Universidad de San Andrés

Departamento de Economía

Maestría en Economía

"Análisis sobre las Diferencias Interraciales en la Brecha Educativa entre Géneros"

Esteban Aucejo

DNI: 29.250.931

Mentor: Ignacio Franceschelli

Buenos Aires,

Argentina

6 de Marzo, 2020

Tesis de Maestría en Economía de

Esteban Aucejo

"*Análisis sobre las Diferencias Interraciales en la Brecha Educativa entre Géneros*"

Resumen

La considerable brecha de género en el enrolamiento universitario, especialmente entre los afroamericanos, constituye una regularidad empírica que puede tener serias consecuencias en la formación familiar, crianza de los hijos y participación de la fuerza laboral masculina. Por ejemplo, solo el 35.7 por ciento de todos los estudiantes afroamericanos universitarios estadounidenses eran hombres en 2004. Los resultados muestran que, mientras las variables de antecedentes familiares no pueden explicar la brecha observada, las variables sobre habilidades no-cognitivas son cruciales para explicarlo. Por otra parte, un modelo secuencial de avance educativo indica que los hombres tienen en realidad "mayores preferencias" por la educación que las mujeres una vez que se controla por factores latentes (habilidades cognitivas y no-cognitivas). El modelo también muestra que si bien las habilidades cognitivas afectan fuertemente la posibilidad de progresar hacia mayores niveles académicos, especialmente después de concluida la escuela secundaria, no pueden explicar las disparidades entre los géneros. Por el contrario, las diferencias sustanciales en la distribución de las habilidades no-cognitivas entre hombres y mujeres son críticas para explicar la brecha de género en el logro educativo.

Palabras clave: Brecha, géneros, enrolamiento universitario, habilidades no-cognitivas, raza

"Analysis of the Racial Differences in the Educational Gender Gap"

Abstract

The sizable gender gap in college enrollment, especially among African Americans, constitutes a puzzling empirical regularity that may have serious consequences on family formation, parenting, and male labor force participation. For instance, only 35.7 percent of all African American undergraduate students were men in 2004. Reduced form results show that, while family background covariates cannot account for the observed gap, proxy measures for non-cognitive skills are crucial to explain it. Moreover, a sequential model of educational attainment indicates that males have actually "higher preferences" for education than females after controlling for latent factors (i.e. cognitive and non-cognitive skills). The model also shows that cognitive skills strongly affect the decision to move from one school level to the next, especially after finishing high school, but cannot account for disparities between genders. On the contrary, the substantial differences in the distribution of non-cognitive skills between males and females are critical to explain the gender gap in educational attainment across and within races.

Keywords: Gender Gap, College Enrollment, Non-cognitive Skills, Race

Códigos JEL: I2, J15, J16

.

1 Introduction

The gender composition of U.S. college campuses has changed dramatically since the 1950's. In 1950, males represented 68% of college enrollees but by 1970 this number had fallen to 52%.[1] Goldin et al. (2006) indicate that the elimination of institutional and social barriers that prevented women from pursuing higher education coupled with changes in expectations, labor force participation, age at first marriage and improvements in high school performance contributed to a convergence between genders. However, not only did women catch up to men in terms of college enrollment, in the last three decades women have overtaken men, and by a substantial margin. According to the National Center of Educational Statistics (NCES), females represented 57% of the total fall enrollment in degree-granting institutions in 2004. Moreover, this empirical regularity masks considerable heterogeneity by race, while 56% of white undergraduates enrolled in 2004 were women, approximately 2 out of 3 African American students in college (64%) were females.[2] Indeed, the fact that the difference in the total population proportion[3] of white and black females attending college (13%) is smaller than the difference between black females and black males (17%) highlights the importance of the gender imbalances.

A detailed analysis of this empirical regularity, especially among African Americans, is key given its economic, demographic, and social implications. First, earnings and employment prospects of less educated workers have shown a sharp decline since the early 1980's [Acemoglu and Autor (2011)], while education has become an increasingly important determinant of lifetime income. This fact suggests that recent cohorts of less-educated males, and in particular black males, have fewer opportunities in the labour market than their predecessors. Second, differences in educational attainment contribute to low marriage rates affecting family formation and parenting [McDaniel et al. (2011)]. Less-educated males are not less likely to have children but they are less likely to settle with a partner. Over the past few decades, marriage rates of black men without high school degree have decreased substantially, with an accompanying sharp rise in out-of-wedlock births and single parent family formation. In this regard, children of less educated males are more likely to come from broken families, facing larger risks of academic achievement deficits[4] that may perpetuate current inequalities.[5] To sum up, current gender differences in educational attainment

[1] U.S. Department of Education, NCES, Digest of Education Statistics, 2008. Table 189

[2] U.S. Department of Education, NCES, Digest of Education Statistics, 2005. Table 23.1. See also McDaniel et al (2011) for an analysis of historical trends in the educational gender gap among African Americans.

