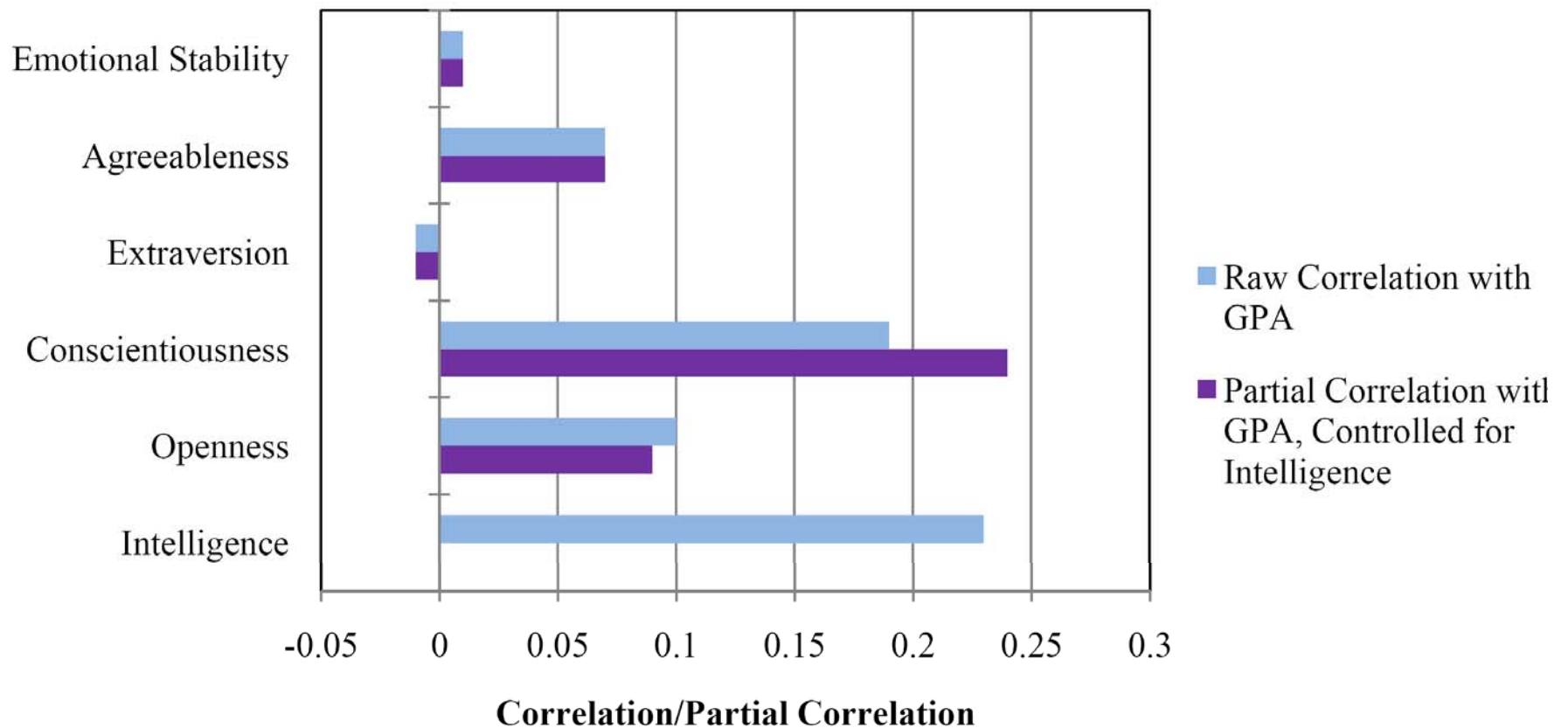


Figure 13: Associations with Standardized Achievement Test Scores

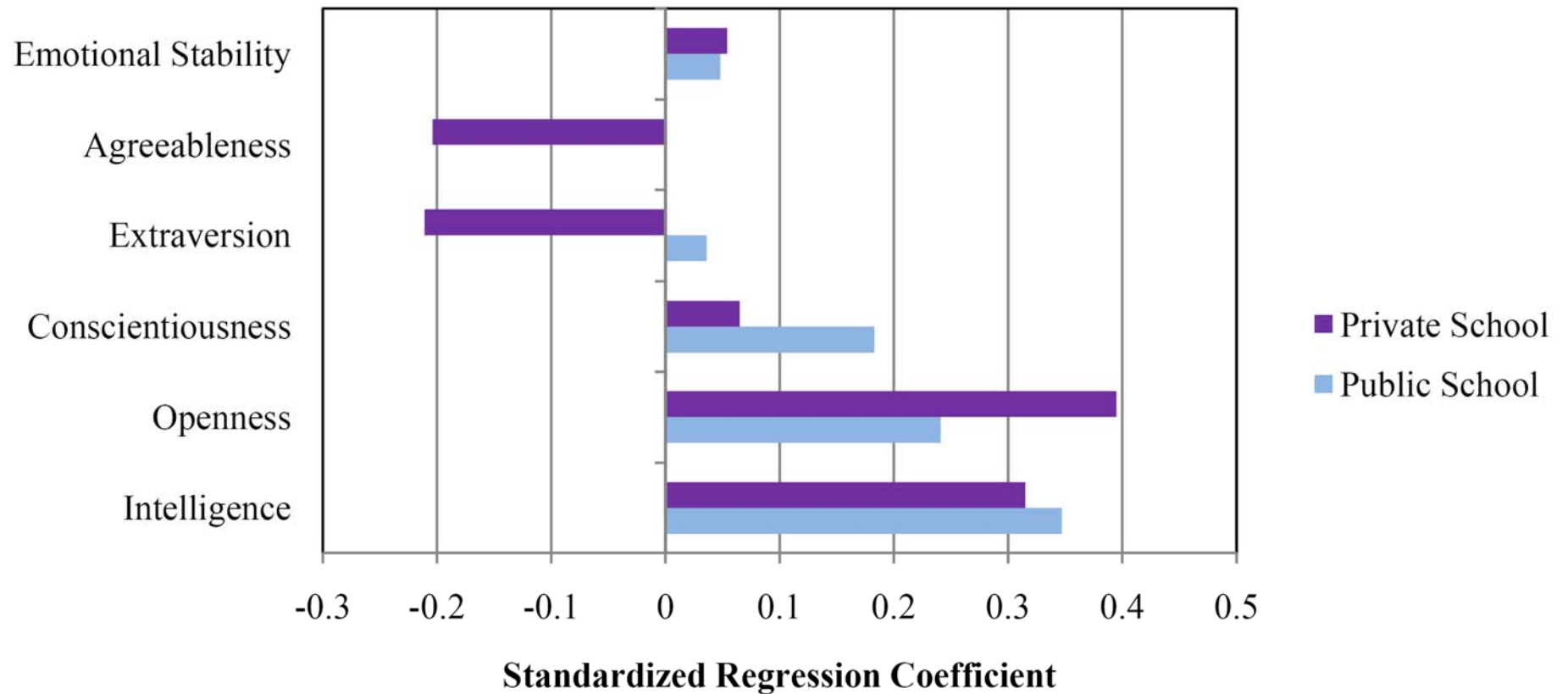


Figure 14: Associations with Job Performance

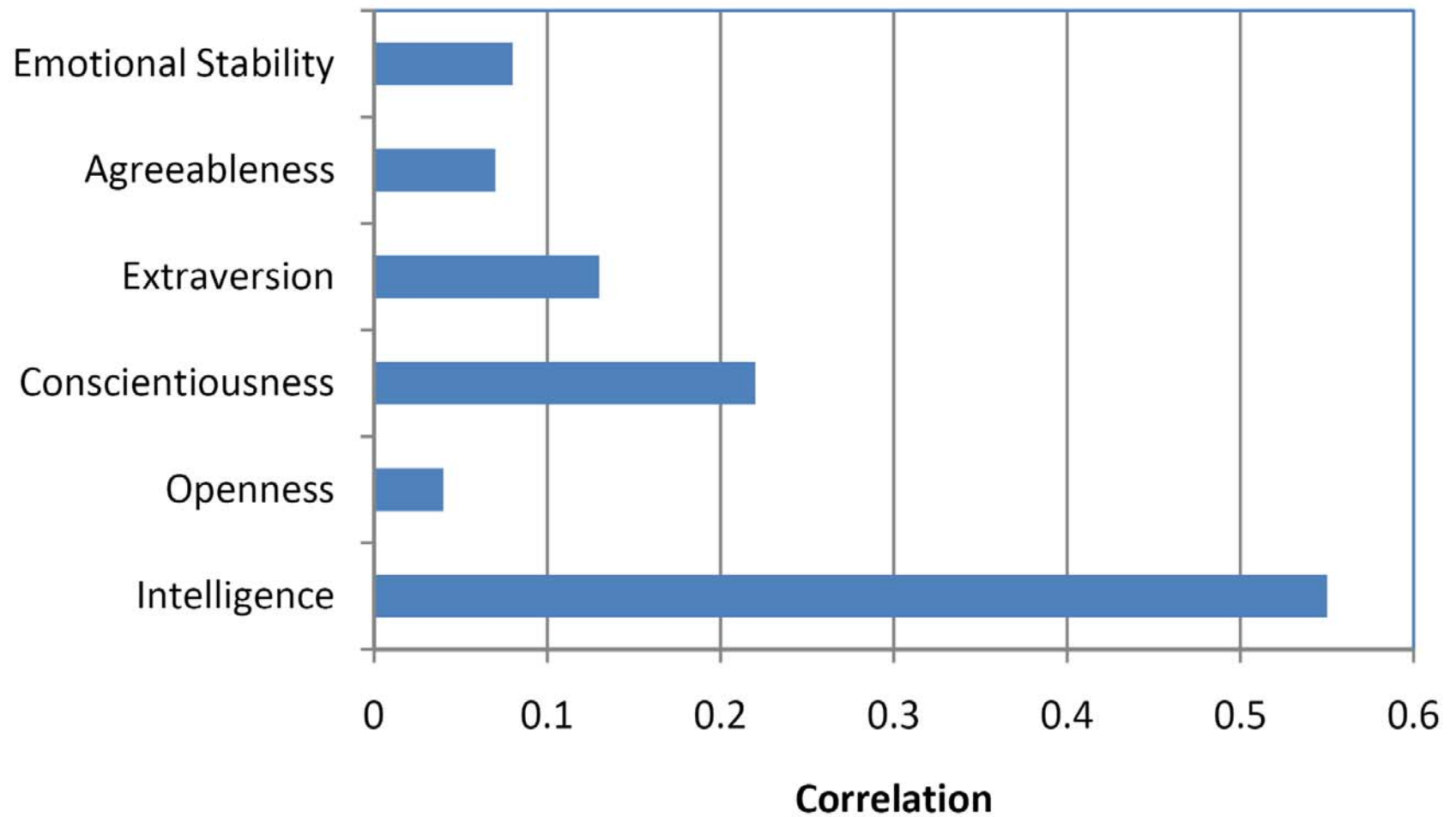


Figure 15: Correlations of mortality with personality, IQ, and socioeconomic status (SES)

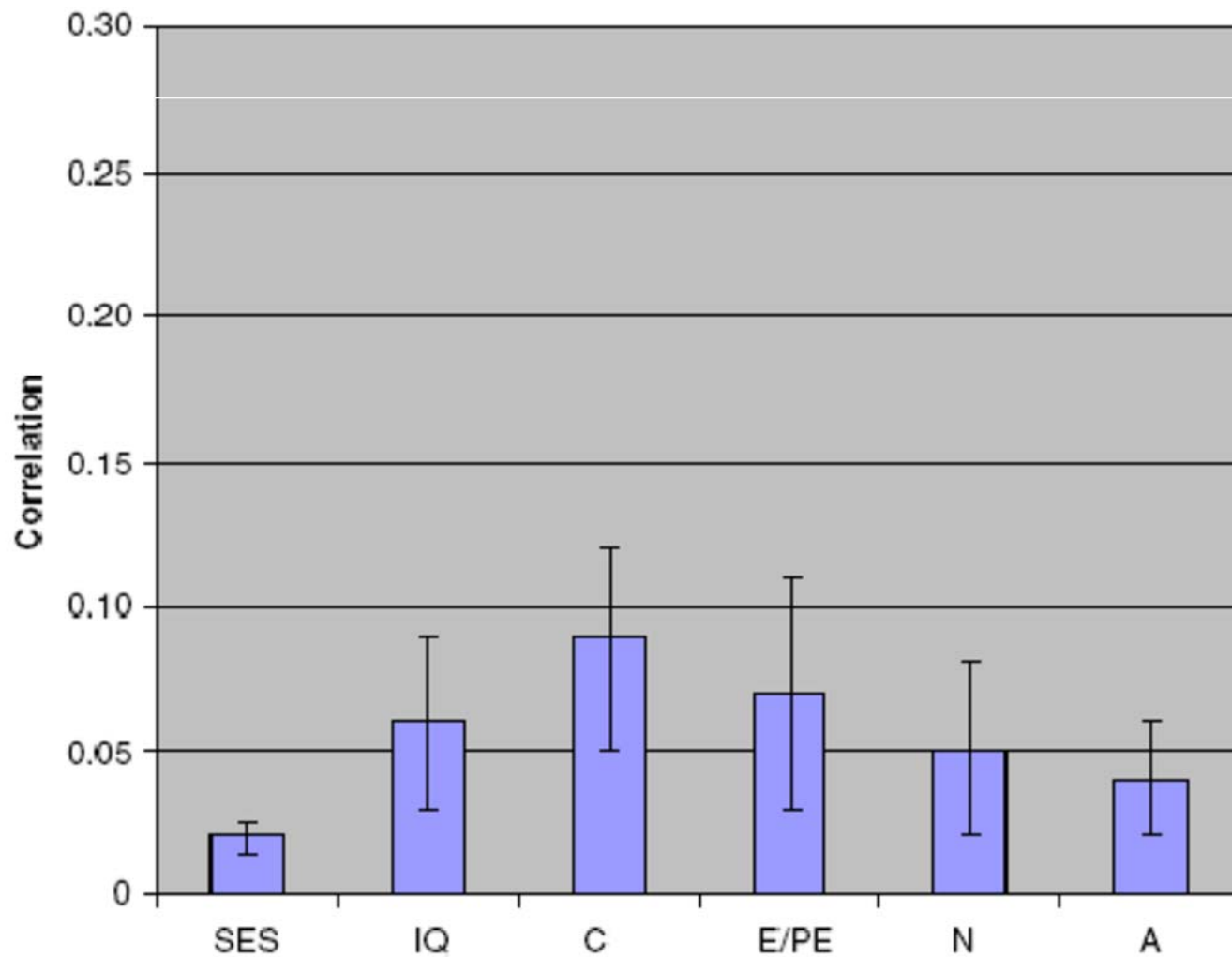
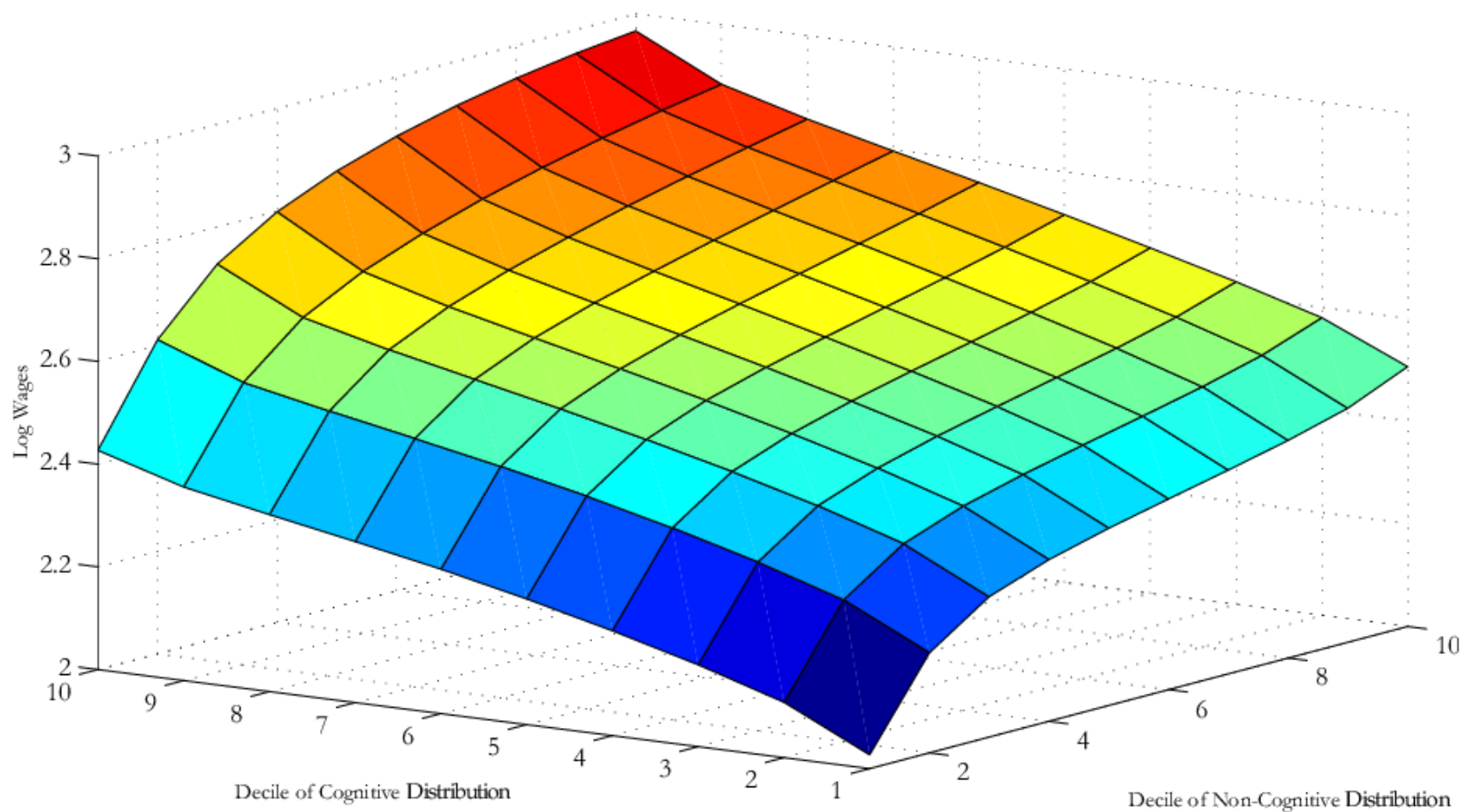
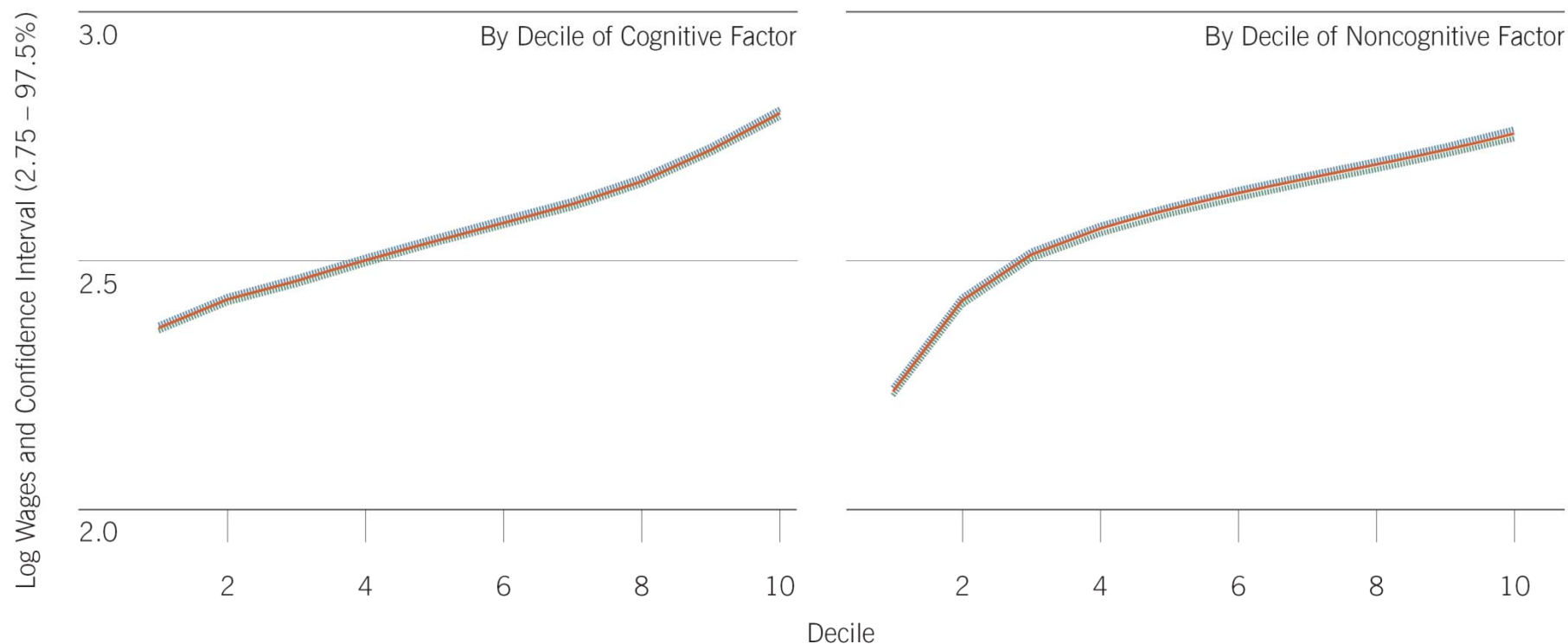


Figure 16: Mean log wages by age 30 (males)



Source: Heckman, Stixrud, and Urzua (2006)

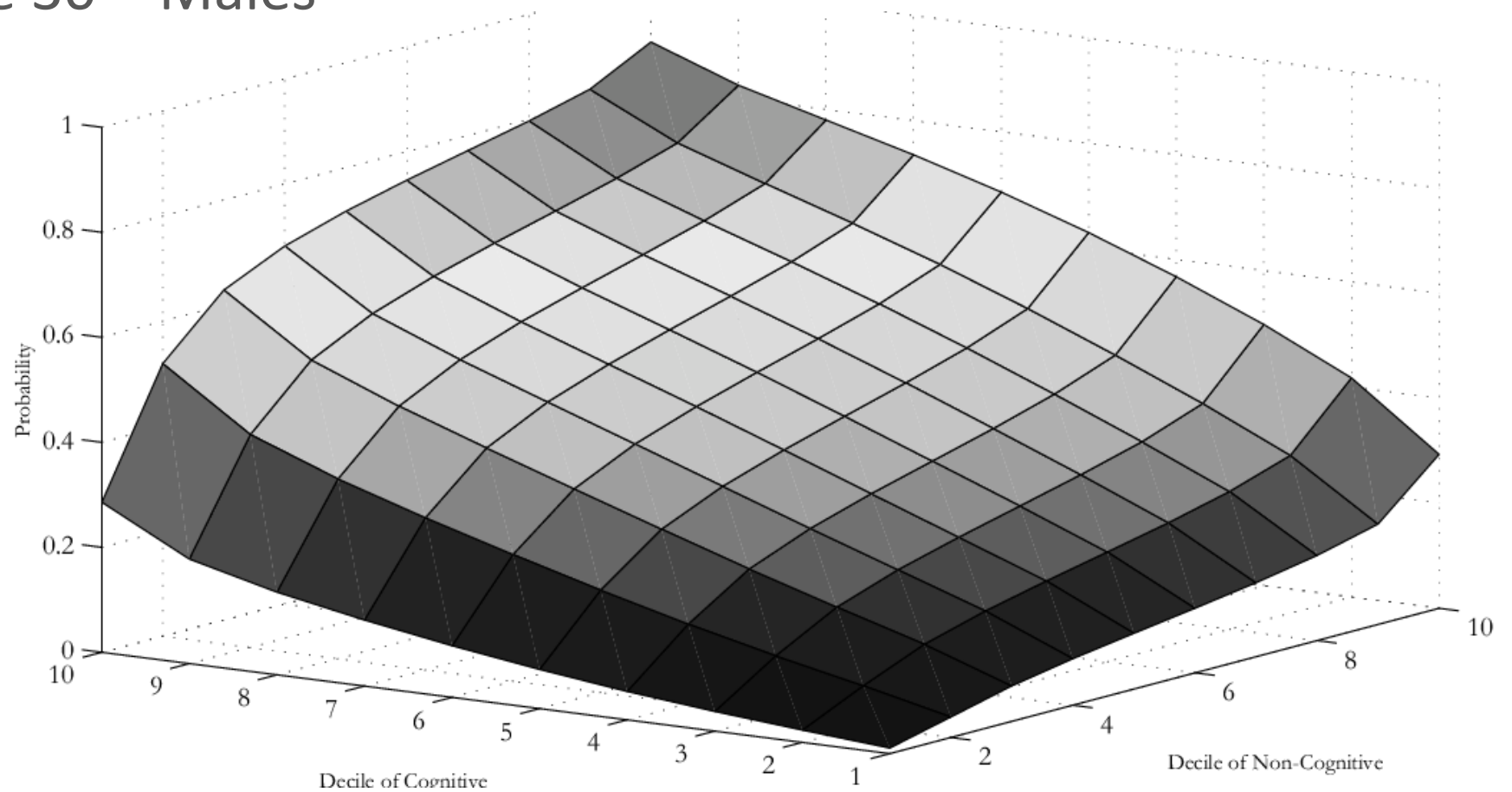
Figure 16: Mean log wages by age 30 (males)



Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (50 draws).

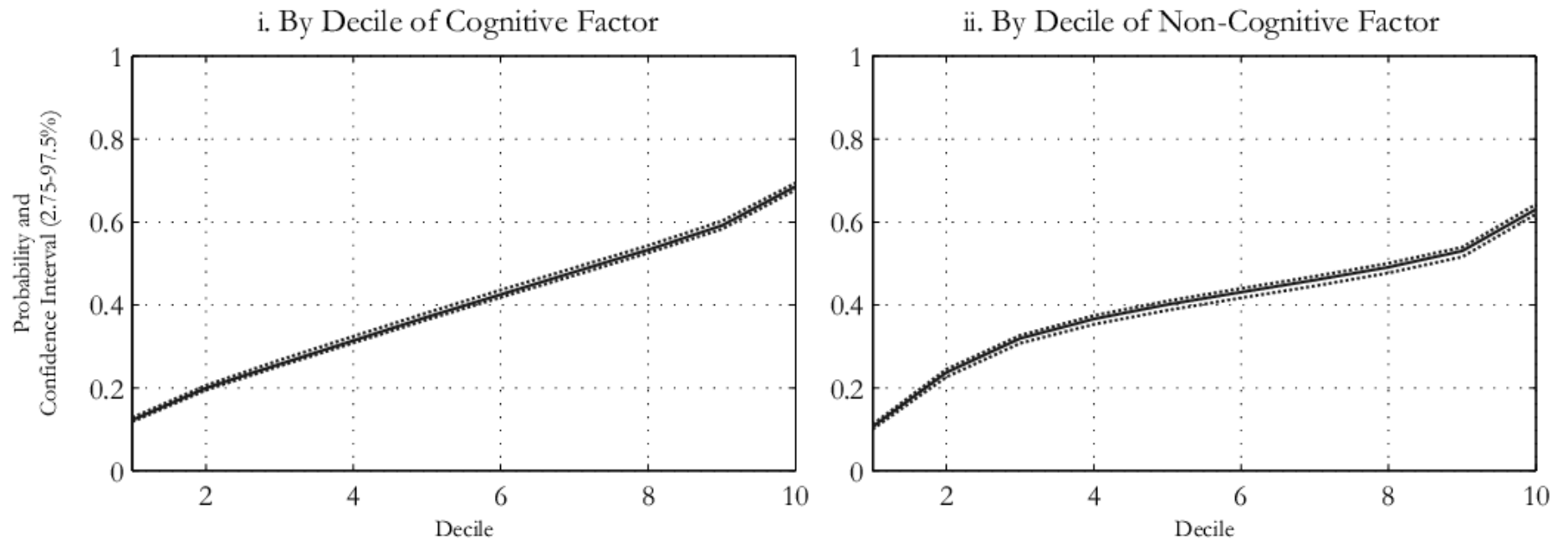
Source: Heckman, Stixrud, and Urzua (2006)

Figure 17: Probability of Being a White Collar Worker by Age 30 – Males



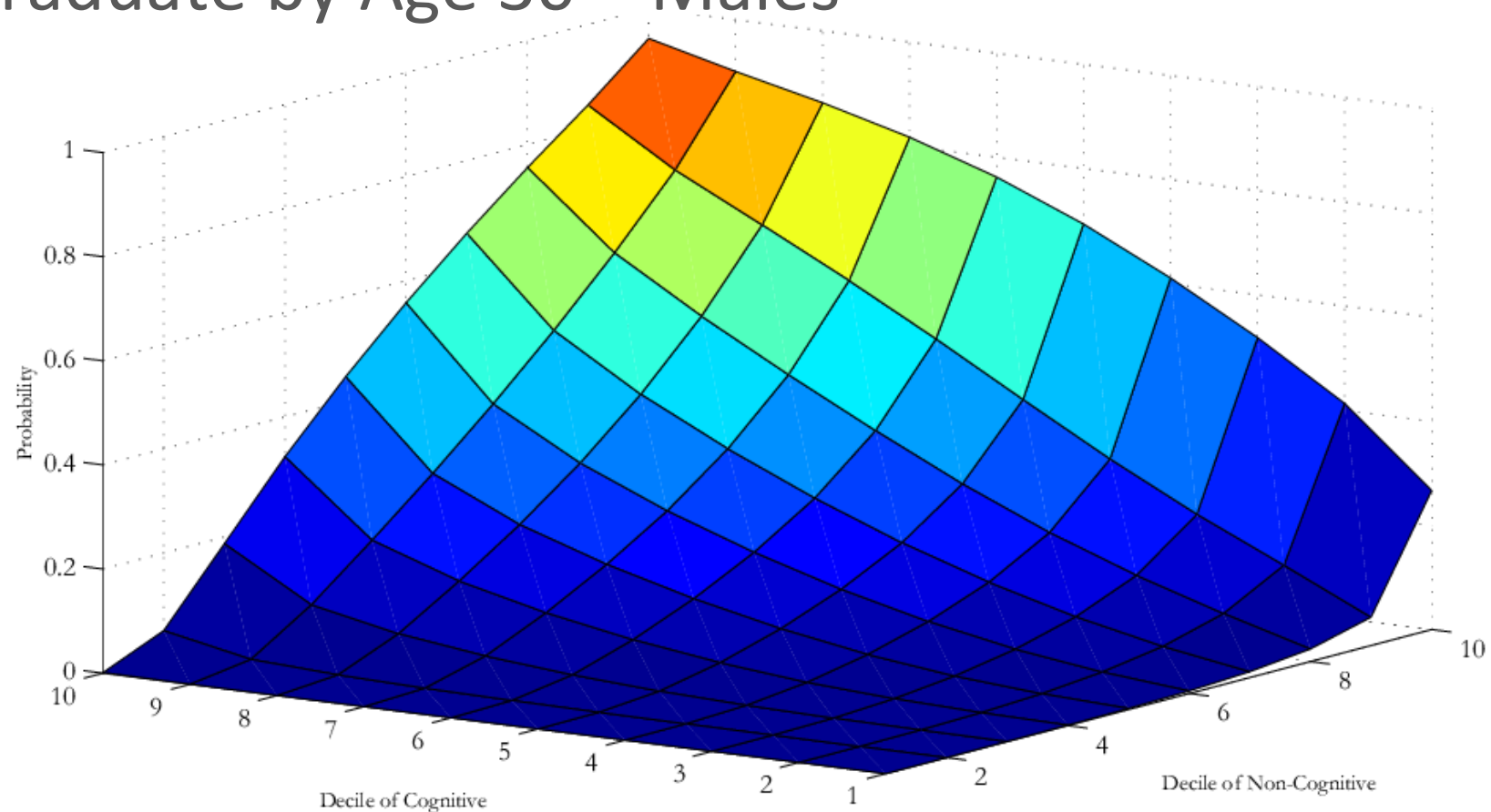
Source: Heckman, Stixrud, and Urzua (2006)

Figure 17: Probability of Being a White Collar Worker by Age 30 – Males



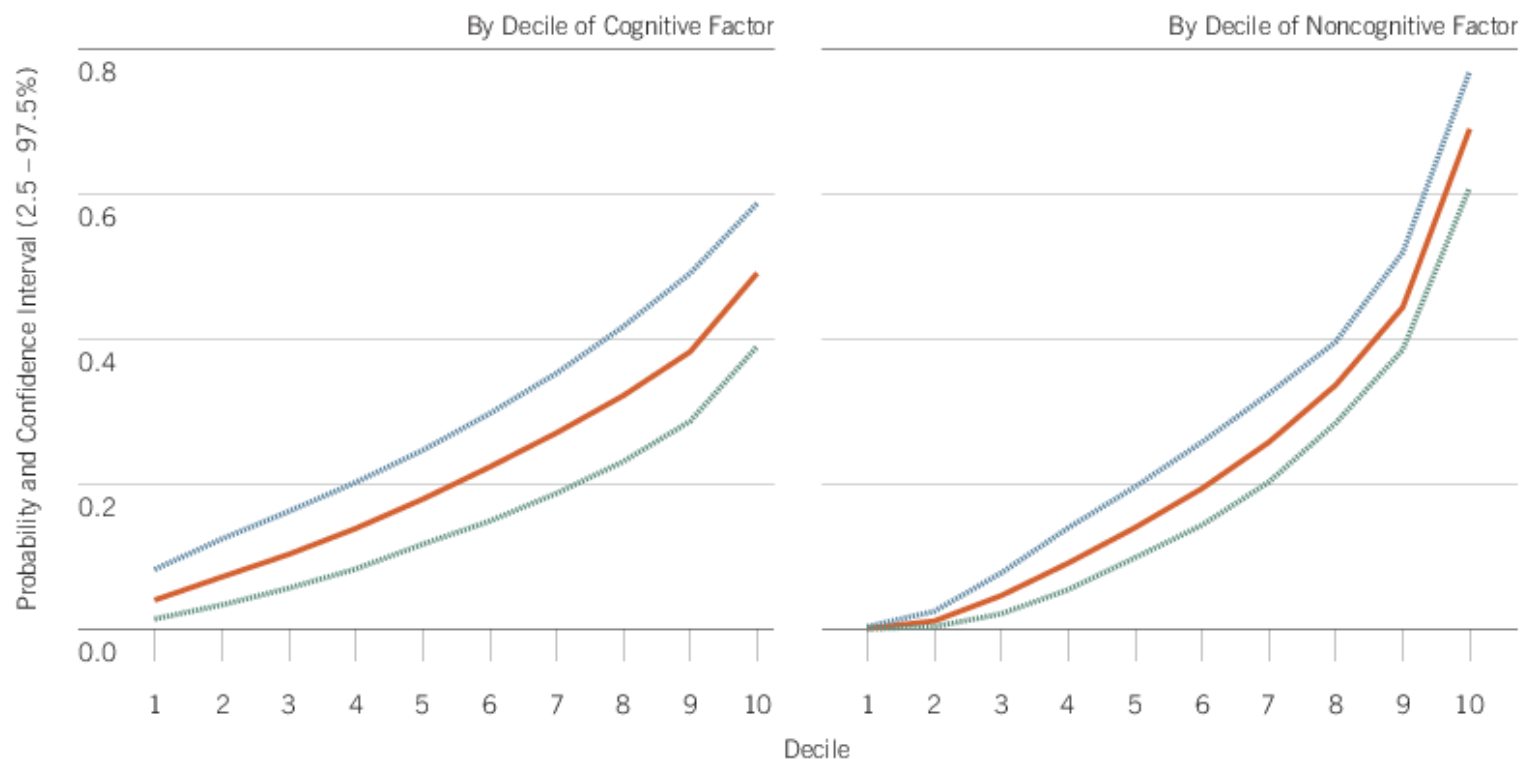
Source: Heckman, Stixrud, and Urzua (2006)

Figure 18: Probability of Being a 4-yr College Graduate by Age 30 – Males



Source: Heckman, Stixrud, and Urzua (2006)

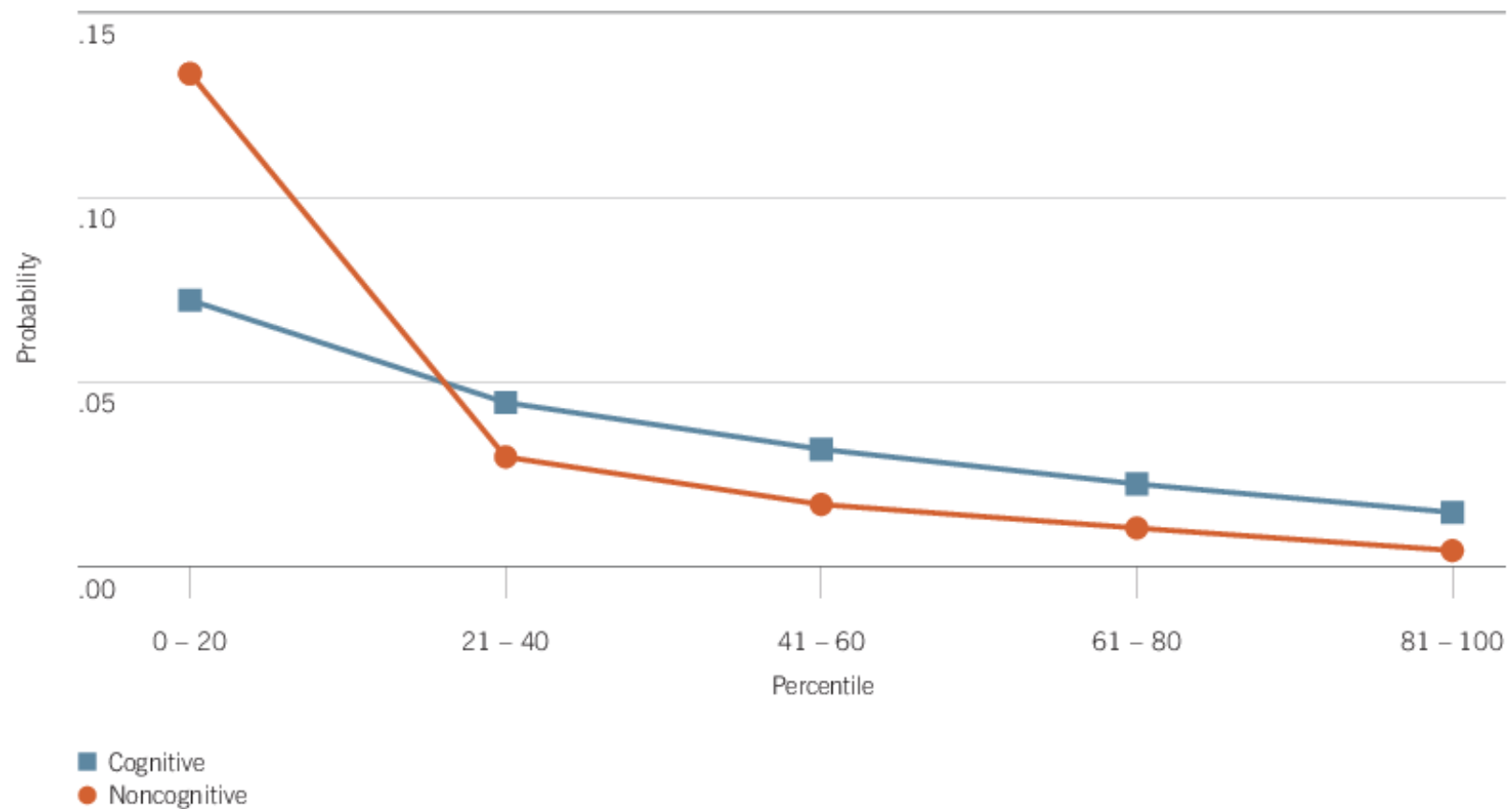
Figure 18: Probability of Being a 4-yr College Graduate by Age 30 – Males



Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws).

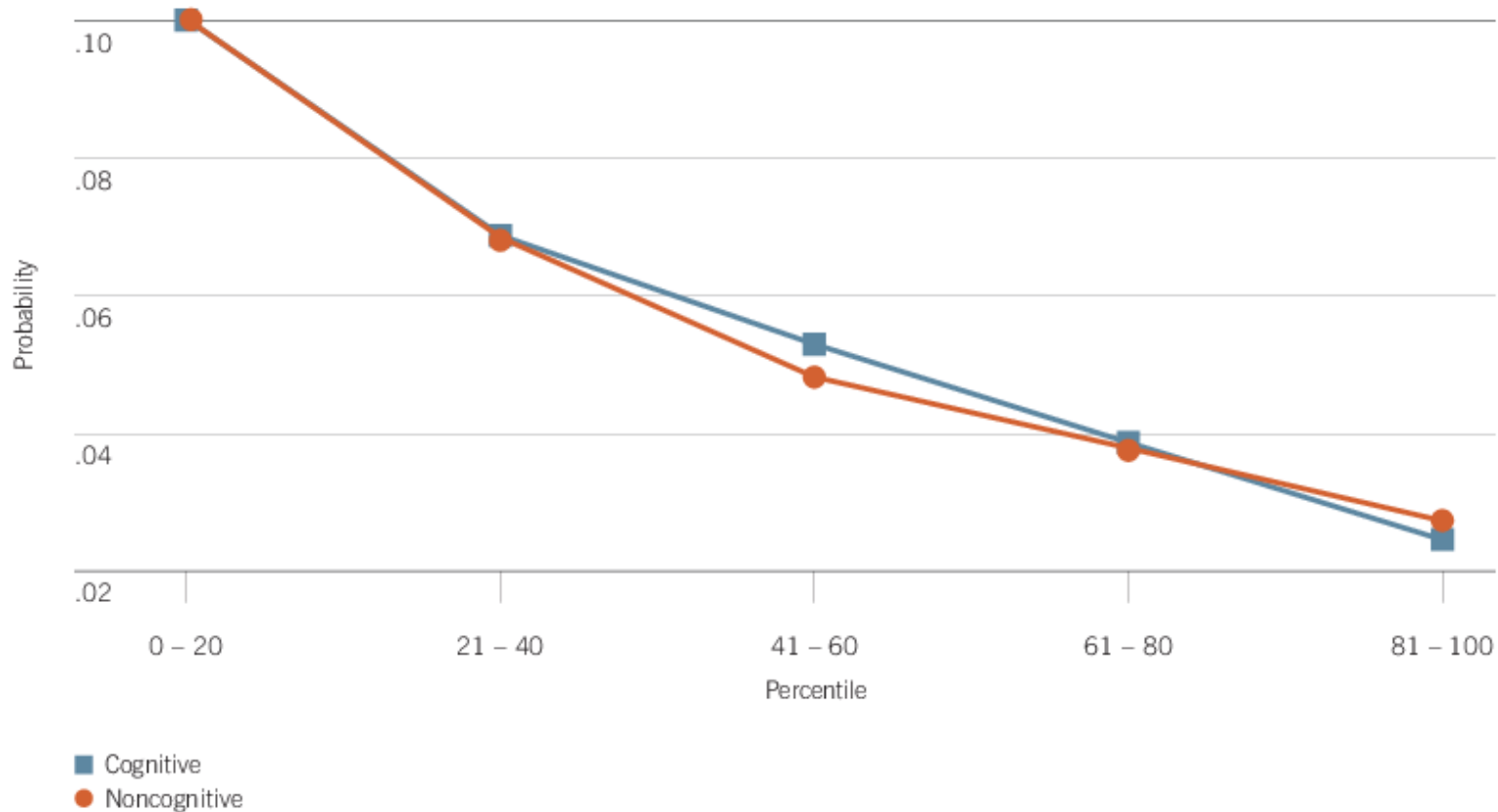
Source: Heckman, Stixrud, and Urzua (2006)

Figure 19: Ever been in jail by age 30, by ability (males)



Source: Heckman, Stixrud, and Urzua (2006)

Figure 20: Probability of being teenage and single with children (females)



Source: Heckman, Stixrud, and Urzua (2006)

Figure 21: Probability of daily smoking by age 18 (males)

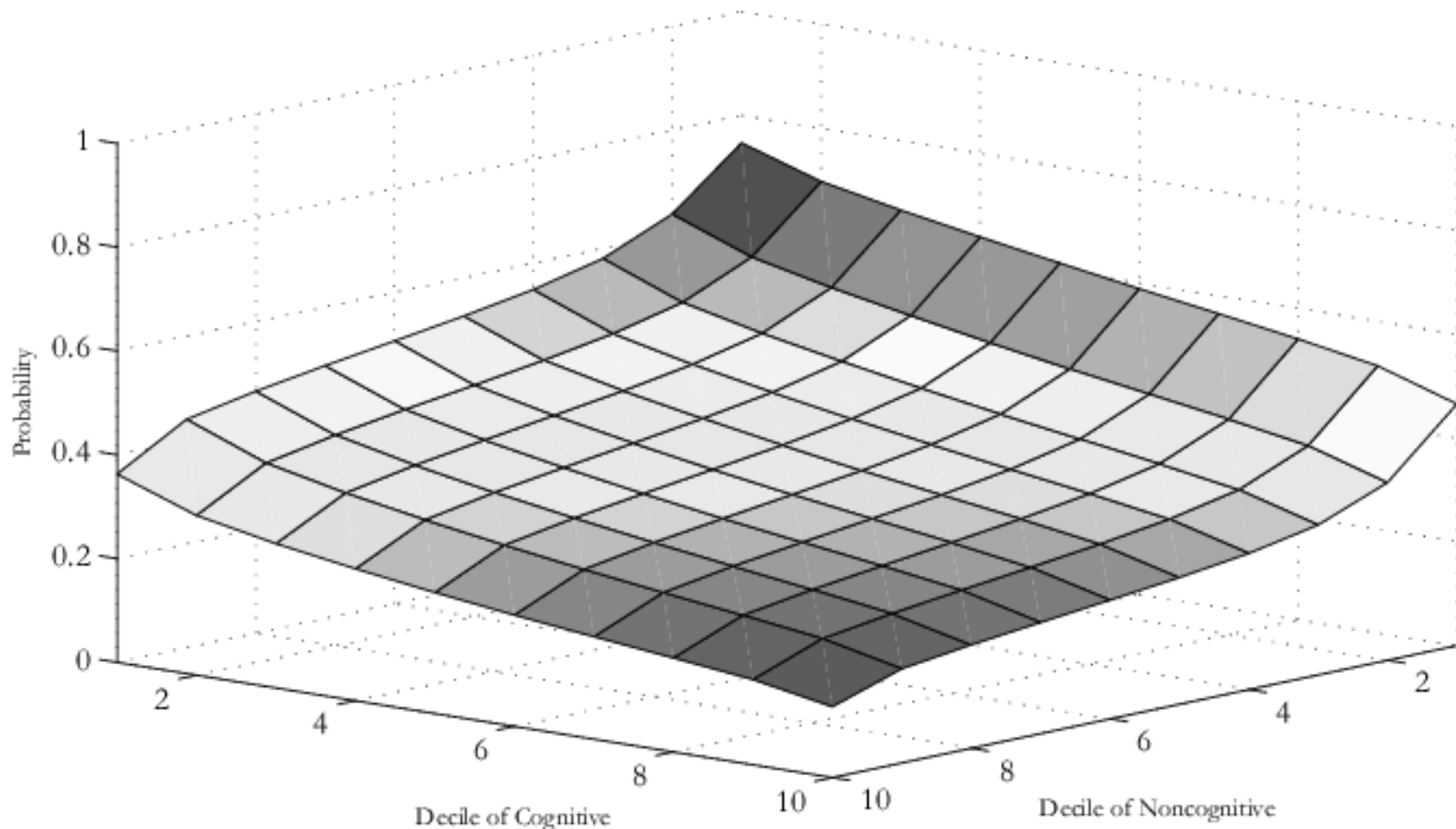
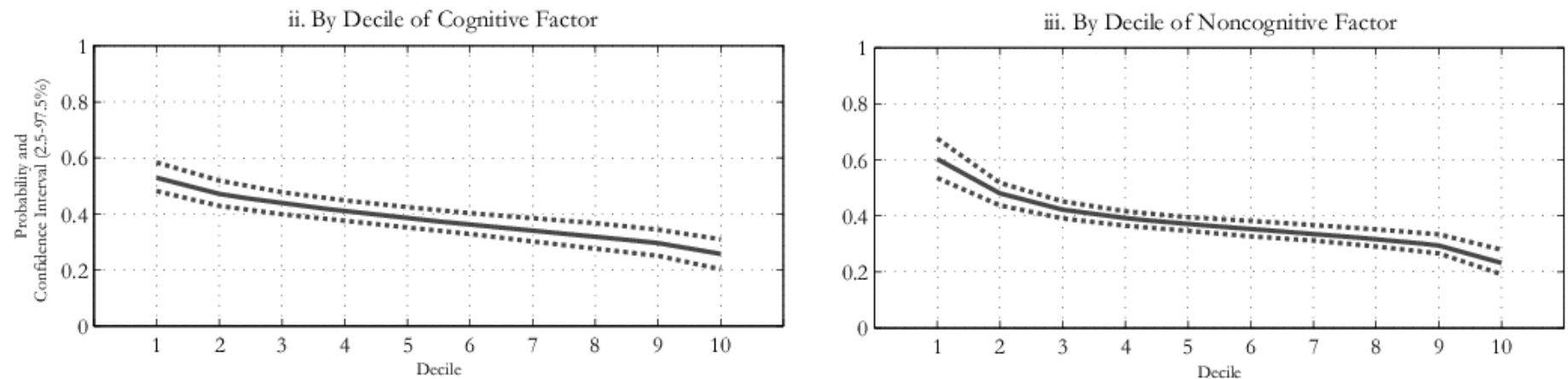


Figure 21: Probability of daily smoking by age 18 (males)

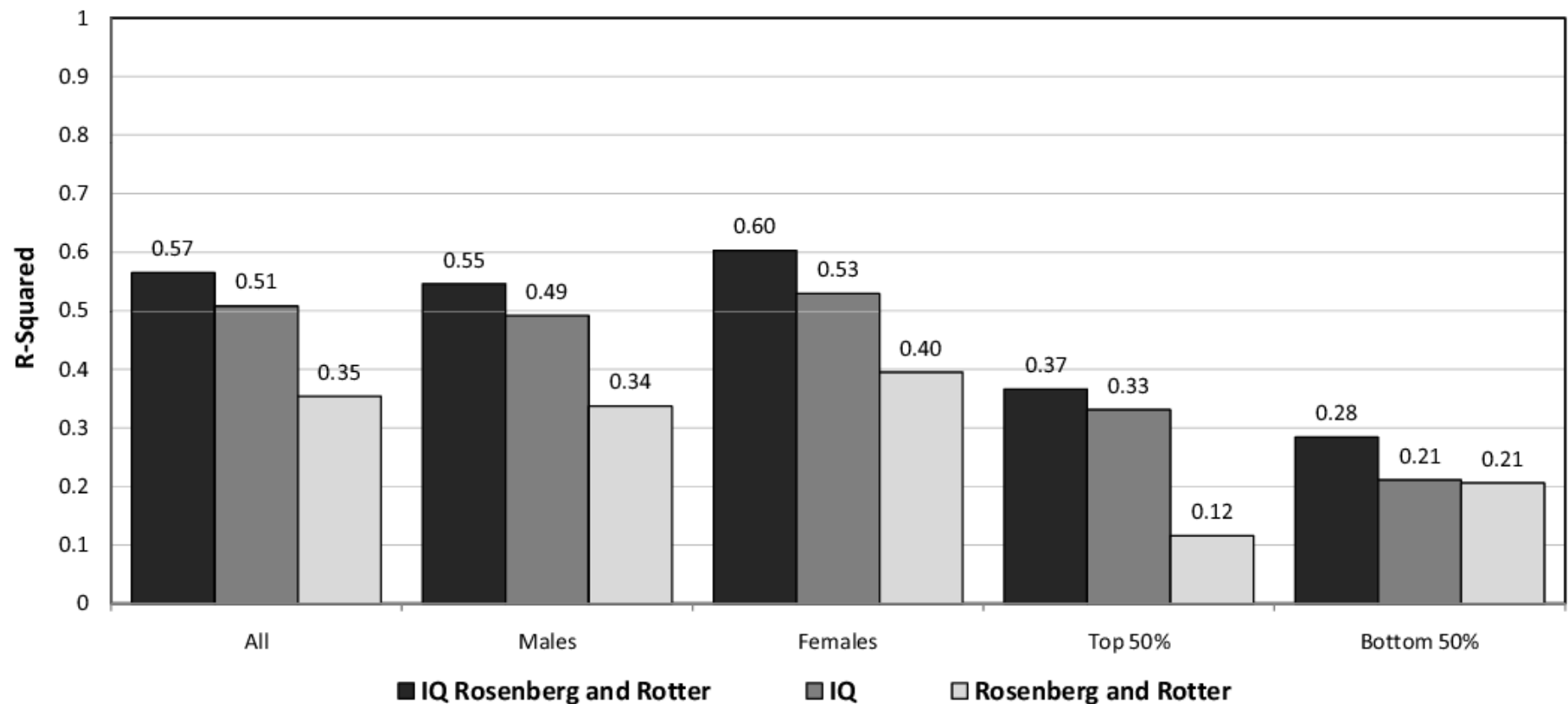


Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws).

Confusion As To What Achievement Tests Actually Measure

- Achievement test scores themselves are explained in part by personality tests.
- Confusion of IQ and achievement (e.g., *The Bell Curve*) is common.
- Achievement tests bundle IQ and personality (recall Table 2).
- The power of “IQ,” as claimed by Herrnstein and Murray, is in part the power of personality.

Figure 22: AFQT Decomposed by IQ, Rosenberg, and Rotter (Unconditional)



Source: Borghans, Golsteyn, Heckman et al. [2010]

Test Performance Can Be Incentivized

Table 6: Incentives and Performance on Intelligence Tests

Study	Sample and Study Design	Experimental Group	Effect size of incentive (in standard deviations)	Summary
Edlund [1972]	Between subjects study. 11 matched pairs of low SES children; children were about one standard deviation below average in IQ at baseline	M&M candies given for each right answer	Experimental group scored <u>12 points</u> higher than control group during a second testing on an alternative form of the Stanford Binet (about 0.8 standard deviations)	"...a carefully chosen consequence, candy, given contingent on each occurrence of correct responses to an IQ test, can result in a significantly higher IQ score."(p. 319)

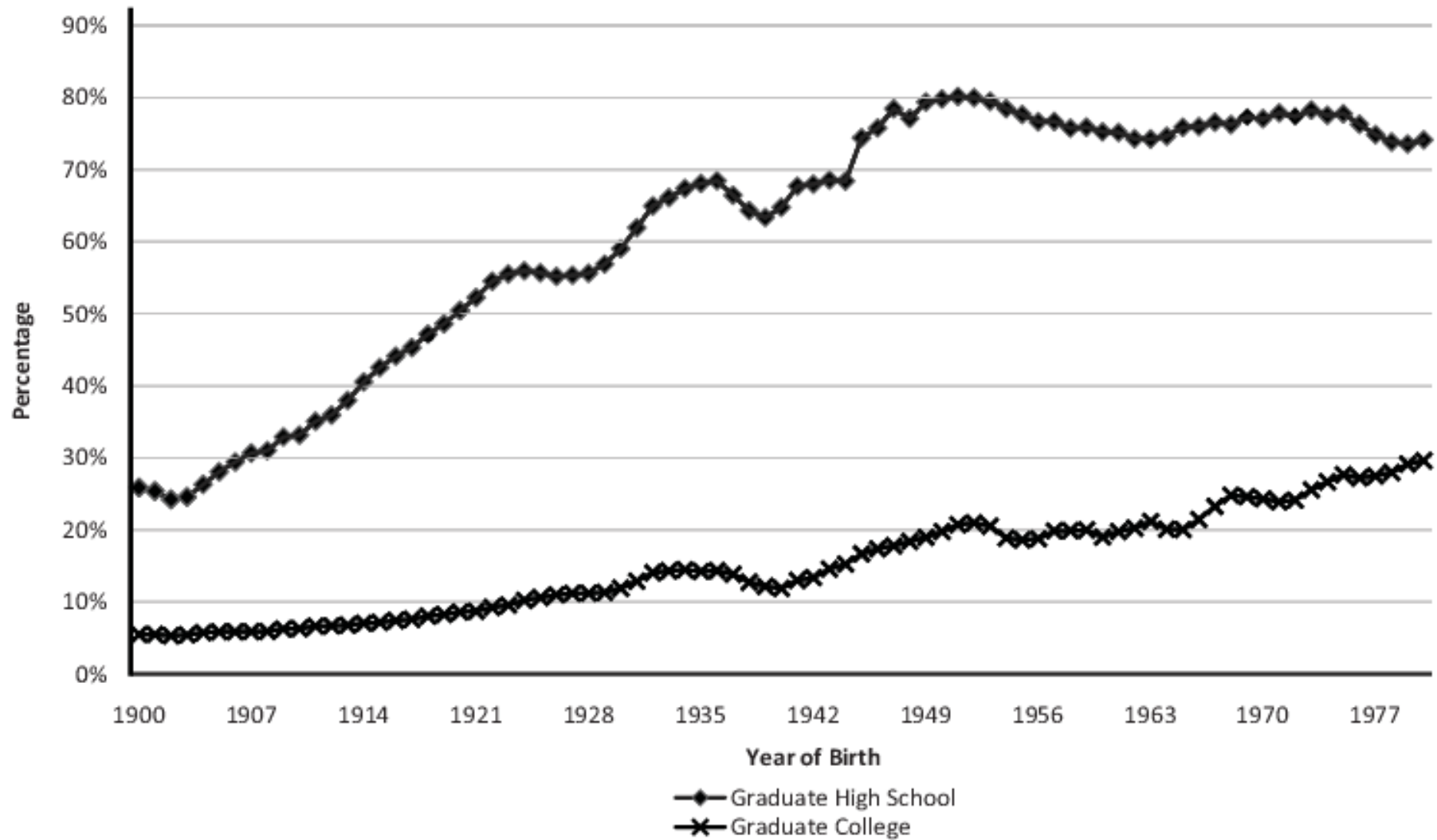
- Incentives operate more effectively on those with lower levels of motivation.
- No lasting effects of incentives (Pay for grades).

III. Costs of Neglecting Soft Skills

- GED conceals and creates major problems in American society.
- The one-dimensional focus of public policy on “smarts” conceals major problems by distorting social statistics on the health of society and by misdirecting effort by institutions and individuals.
- GEDs earn at the rate of high school dropouts and resemble dropouts in many other ways, even though they are as smart as high school graduates as measured by achievement tests.

- By counting GEDs as high school graduates, Americans deceive themselves about the health of their society.
- If GEDs are properly counted as high school dropouts, the U.S. high school dropout rate has *increased* since the early 1970s.
- The rising dropout rate is a worry because the market value of education has risen in the past thirty years.

Figure 23: Trends in Educational Attainment



- The rising high school dropout rate helps to explain the recent slowdown in the growth of skills in the American economy and the rise in inequality.
- At a time when skilled labor has become more valuable and when a high skilled work force is needed to compete in the world economy and to meet fiscal challenges, America's rate of producing high skills has decelerated.

- The U.S. has fallen from first in the world in percentage of population graduating college 30 years ago to twelfth today.
- The GED credential conceals this trend because many nominal high school completers are actually GEDs.
- The conventional measures of educational attainment *overstate* the true growth of skill in the American economy.

Tests Conceal Problems of Racial Disparity

- The GED program makes it easy to hide serious social problems.
- The official statistics suggest that the black-white gap in high school attainment has been substantially reduced.
- However, racial convergence in educational attainment is an illusion manufactured in part by the GED program.
- For males, the apparent educational advance of black males is largely due to their higher incarceration rates as well as to the production of GED credentials in prison.

- Prisoners are not counted in many official statistics on educational attainment.
- Since most prisoners are high school dropouts, their removal from social statistics inflates the measured high school graduation rate for black males.
- In addition, for many black males, prison is their high school, and many obtain GED credentials while incarcerated.
- Count these GEDs as dropouts, as the data on the socioeconomic performance of the GEDs says we should, there is *no* progress in black high school attainment rates over the past 40 years.

Making Test Scores a Goal Diverts Attention from Problems and Give a False Sense of Success

- The persistence of a large-scale GED testing program, despite all of the evidence that it is not working, exposes deep cynicism in many quarters of American public life.
- The quick fix is the rule of the day despite a growing awareness that we need to address long-run structural deficits.
- If a problem is papered over, it is out of sight, and hence out of mind.

Not All GEDs Are Alike: The Program Is Not Universally Bad

- It benefits some: who are they?
- Those high in both cognitive and noncognitive skills.
- For white males the respective probabilities of being a terminal dropout and of obtaining a GED based on population deciles of cognitive and noncognitive ability.
- “1” represents the lowest decile and “10” represents the highest.
- Dropouts low in both cognitive and non-cognitive skill.
- GEDs high in cognitive and low in noncognitive skill.

Figure 24: Distribution of Probability of Dropping Out by Cognitive and Noncognitive Ability (white males)

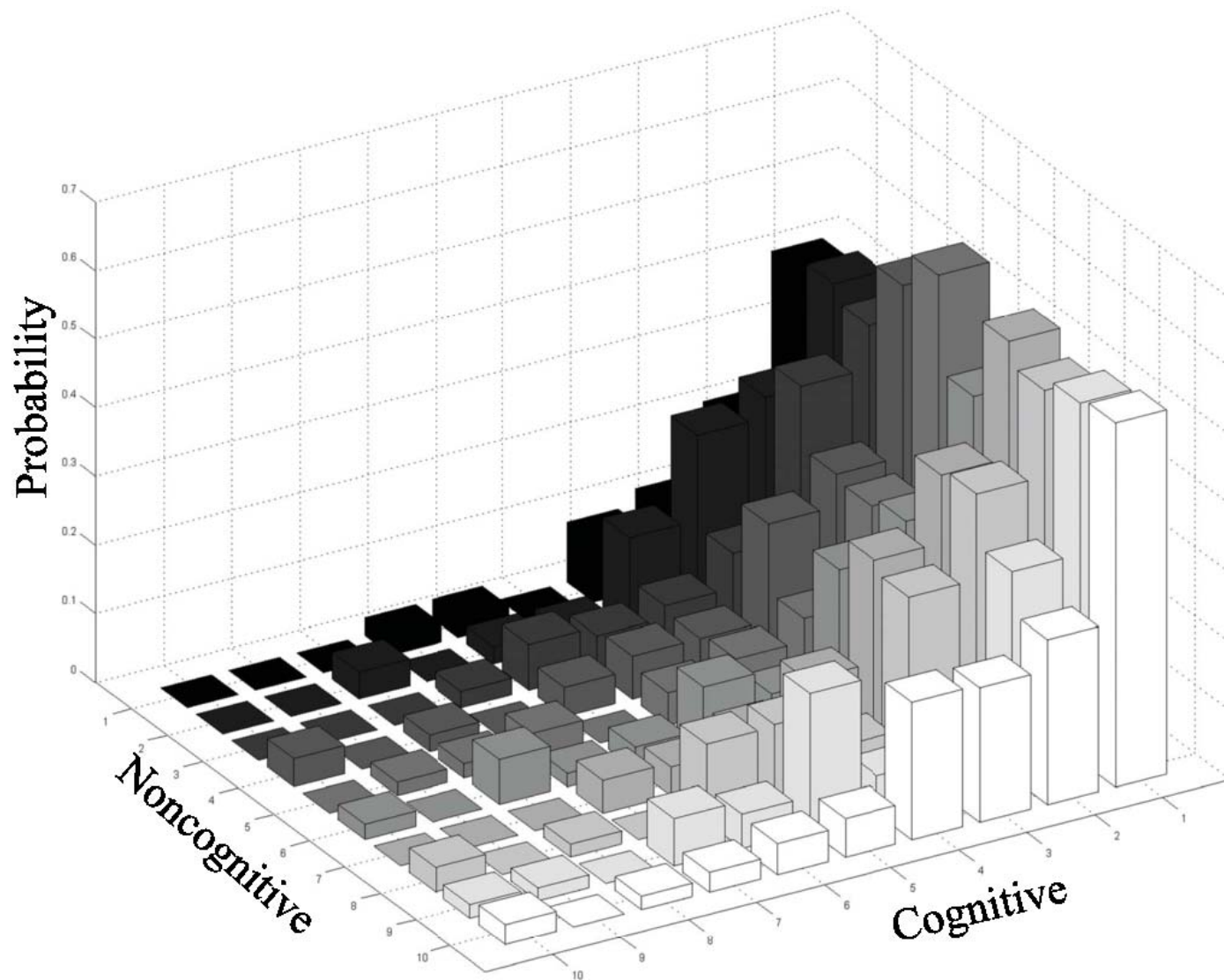
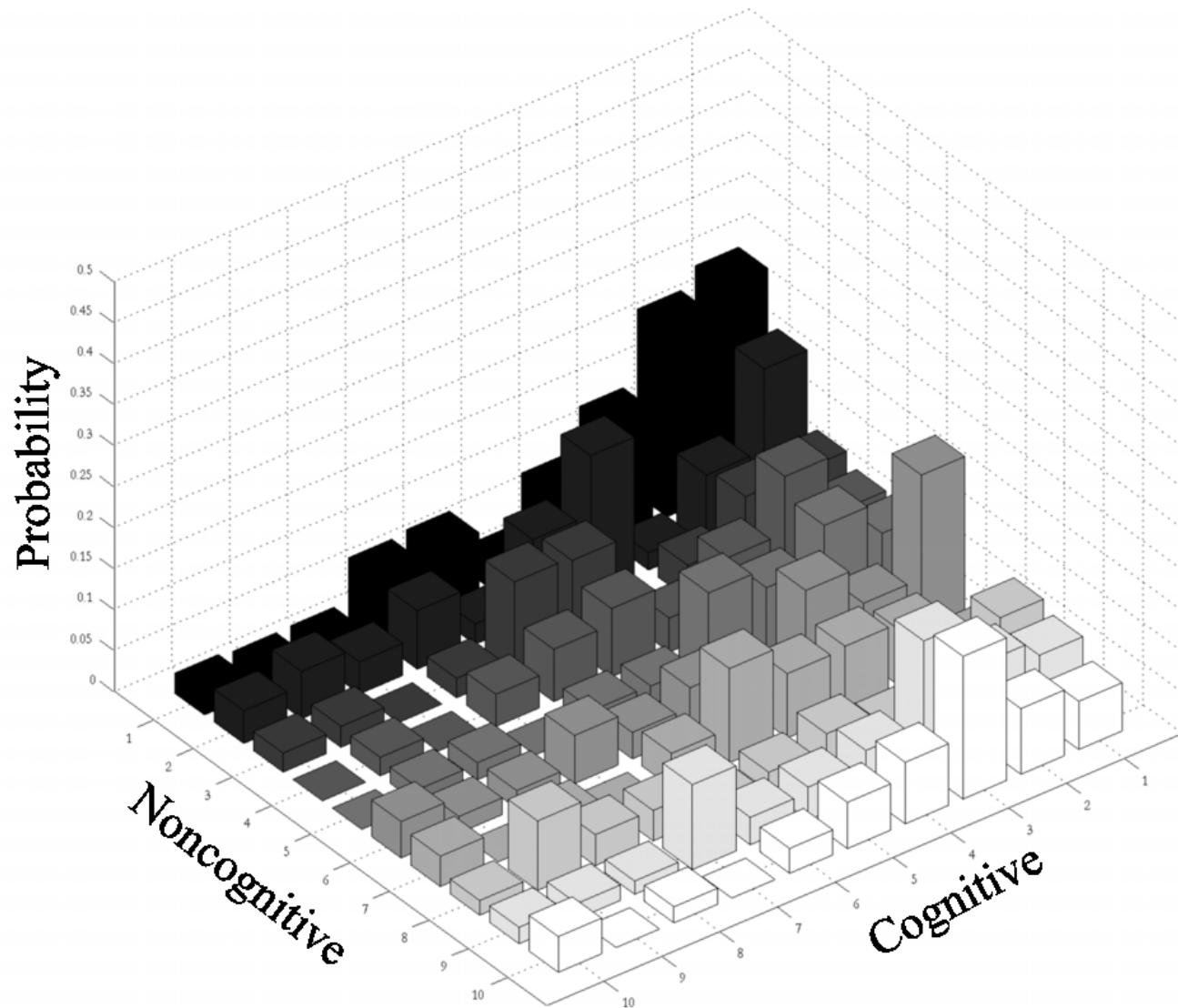


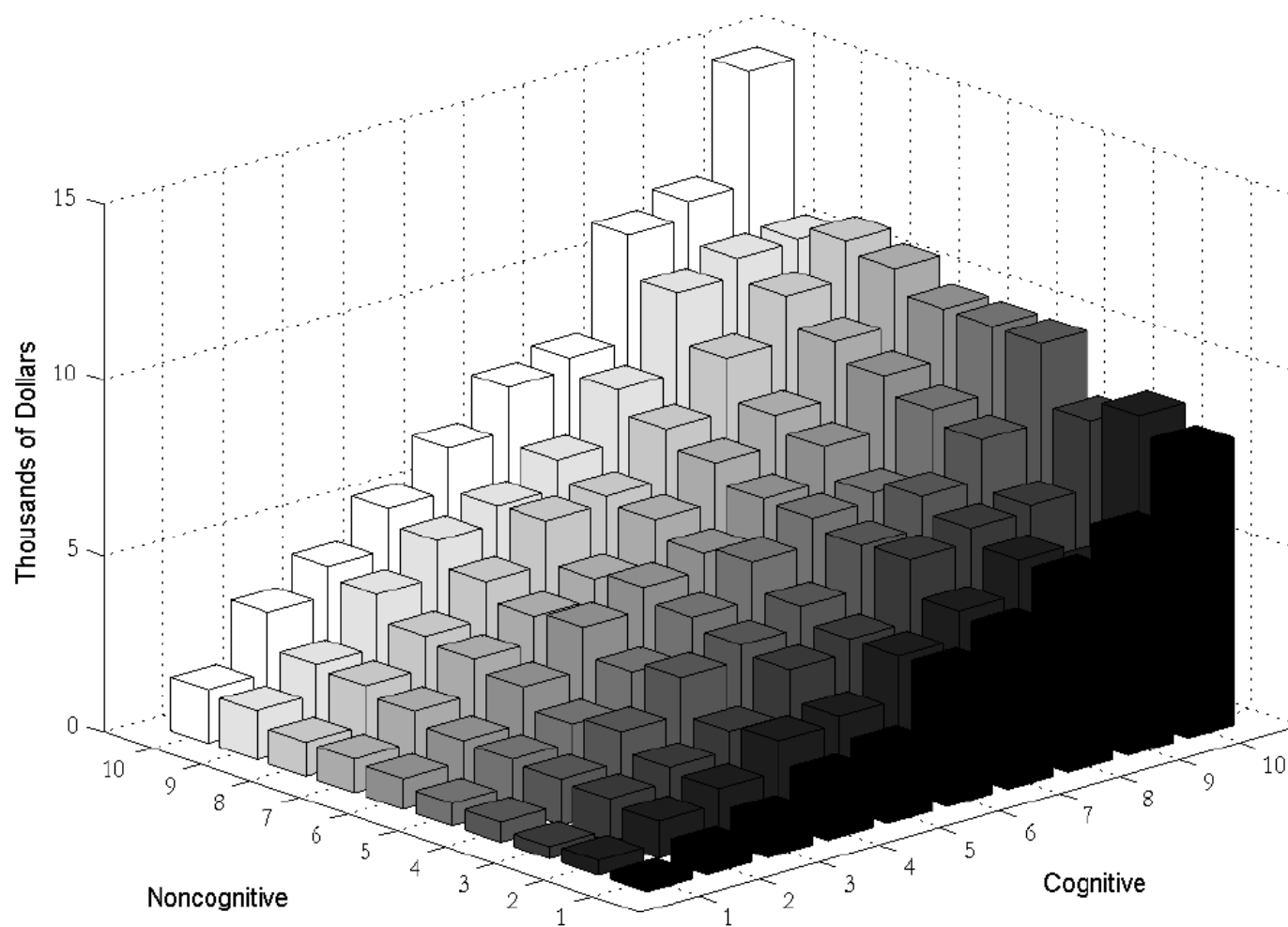
Figure 25: Distribution of Probability of GED Receipt by Cognitive and Noncognitive Ability (white males)



The Option Value of the GED

- The GED creates options for high school dropouts.
- It opens the doors to higher education for them.
- 60% go on to college.
- Yet only 4% graduate a 4-year college.
- Who benefits?
- Those with high levels of cognitive and noncognitive skills.
- The estimated option values of the GED as a function of cognitive and noncognitive ability.

Figure 26: Distribution of GED Option Values by Cognitive and Noncognitive Ability Deciles for white males



- This is a new approach to policy evaluation.
- Instead of saying “good” or “bad,” saying which programs work for which people.
- Targeted programs

But on Net Is There a Benefit or a Loss from the GED Program?

- Consider the growth in GED certification and what fuels it.
- Lessons from the U.S. apply more generally.

Changes and Growth in the GED Test Taking Population

- Government Education and Training Programs introduced in the War on Poverty in 1960s.
- Accountability Movement spawned by McNamara created a demand for numbers (body count, etc.), “performance driven” policies.
- GEDs were perceived to be true equivalents of ordinary high school graduates.

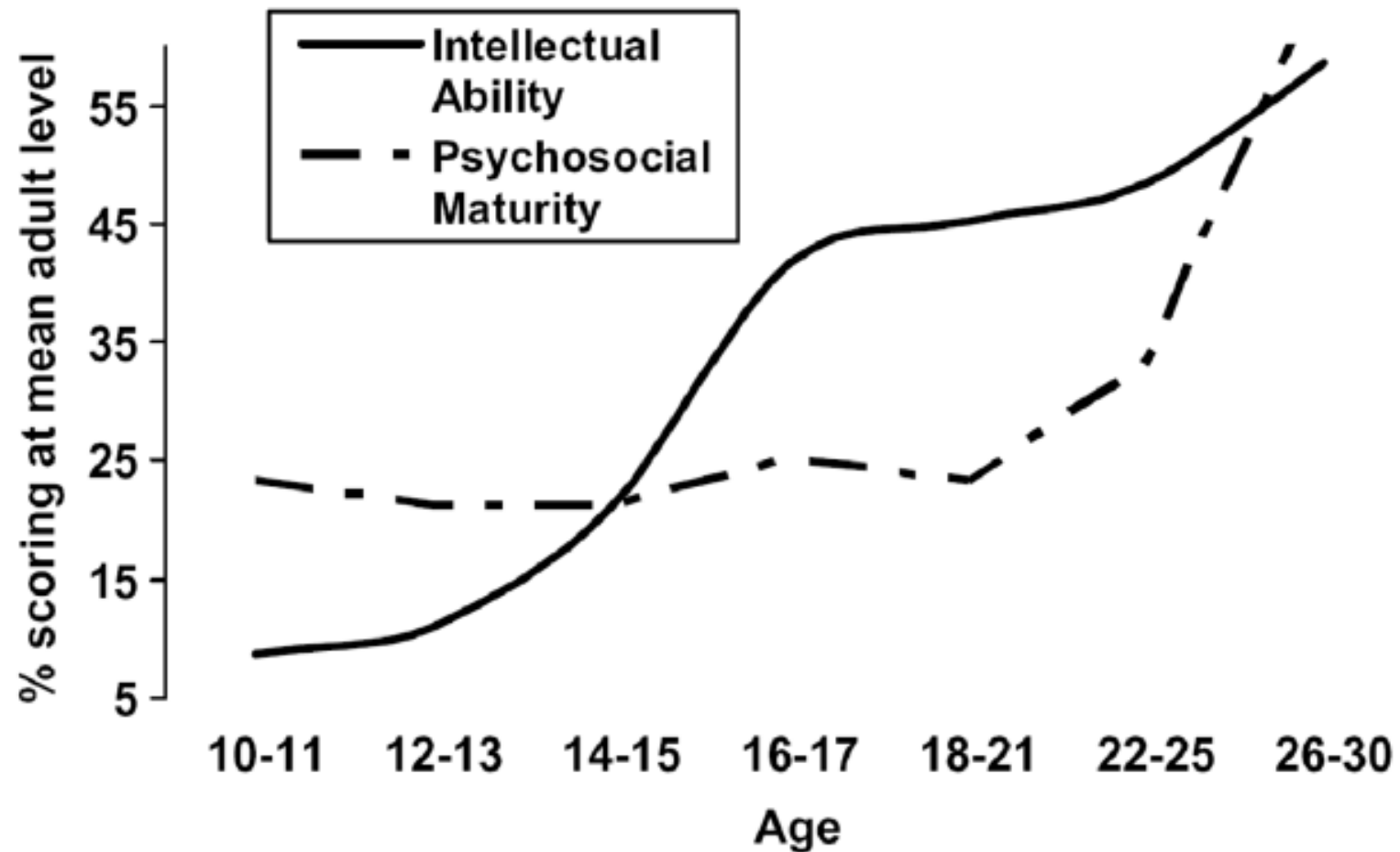
Growth in High School-Age Test Takers

- In 1955, as more states began offering the GED to civilians, the American Council of Education implemented a minimum age of 20 for taking the GED test to prevent teen-aged students from seeking the GED as a replacement for high school.
- Schools are under pressure to comply with new accountability laws (NCLB) have strong incentives to dump dropouts into GED programs.
- High stakes testing drives people into taking GEDs.

Vulnerable Youth

- Youth are very vulnerable, often make bad choices.

Figure 27: Proportion of individuals in each age group scoring at or above the mean for 26- to 30-year-olds on indices of intellectual and psychosocial maturity.



Source: Steinberg, Graham, O'Brien et al. [2009].

- The decision-making process of teens may lead them to make choices that restrict their educational paths and earnings in a way that they later regret.

Adverse Social Consequences of the GED

- The GED Induces Would-Be High School Graduates to Drop Out

Table 7: Simulated Response of Educational Attainment to Elimination of the GED

Schooling Level	Simulated with GED (1)	No GED (2)	Change in Rate (2)-(1)	% Change ((2)/(1)-1)%
Four Year College	25.5%	26.0%	0.5%	2.1%
Some Four Year College	7.0%	7.1%	0.1%	1.3%
Two Year College	7.2%	7.8%	0.6%	8.0%
Some Two Year College	10.2%	10.7%	0.6%	5.5%
Some College GED	2.5%	-		
High School Graduates	31.9%	34.0%	2.1%	6.5%
GEDs	3.7%	-		
High School Dropouts	12.0%	14.3%	2.4%	19.6%

Note: The numbers in columns (1) and (2) are computed as fractions of the overall population

Source: Heckman and Urzua (2010).

- Similar Results from the Introduction of GED Into California in 1970s

The GED Obscures the True Returns to Education

Table 8: The Role of the GED in Explaining Rising Educational Wage Gaps

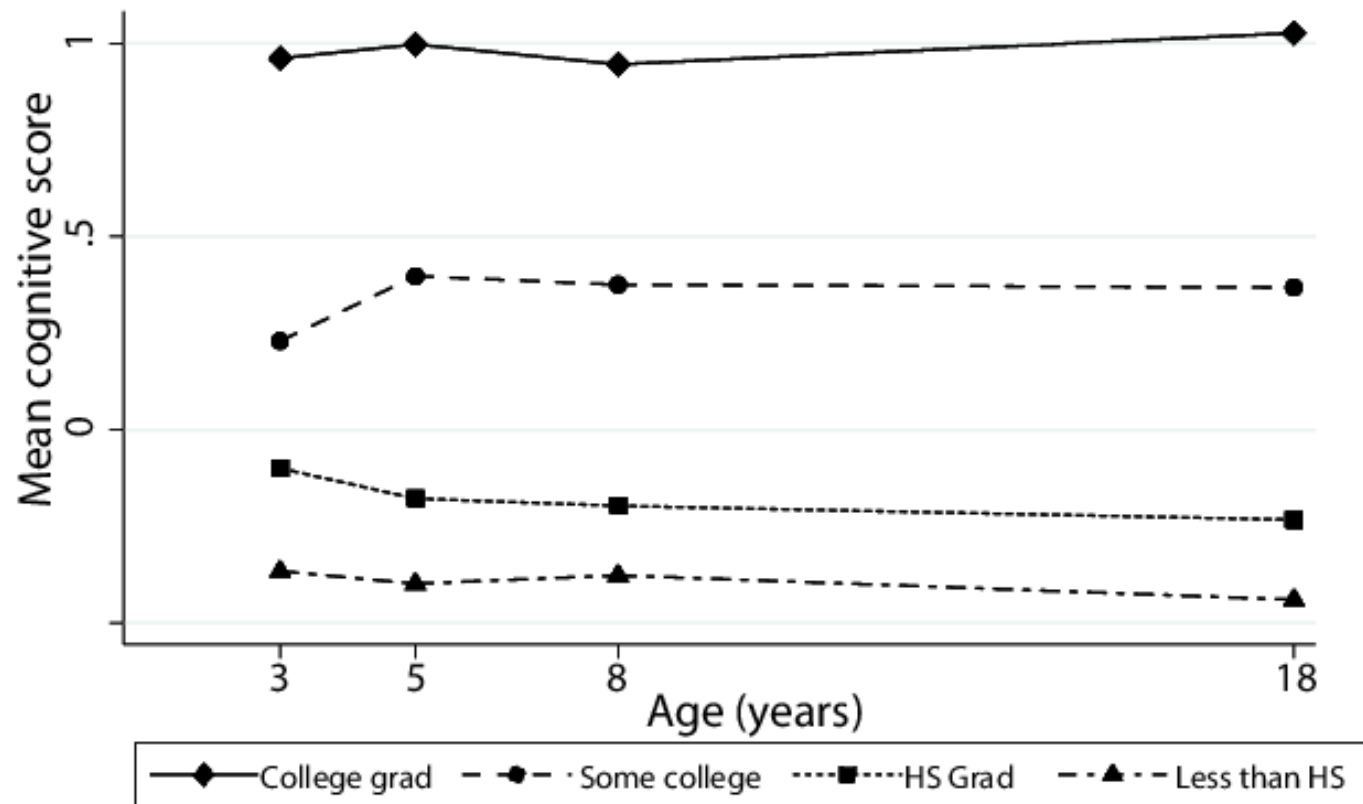
Percentage of Overall Change Explained by GED Misclassification from 1957 to 1964

	Annual Earnings	Weekly Wage	Hourly Wage
Growth in College-HS Gap	18.4%	12.8%	14.4%
Growth in College-Dropout Gap	23.7%	13.6%	5.5%

IV. Cognitive and Social and Emotional Skills Can Be Fostered

- Gaps in cognitive and noncognitive skills open up early across social and economic groups.
- For both cognitive and socioemotional traits, ability gaps across socioeconomic groups open up at early ages and persist.

Figure 28: Trend in mean cognitive score by maternal education



Each score standardized within observed sample. Using all observations and assuming data missing at random. Source: Brooks-Gunn et al. (2006).

- Family investment and early childhood programs promote both cognitive and noncognitive skills.
- IQ becomes rank stable by the early teenage years.
- Achievement (crystallized intelligence) can be fostered.
- Personality skills are more malleable.
- Controlling for early family environments using conventional statistical methods substantially narrows the gaps.
- There are comparable phenomena in the evolution of gaps in behavioral problems.
- Schools foster these skills.

Figure 29: Causal Effect of Schooling on Measures of Cognition (from ASVAB)

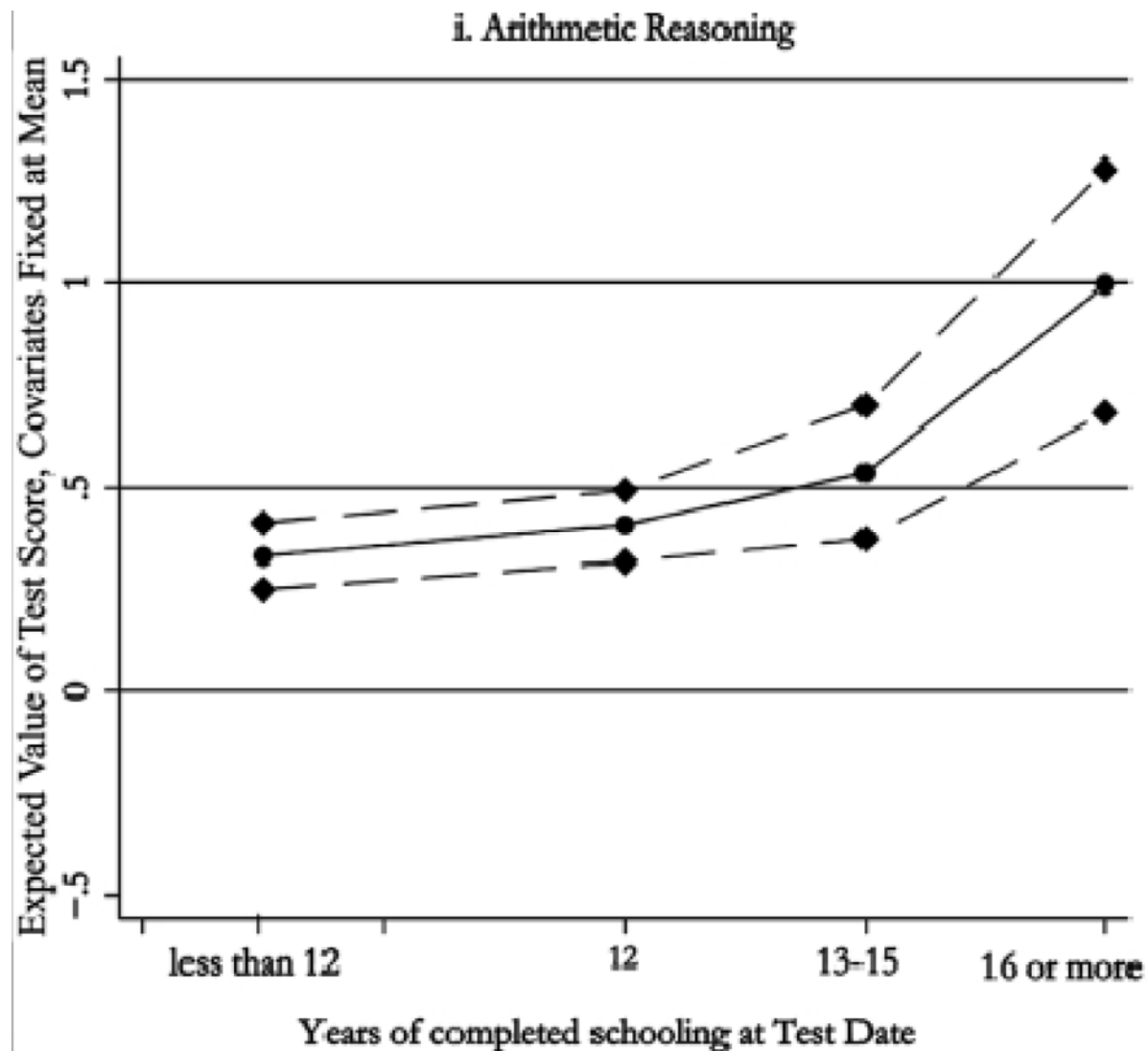


Figure 30: Causal Effect of Schooling on Measures of Cognition (from ASVAB)

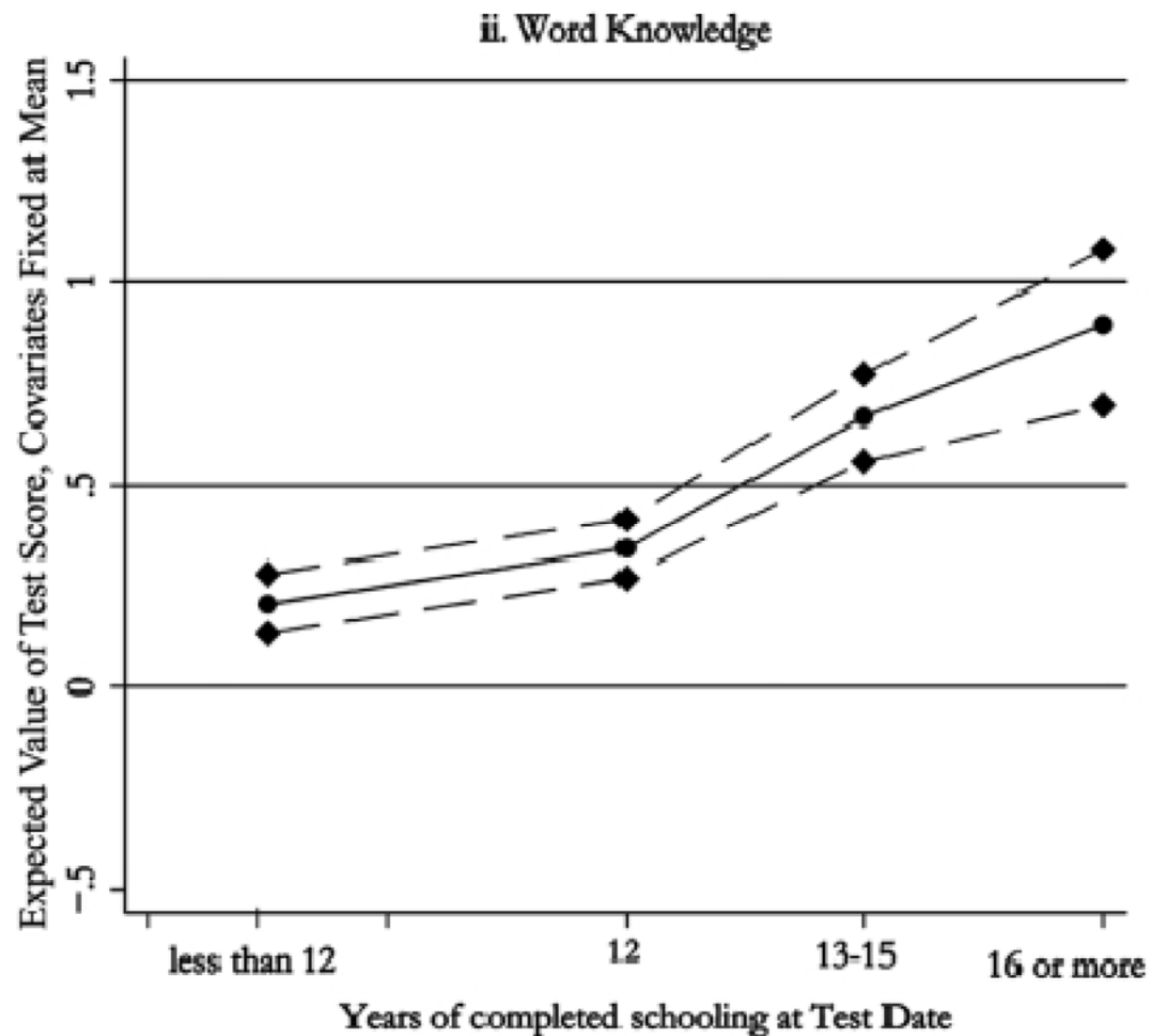


Figure 31: Causal Effect of Schooling on Two Measures of Socioemotional Skills

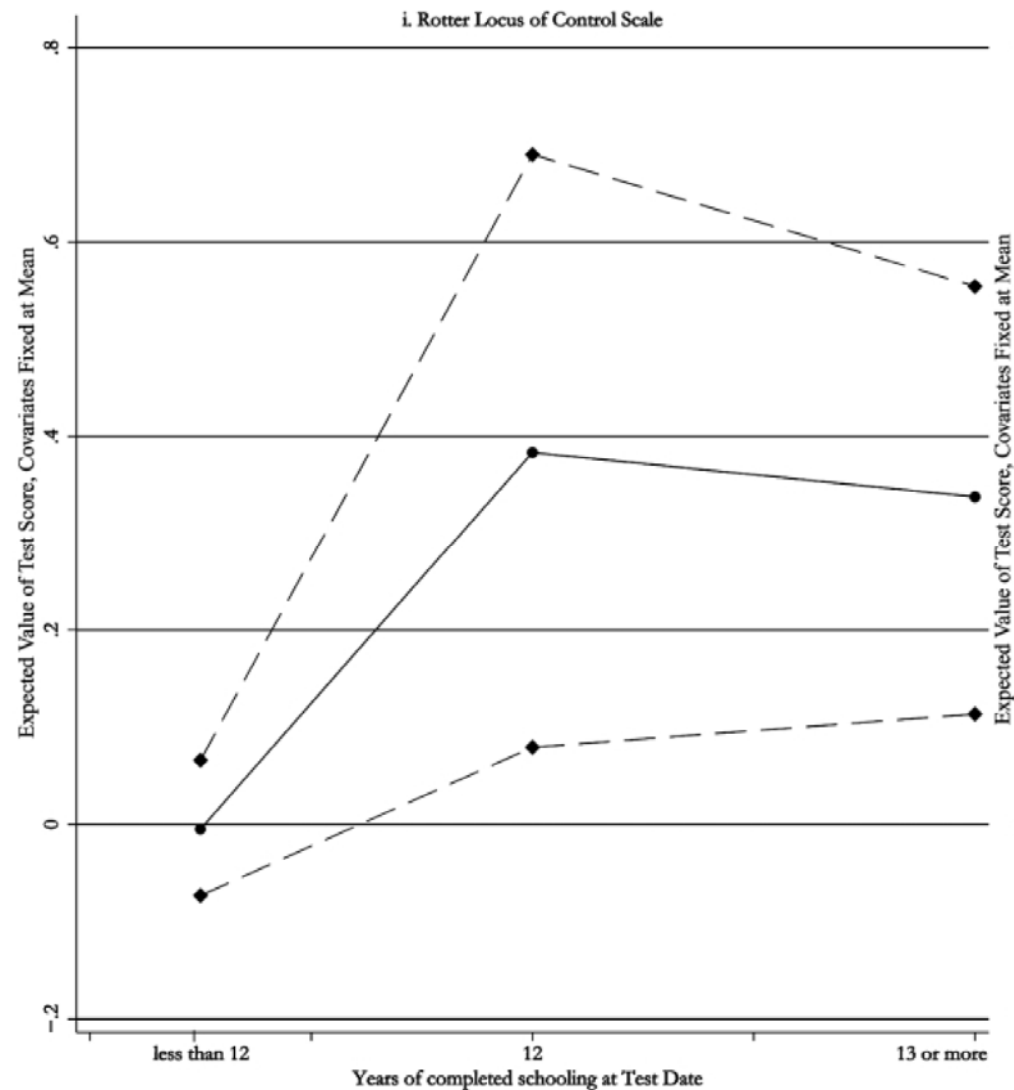
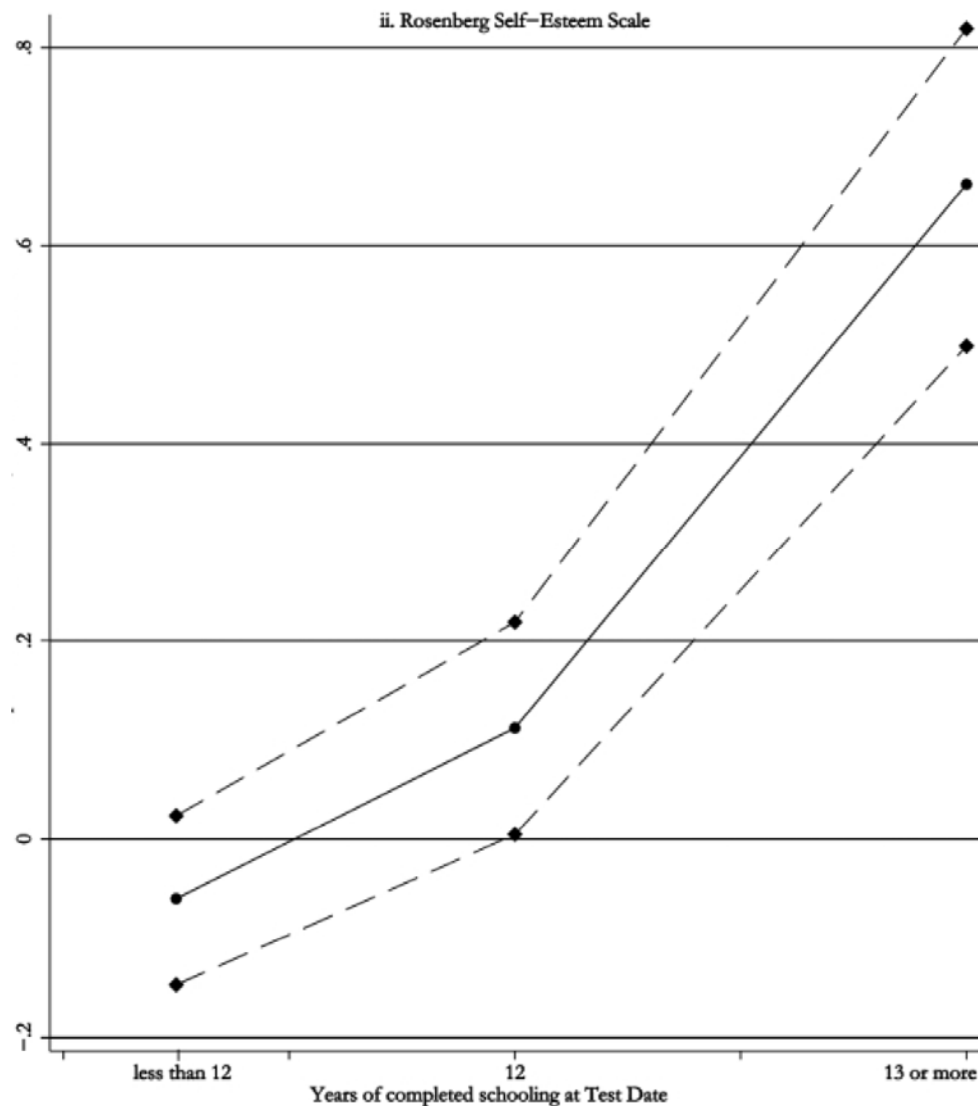


Figure 32: Causal Effect of Schooling on Two Measures of Socioemotional Skills



Enriched Early Environments Foster Social and Emotional Skills

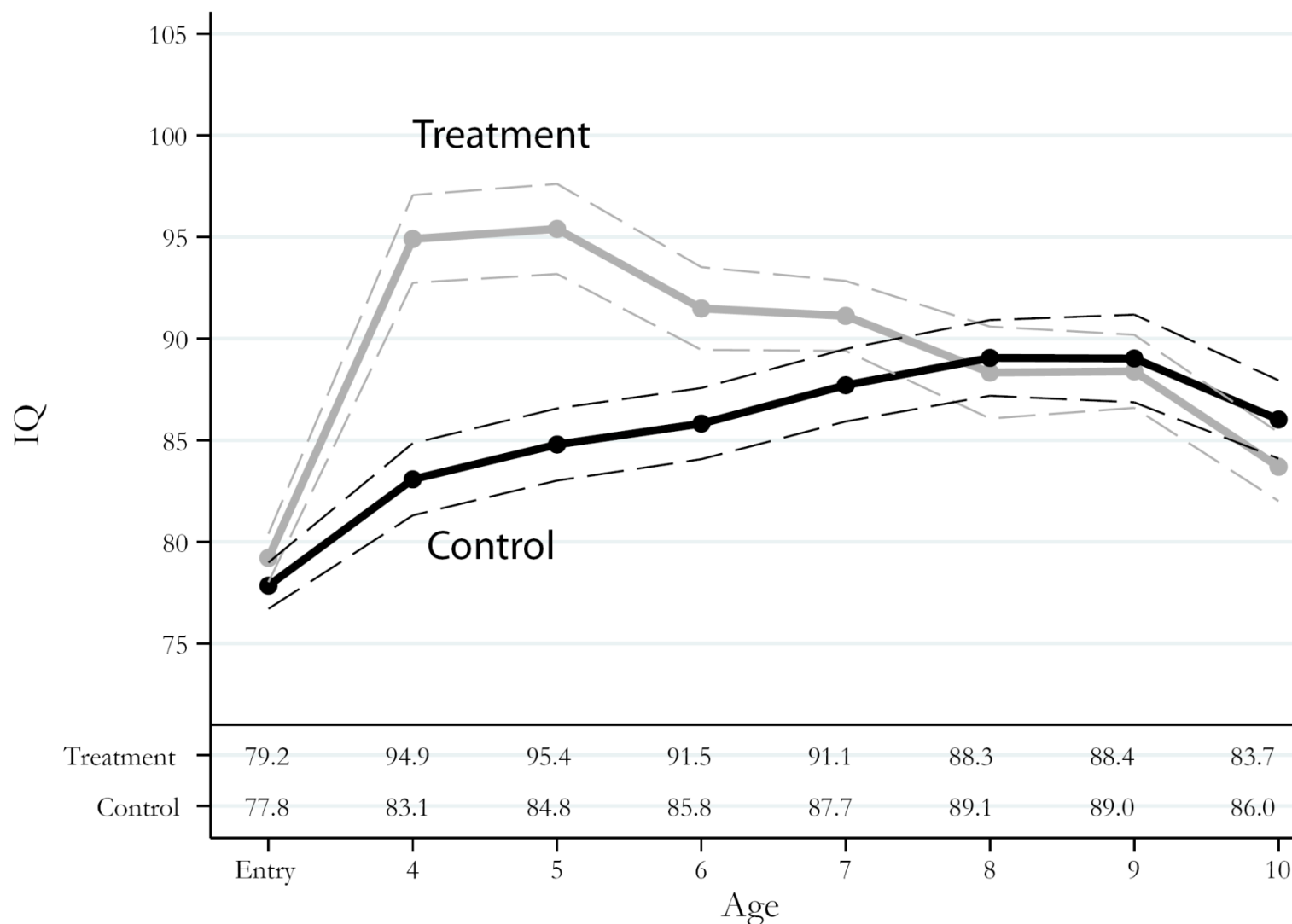
High/Scope Perry Preschool Program

- The Perry preschool program enriched the lives of low income black children with initial IQs below 85 at age 3.
 - 2.5 hours per day
 - 5 days per week
 - 2 years during each school year (mid-October to May).
 - home visits
 - program stops after two years
 - the program taught planning and persistence as well as social adjustment
 - “Plan, Do, Review”: Plan a project, do it, review it collectively

- Evaluated by the method of random assignment.
- Strong effects are found for both boys and girls, although different effects are found at different ages for different outcomes.
- Did not lead to sustained gains in IQ for males, and only slight effect for females.

Figure 33: Cognitive Evolution Through Time, Perry Males: Male Cognitive Dynamics

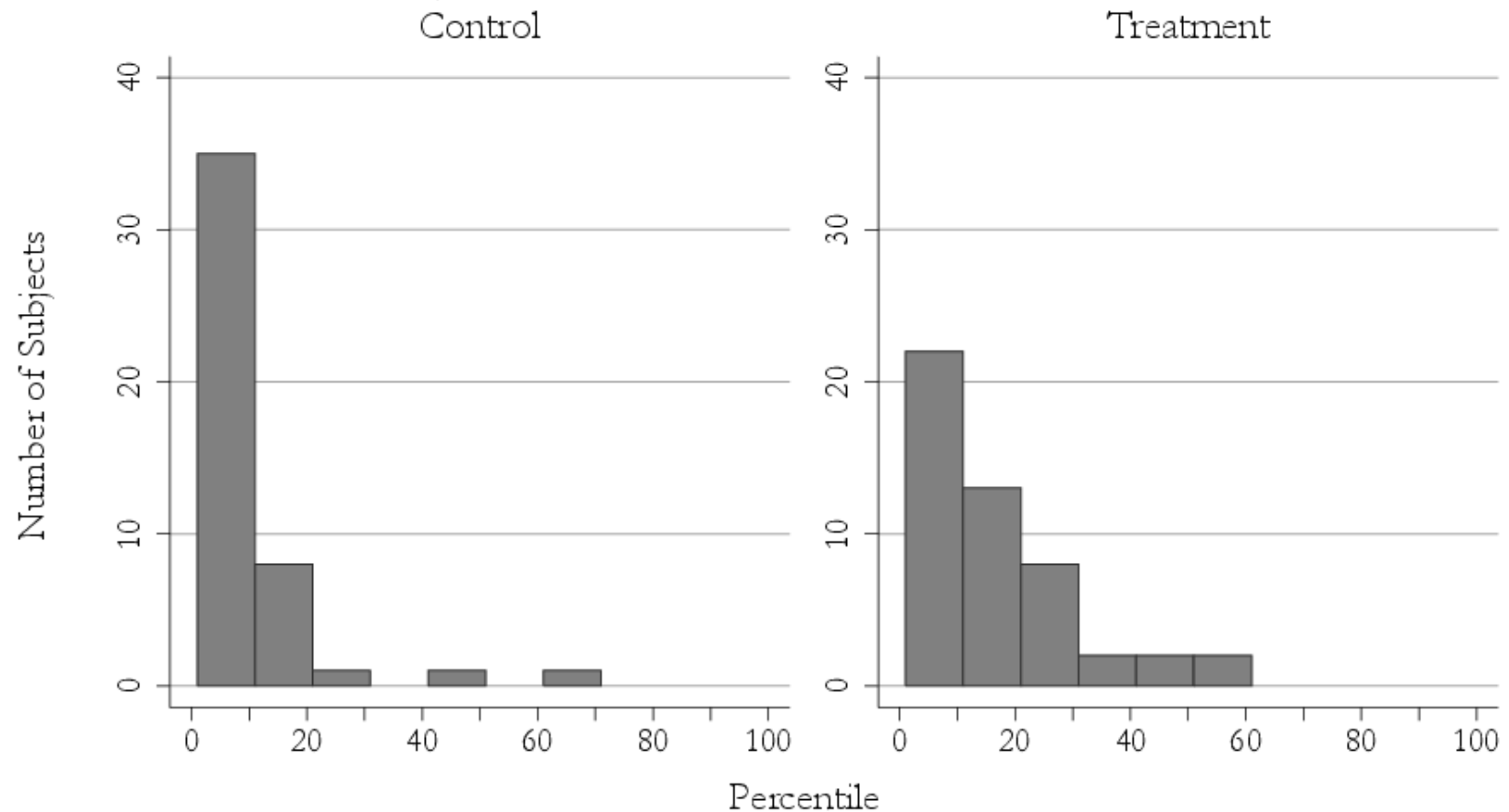
Male Cognitive Dynamics



- Yet the Perry Program has a statistically significant annual rate of return of around 6-10% per annum—for both boys and girls—above the post World War II stock market returns to equity in U.S. labor market estimated to be 5.8%.

- The Perry Preschool Program worked primarily through socioemotional channels.
- Raises scores on achievement tests but not IQ tests.
- Socioemotional factors and cognitive factors both explain performance on achievement tests (Duckworth, 2006; Borghans et al., 2008; Borghans et al., 2009).
- Personality factors substantially affect performance on achievement tests.

Figure 34: Perry Age 14 Total CAT Scores, by Treatment Group



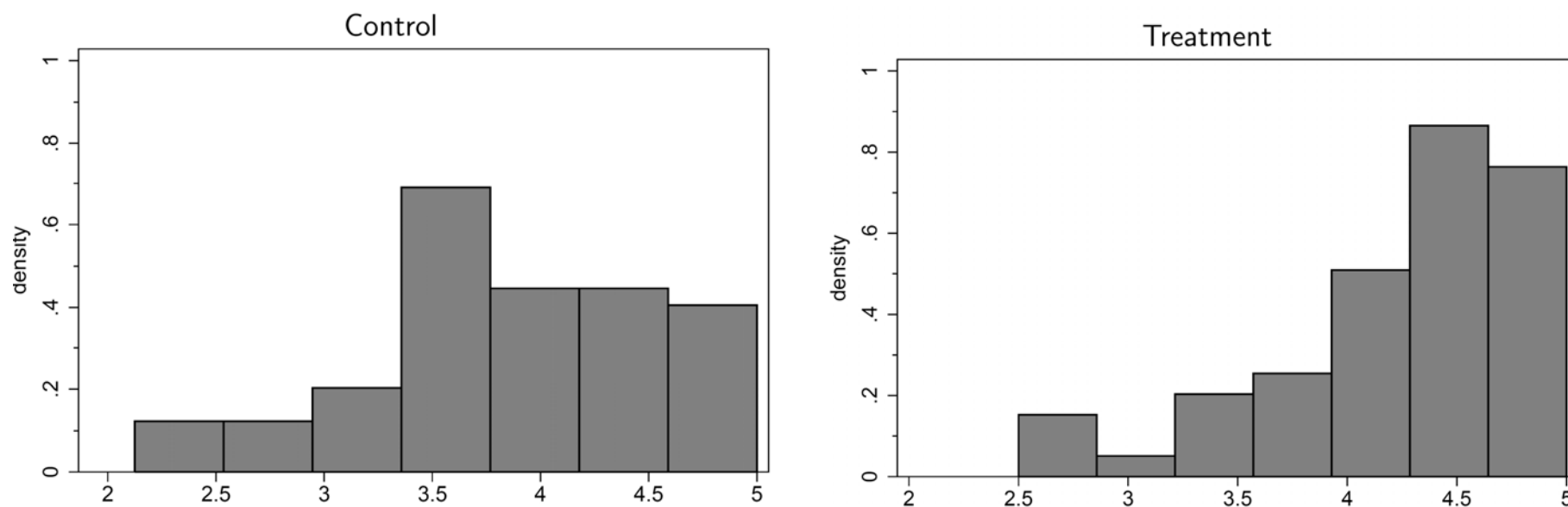
CAT = California Achievement Test

Treatment: $N = 49$; Control: $N = 46$

Statistically Significant Effect for Males and Females (p-values 0.009, 0.021 respectively)

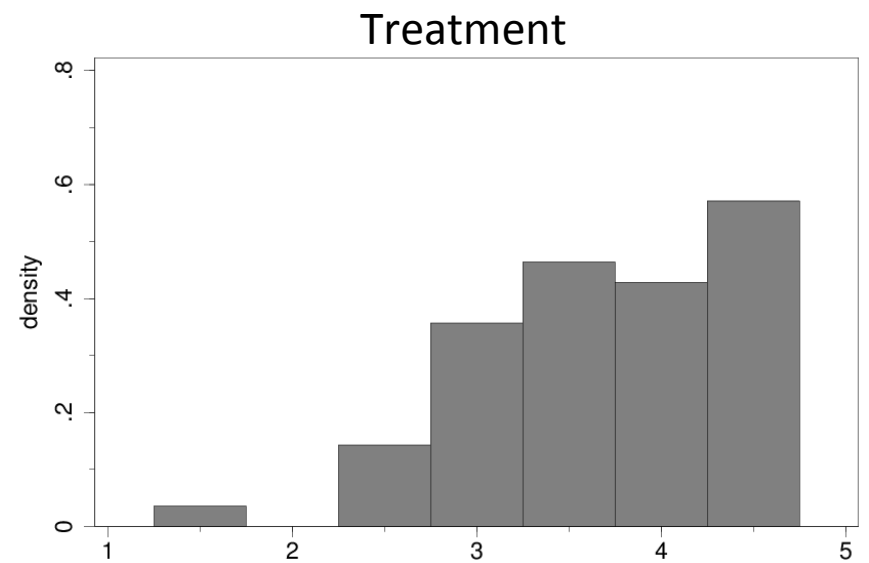
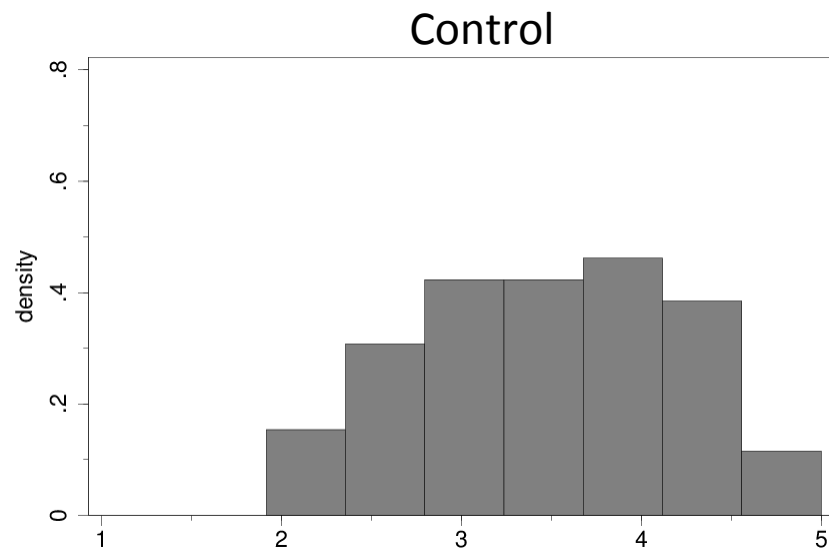
Source: Heckman, Malofeeva, Pinto, and Savelyev (2008).

Figure 35: Personal Behavior Index, by Treatment Group



- Treatment shifts the distribution upwards (1=bad;...;5=good).
- Statistically significant treatment effect is observed:
 $p = 0.002$.
- The Personal Behavior Index is an unweighted average of four items: “absences and truancies,” “lying or cheating,” “steals” and “swears or uses obscene words.”

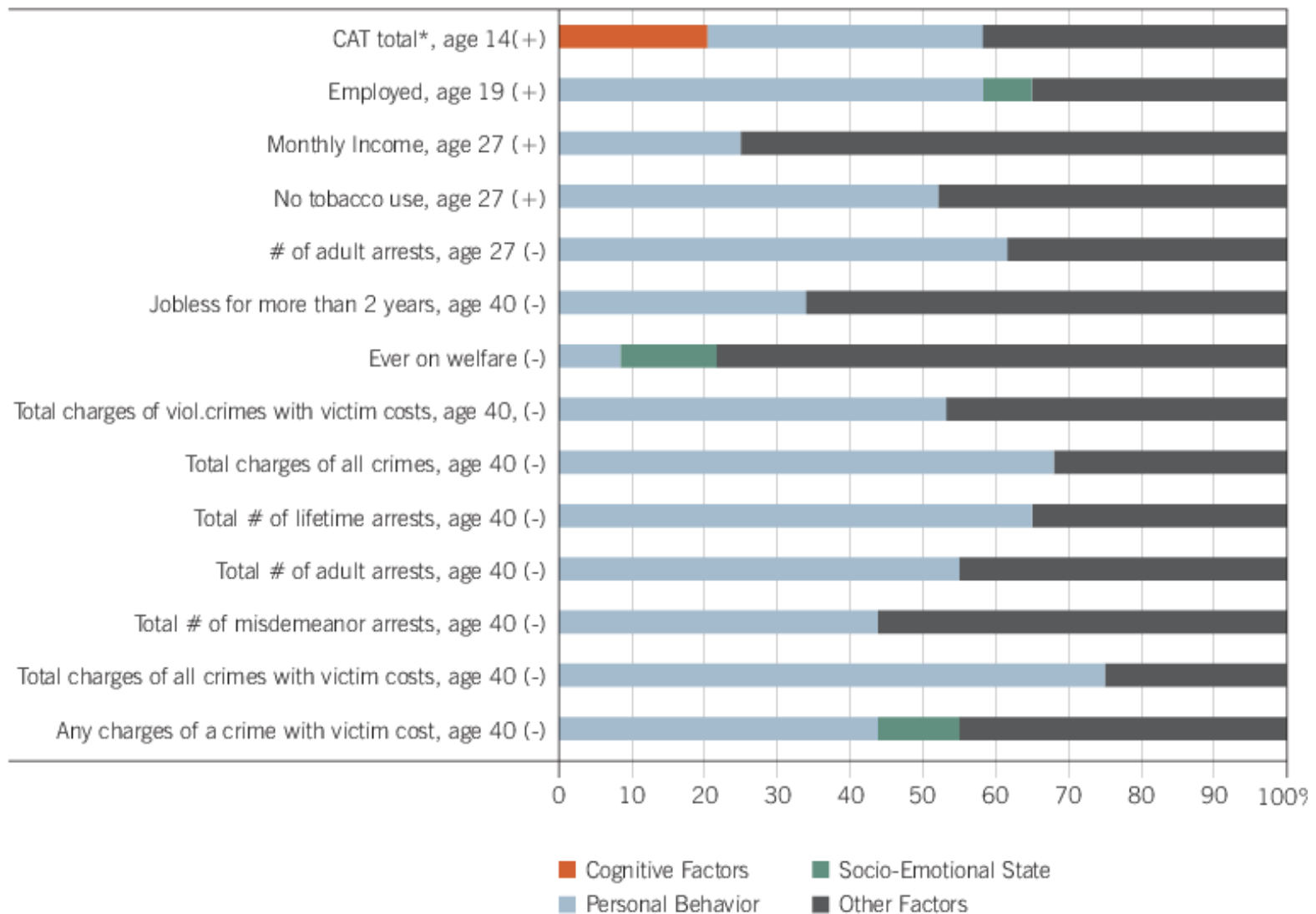
Figure 36: Socio-Emotional Index by Treatment Group



- Treatment shifts the mean upwards (1=bad;...;5=good).
- Treatment effect one-sided p-values is 0.096 (borderline statistically significant).
- The Socio-Emotional index is an unweighted average of four items: “appears depressed,” “withdrawn and uncommunicative,” “friendly and well-received by pupils,” and “appears generally happy.”

Decomposing Treatment Effects

Figure 37: Decomposition of Treatment Effects, Males



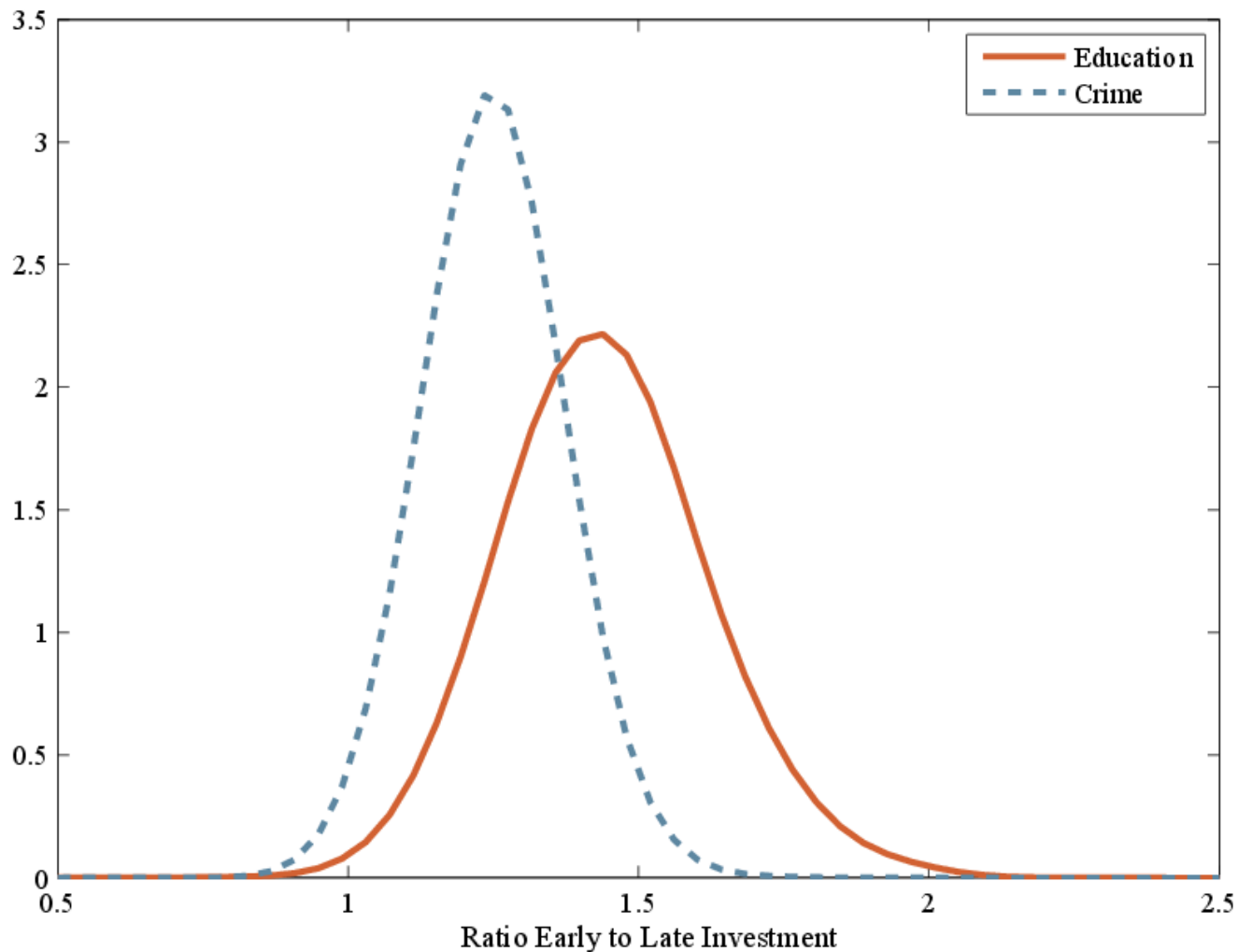
Evidence from a Substantial Body of Work on Parental Investment: The Following Lessons Emerge

- Investment most productive in early years for IQ and fluid intelligence.
- Investment relatively more productive in middle years of childhood for fostering personality.
- Associated with the slowly developing pre-frontal cortex.
- Criminal activity depends relatively more on personality: self-control, anger management, etc.

Evidence from a Substantial Body of Work on Parental Investment: The Following Lessons Emerge

- Educational attainment depends relatively more on cognitive skills.
- Personality fosters cognitive skills.
- Leads to following optimal policy (early vs. late) (Cunha, Heckman, and Schennach, 2010).
- It is a policy tailored to individual patterns of disadvantage.
- Not a “one size fits all” type of policy.

Figure 38: Densities of Ratio of Early to Late Investments:
Maximizing Aggregate Education versus Minimizing
Aggregate Crime



Conclusions

- Focus on cognitive test scores deceives us.
- Soft skills matter.
- They can be shaped, even into the adolescent and young adult years.
- Improving them is a productive avenue of social policy.
- GED is a good illustration of the importance of soft skills.
- GED program tests a part of the skills that matter.

Conclusions

- The GED program conceals and creates major social problems.
- Distortions created by No Child Left Behind.
- Schools now focus only on math and reading.
- Policies focusing on promoting test scores, tuition and college application policies, and the like miss a basic point about what matters for success and how to foster it.
- A policy of prevention rather than remediation is more cost effective and is a more equitable way to promote efficiency.

Appendix

Table A1: Correlations, Partial Correlations, and Explained Variance of IQ and Personality with Later-life Outcomes

Variables Included	Correlations and Partial Correlations			Explained Variance (R^2)		
	IQ	Locus of Control/ Self-Esteem	All	IQ	Locus of Control/ Self-Esteem	All
High School Diploma	0.037***	0.049*** 0.044***	0.017 0.045*** 0.039***	0.001	0.005	0.005
Highest Grade Completed	0.390***	0.013 0.199***	0.358*** -0.049*** 0.118***	0.151	0.042	0.165
12 th Grade GPA	0.486***	0.018 0.180***	0.464*** -0.065*** 0.075***	0.236	0.035	0.242
Hourly Wage at 35	0.198***	0.063*** 0.058***	0.179*** 0.033** 0.013	0.039	0.001	0.041
Weeks Unemployed at 35	-0.137***	0.023* -0.043***	-0.137*** 0.046*** -0.009	0.019	0.002	0.021
Any Welfare at 35	-0.235***	-0.038*** -0.103***	-0.211*** -0.003 -0.051***	0.055	0.014	0.058
Depression at 40	-0.097***	-0.024* -0.098***	-0.070*** -0.012 -0.079***	0.009	0.011	0.016
Physical Health at 40	-0.040***	-0.143*** 0.024*	-0.019 -0.138*** 0.027**	0.002	0.020	0.021
Mental Health at 40	0.023*	0.056*** 0.021	0.007 0.054*** 0.019	0.001	0.004	0.004

Figure A1: Juvenile delinquency and The Big Five

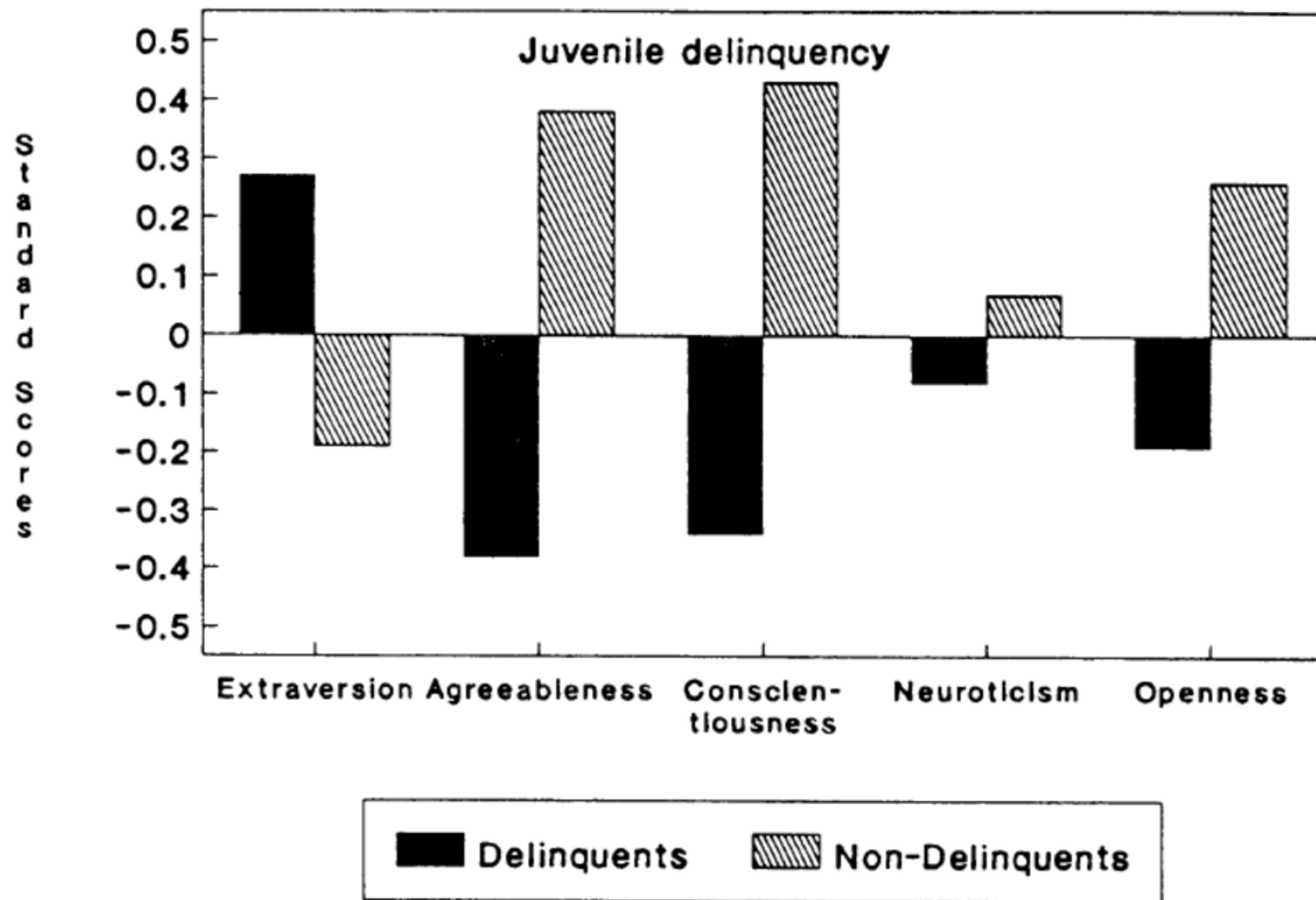


Figure A2: Growth in Incarcerated Populations by Race

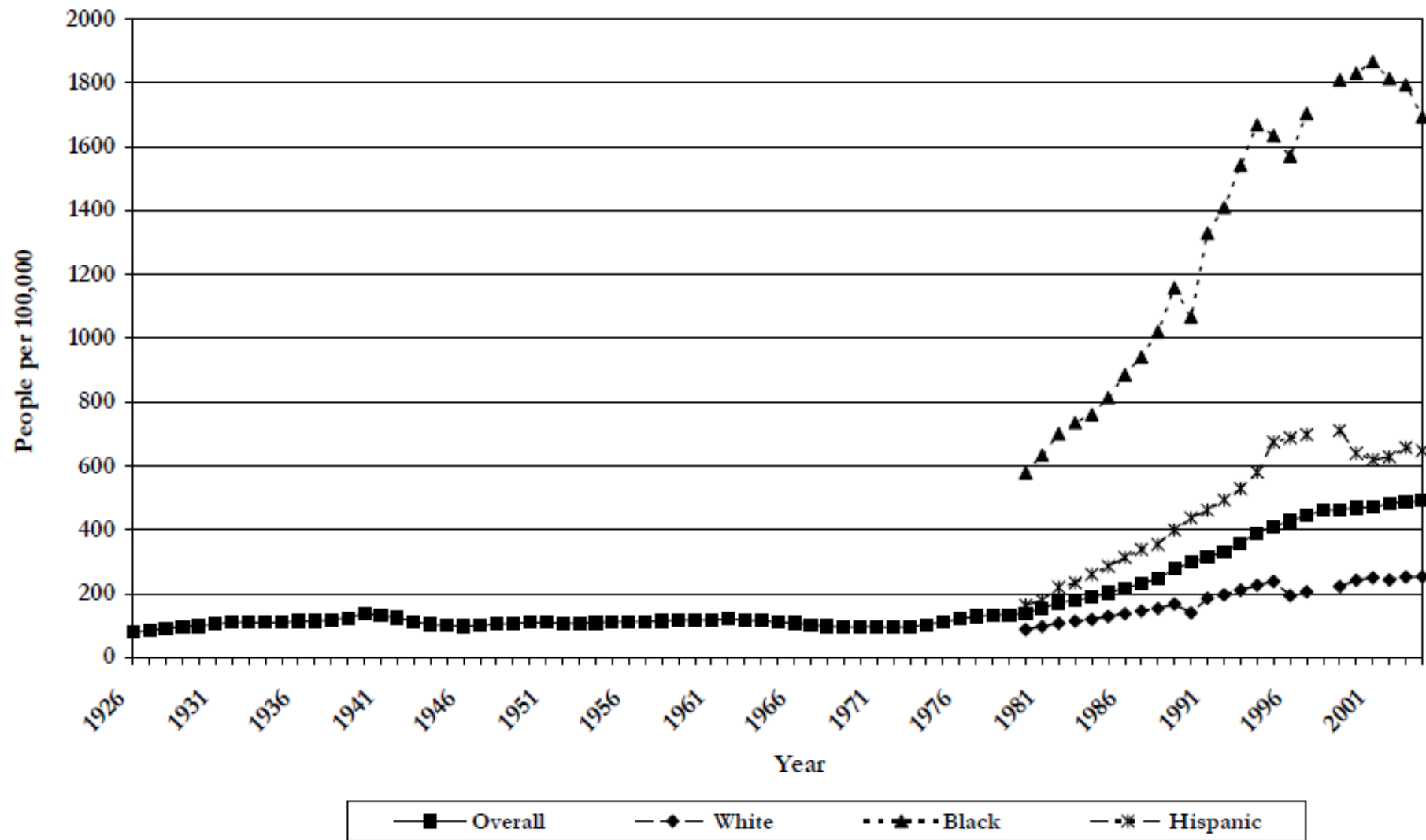
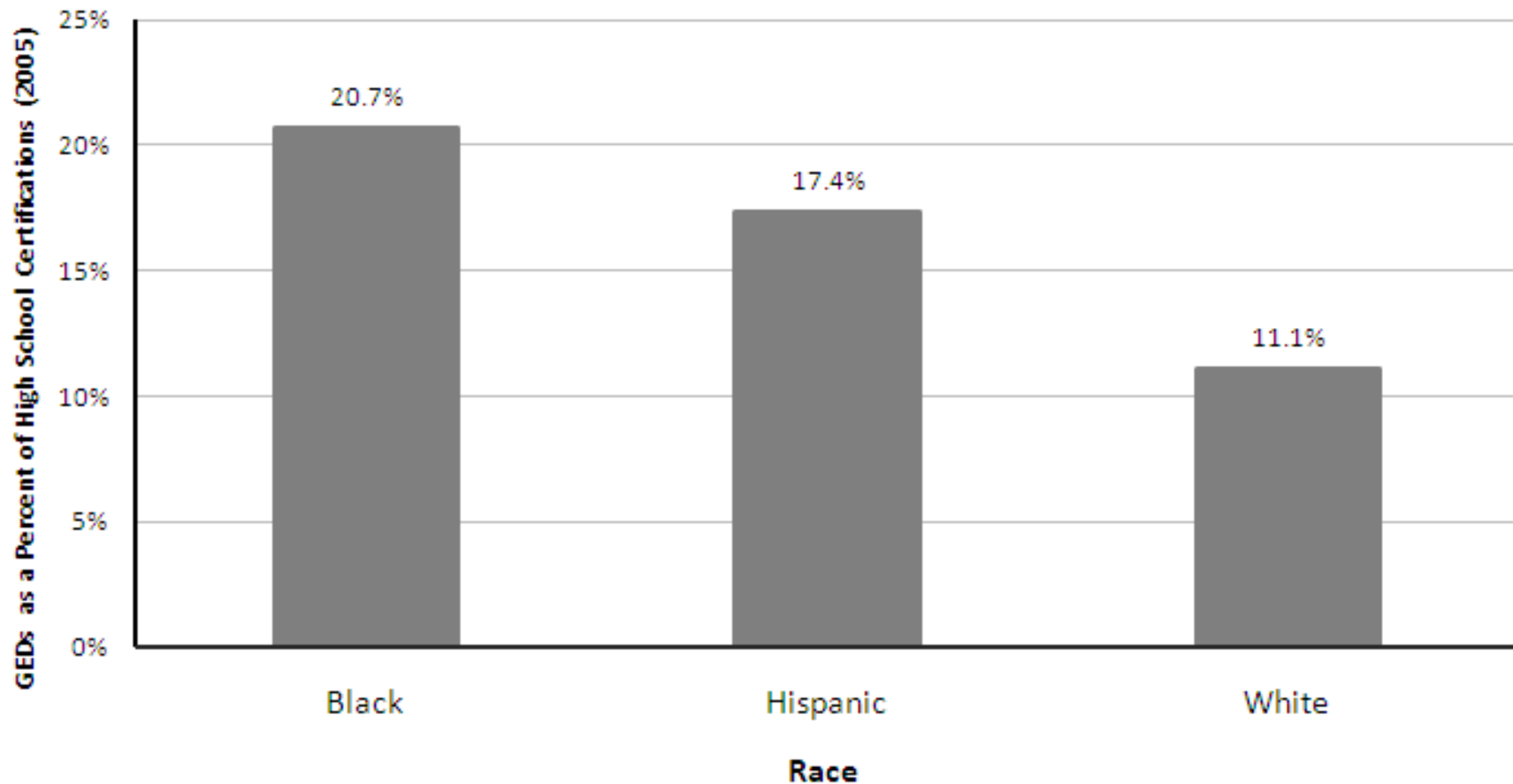


Figure A3: GEDs as a Percent of HS Credentials by Race, 2005



Source: Heckman and LaFontaine (2010).

Table A2: The Effect of Interventions on Personality

Author(s)	Main Variable(s)	Data and Methods	Causal Evidence	Main Result(s)
Barnett, Jung, Yarosz et al. [2008]	<u>Outcome(s)</u> : <i>internalizing and externalizing behavior</i> – teacher-assessed Problem Behaviors Scale of the Social Skills Rating System (SSRS) <u>Intervention</u> : participation in a year-long Tools of the Mind preschool program compared to a generic program	<u>Data</u> : collected by authors; 210 children aged 3 and 4 <u>Methods</u> : random assignment	<u>Controls</u> : n/a <u>Timing of Measurements</u> : Children were first assessed in the fall before the program and then again in the spring.	Participants in the program had a 0.47 standard deviation lower score for the behavioral problems index ($p < 0.05$).
Behncke [2009]	<u>Outcome(s)</u> : <i>cognitive ability</i> – performance on a diagnostic math test for a college economics class <u>Intervention</u> : verbal encouragement before the test	<u>Data</u> : Collected by author; 440 students from a Swiss University <u>Methods</u> : random assignment, randomization inference	<u>Controls</u> : n/a <u>Timing of Measurements</u> : The noncognitive skill shock directly proceeded test.	Verbal encouragement raised test scores by 2.5% amongst all students ($p < 0.05$) and by 8.0% amongst students who reported difficulties with math ($p < 0.01$).

Bierman, Coie, Dodge et al. [2010]	<u>Outcome(s):</u> <i>teacher-assessed behavior</i> –Social Health Profile (SHP) including authority acceptance, cognitive concentration, and social competence; <i>peer-assessed behavior</i> – survey questions about behavior labeled as aggressive, prosocial, and hyperactive <u>Intervention:</u> – participation in three-year-long Fast Track PATHS program focused on improving self-control and positive social behavior	<u>Data:</u> 2,937 children (grades 1-3) <u>Methods:</u> three-level ordered logistic regression with clustering at the school and individual level, random assignment	<u>Controls:</u> time, time squared, individual baseline, school baseline, city fixed effects, poverty level, interactions of intervention with time, time squared, individual baseline, poverty, and poverty and time <u>Timing of Measurements:</u> The baseline outcomes were measured in kindergarten.	Participation in the intervention was associated with a 0.24 standard deviation increase in authority acceptance ($p<0.001$), a 0.12 standard deviation increase in cognitive concentration ($p<0.001$), and a 0.34 standard deviation increase in social competence ($p<0.0001$) compared to the control group. The effects were stronger in more disadvantaged schools. They find similar but weaker results for the peer-assessed measures.
Bloom, Gardenhire-Crooks and Mandsager [2009]	<u>Outcome(s):</u> <i>educational attainment</i> – high school diploma, <i>labor force participation</i> – whether working at a job, <i>personality</i> – self-efficacy and social adjustment <u>Intervention:</u> participation in the ChalleNGe program consisting of a 2-week assessment period, 20-week residential program often conducted at a military base, and a one-year mentoring program.	<u>Data:</u> 1,018 young people between the ages 16 and 18 who have dropped out of school <u>Methods:</u> random assignment	<u>Controls:</u> sample member characteristics <u>Timing of Measurements:</u> The participants and controls were compared approximately 9 months after entering the study.	Participants in the program were 12.0 percentage points more likely to earn a high school diploma ($p<0.01$), 9.1 percentage points more likely to be working ($p<0.01$), and 9.6 percentage points less likely to report a self-efficacy and social adjustment score one standard deviation below the mean ($p<0.01$). The program also improved measures of criminality and health.

Chetty, Friedman, Hilger et al. [2010]	<u>Outcome(s):</u> <i>non-cognitive skills</i> — an index based on the teacher's observations of the students <u>Intervention:</u> randomly assigned kindergarten class quality as measured by difference in percentiles of the mean end-of-year test scores of the students' classmates and the scores of the other kindergarteners at the same school	<u>Data:</u> Project STAR; 1,671 4 th grade students and 1,780 8 th grade students <u>Methods:</u> OLS, random assignment	<u>Controls:</u> wave fixed effects, student gender, free-lunch status, age, race, a quartic in the claiming parent's household income interacted with parent's marital status, mother's age at child's birth, whether the parents own a home, and whether the parents made a 401 (k) contribution between 1996 and 2008 <u>Timing of Measurements:</u> Classes were randomly assigned in kindergarten, and the behavioral indices were based on 4 th and 8 th grade teacher observations.	A 1 percentile improvement in kindergarten class quality increases an index of non-cognitive skills by 0.15 percentiles in 4 th grade ($p < 0.05$) and 0.13 percentiles in 8 th grade ($p < 0.05$). Better classrooms were also associated with better life outcomes.
Diamond, Barnett, Thomas et al. [2007]	<u>Outcome(s):</u> <i>Executive Function</i> – Dots-Mixed task, Reverse-Flanker task <u>Intervention:</u> participation in a Tools of the Mind program instead of the regular school district's balanced literacy program	<u>Data:</u> 147 preschoolers <u>Methods:</u> random assignment	<u>Controls:</u> age, gender, years in program <u>Timing of Measurements:</u> The tasks were given at the end of the second year of the program.	84% of students in Tools were successful in the Reverse Flanker task compared to 65% in the control group. Almost twice as many students in the Tools program achieved greater than 75% accuracy on the Dots-Mixed task compared to the control group.

Durlak and Weissberg [in press]	<u>Outcome(s)</u> : social and emotional learning skills, attitudes, positive social behavior, conduct problems, emotional distress, academic performance <u>Intervention</u> : Meta-analysis of school-based, universal social and emotional learning program	<u>Data</u> : 270,034 kindergarten through high school students <u>Methods</u> : meta-analysis	<u>Controls</u> : n/a <u>Timing of Measurements</u> : n/a	The mean difference in standard deviations between the treatment and control groups are as follows: social and emotional learning skills = 0.57 ($p < 0.05$); attitudes = 0.23 ($p < 0.05$); positive social behavior = 0.24 ($p < 0.05$); conduct problems = 0.22 ($p < 0.05$); emotional distress = 0.24 ($p < 0.05$); academic performance = 0.27 ($p < 0.05$). All variables are coded so that positive numbers reflect better outcomes.
Gottschalk [2005]	<u>Outcome(s)</u> : <i>Personality</i> – four measures of locus of control based on whether the respondent agrees strongly, agrees, disagrees, or strongly disagrees with various statements <u>Intervention</u> : randomly assigned work subsidies	<u>Data</u> : Self-Sufficiency Project; 4,958 single parents over the age of 19 in New Brunswick and British Columbia <u>Methods</u> : random assignment, probit, IV	<u>Controls</u> : age, age squared, region, gender, speaks French, number of children <u>Timing of Measurements</u> : Participants were interviewed at baseline and 36 months after baseline.	Using whether the participant received the subsidy as an instrument for hours worked, the authors find that working tends to improve locus of control.

Heckman, Malofeeva, Pinto et al. [2010]	<u>Outcome(s)</u> : <i>externalizing behavior, internalizing behavior</i> – measured using Pupil Behavior Inventory (PBI) of teacher reports <u>Intervention</u> : participation in the Perry Preschool Program, an intervention that lasted two years and enriched the lives of low income black children	<u>Data</u> : Perry Preschool Program; 123 preschool students <u>Methods</u> : random assignment	<u>Controls</u> : n/a <u>Timing of Measurements</u> : The measure of externalizing and internalizing behavior are taken at ages 7 to 9.	The intervention improved mean externalizing behavior for both males and females ($p < 0.05$). It had a borderline statistically significant impact on internalizing behavior. The program also benefited a wide range of later life outcomes primarily through noncognitive skills.
Holmlund and Silva [2009]	<u>Outcome(s)</u> : <i>academic performance</i> – average of standardized test scores in English, Math, and Science <u>Intervention</u> : participation in the “xl programme” targeting the noncognitive skills of secondary school students	<u>Data</u> : “xl club programme,” National Pupil Database (NPD), Pupil Level Annual Schools Census (PLASC) ; 2,333 and 259,189 treated and control students aged 14 in England (2004) <u>Methods</u> : logit, propensity score matching, OLS, difference-in-difference, double differences, random-growth model	<u>Controls</u> : sex, language, eligibility for school meals, special needs status, and race <u>Timing of Measurements</u> : The data contains test scores from age 11, age 14 (both before the program), and age 16 (after the program).	Unconditional on observables, the performance of the students in the xl club is 1.2 to 1.4 standard deviations lower than the control subjects ($p < 0.01$). Using OLS, the effect is -0.17. The propensity score estimates are -0.13 and -0.15. For the difference-in-difference models estimated using OLS and propensity score matching, there is no longer a significant effect of the program in either direction. Overall the program had little effect.

Jackson, Hill, Payne et al. [2010]	<u>Outcome(s)</u> : <i>Personality</i> – Openness to Experience <u>Intervention</u> : participation in a 16-week inductive reasoning training program coupled with 10 hours of puzzle solving per week	<u>Data</u> : collected by the authors; 183 adults aged 60 to 94 <u>Methods</u> : random assignment, latent growth model	<u>Controls</u> : n/a <u>Timing of Measurements</u> : Openness to Experience was measured pre-test, at week 5, at week 10, and post training.	On average, participants in the program were 0.39 standard deviations higher in Openness to Experience after the program relative to people in the control group ($p < 0.05$).
Martins [2010]	<u>Outcome(s)</u> : <i>Educational attainment</i> – grade retention <u>Intervention</u> : participation in the EPIS program that boosts non-cognitive skills including motivation, self-esteem, and study skills	<u>Data</u> : EPIS database; 15,307 students in 7 th - 9 th grade in Portugal <u>Methods</u> : linear probability model, quasi-randomization	<u>Controls</u> : student fixed effects, time fixed effects <u>Timing of Measurements</u> : Information on each student is tracked for each quarter.	The program reduced annual grade retention by at least 10.1 percentage points ($p < 0.001$).

Rodríguez -Planas [2010]	<u>Outcome(s):</u> <i>educational attainment</i> – high-school completion and post-secondary education; <i>academic achievement</i> – math test score percentile, reading test score percentile , GPA; <i>labor market success</i> – earnings during the last year of the program, three years after the program, and five years after the program <u>Intervention:</u> – participation in the Quantum Opportunity Program (QOP), centered around mentoring and providing incentives for academic success	<u>Data:</u> Quantum Opportunity Program (QOP); 1,069 students from seven large US cities <u>Methods:</u> random assignment	<u>Controls:</u> n/a <u>Timing of Measurements:</u> The program was offered for a cohort of ninth graders and was available for five years. Follow-up interviews were conducted during the last year of the program, three years after the program, and five years after the program.	<u>During last year of the program:</u> Participation in the program was associated with a 7 percentage point increase in the probability of graduating high school ($p < 0.10$) and 6 percentage point increase in the probability of attending college ($p < 0.10$). There were no differences in academic achievement. <u>Three years after the program:</u> Participation in the program was associated with a 7 percentage point increase in the probability of ever attending college ($p < 0.10$), 9 percentage point increase in the probability of attending college ($p < 0.05$), and a 7 percentage point decrease in the probability of having a job ($p < 0.10$). <u>Five years after the program:</u> There are no significant differences five years after the program. <u>Findings for sub-populations:</u> The program benefited people who were 14 or less upon entering high school significantly more than older students. It also tended to benefit girls more than boys.
--------------------------	--	--	--	---

Stevens et al. [2008]	<u>Outcome(s)</u> : <i>attention</i> – ERP index of selective auditory attention; <i>language skills</i> – Clinical Evaluation of Language Fundamentals-3 <u>Intervention</u> : Participation in a six-week (100 min/day) computerized training program for boosting language skills (Fast ForWord program)	<u>Data</u> : collected by the authors; 33 children aged 7 on average <u>Methods</u> : random assignment	<u>Controls</u> : Test scores were normalized by age <u>Timing of Measurements</u> : Measures were taken before the beginning of the program and then six weeks later.	The increase in the attention was 0.81 standard deviations higher for the participants than for the non-participants ($p < 0.01$). The increase in the receptive language scores was 0.91 standard deviations higher in the participants than for the control group ($p < 0.01$). There was no significant effect on expressive language scores between the participants and the control group.
-----------------------	--	---	---	--

Source: Almlund, Duckworth, Heckman, and Kautz (2010)