[3] The previous percentages only consider the sample of college enrollees.

[4] See PISA 2009 Results Volume II, Overcoming Social Background: Equity in Learning Opportunities and Outcomes, Heckman (2011), and Autor and Wasserman (2013).

[5] The shortage of black males in post-secondary education may also weaken efforts to increase college campus

pose significant challenges from many policy perspectives. Therefore, a thorough analysis of this empirical regularity will contribute to the design of effective interventions that aim to reduce current inequalities across genders and races.

The empirical strategy of this paper is twofold. First, reduced form results, using the National Longitudinal Survey of Youth 1997 (NLSY97),[6] indicate that family background covariates can fully account for differences in college enrollment across races; however, these variables do not seem to have an impact on the gender gap. By contrast, the inclusion of proxy measures for non-cognitive/socio-emotional[7] skills at quite early stages of schooling career (i.e. grade retention, suspensions at school, academic performance in middle school, involvement in fights, and precocious sex) can fully explain gender disparities in college enrollment for all racial groups.

Second, in order to further investigate the importance of non-cognitive and cognitive skills, a sequential model of educational attainment, together with a measurement system for the identification of latent factors (i.e. cognitive and non-cognitive skills) is estimated. This model specification provides four main advantages. First, it incorporates into the analysis the key fact that postsecondary attainment is the result of previous educational decisions; enrolling in college depends on graduating from high school which also depends on finishing grade 10 and so on. Second, a complete profile of the schooling career path of males and females can be recovered, thus helping to identify the educational levels at which boys are more likely to leave education. Third, the inclusion of two latent factors associated with cognitive and non-cognitive skills makes it possible to control for dynamic selection and to deal with measurement error, given that noisy proxies likely provide biased estimates. Fourth, the effects of non-cognitive skills can be distinguished from the cognitive ones.[8] Therefore, changes in the relative importance of these skills can be determined at the different educational transitions.[9] Cameron and Heckman (1998) and (2001), Heckman, Stixrud and Urzua (2006), and Urzua (2008) among others, have shown the relevance of factor models in accounting for the effects of skills and family background characteristics in educational attainment. For instance, Heckman, Stixrud and Urzua (2006) show that both cognitive and non-cognitive skills

diversity, which is "*a compelling state interest,*" as Justice Sandra O'Connor wrote for the majority in *Grutter v. Bollinger* (2003).

[6] NLSY97 shows gender imbalances in educational attainment similar to those highlighted in the NCES.

[7] Cognitive skills reflect an individual's ability to think. The terms socio-emotional and non-cognitive skills, used as synonyms in this manuscript, reflect the ability to: understand and manage feelings, follow appropriate social behaviors and develop manners. According to psychologists, socio-emotional skills are critical because they facilitate engagement in learning, promote positive peer relationships, buffer children against risk and benefit mental health.

[8] Covariates such as school GPA, which are generally included in OLS strategies, are a function of cognitive and non-cognitve skills [Poropat (2009), and Duckworth and Seligman (2005)].

[9] Isolating the effect of each skill may also help to explain differences in socioeconomic outcomes. Moreover, policy recommendations may be different if non-cognitive skills turn out to be important for educational attainment. According to Cunha et al. (2005) these skills are more malleable at later stages of life than the cognitive ones.

have considerable effects on graduating from four-year institutions. In a similar vein, Urzua (2008) finds, using the male sample of the NLSY79, that unobserved non-cognitive skills play a key role in closing the black-white gap in incarceration rates. In this regard, Heckman et al's approach is (partially) methodologically extended in this manuscript to examine differences across genders.[10]

Estimation results reveal that disparities in skills, especially in non-cognitive ones, can explain the gender gap across all races. Indeed, males are shown to have "higher preferences" for educational attainment than females after controlling for the latent factors. For example, gender differences in college enrollment among African Americans would be 29% greater if black males had preferences for schooling equivalent to those of black females. This result is consistent with part of the economics literature that finds disparate incentives for educational attainment between males and females. For instance, Becker et al. (2010) argue that the expected benefits of schooling are still higher for males than for females; and Hubbard (2011) shows that college premium for women is not larger than for men once topcoding biases (in CPS survey) are corrected for.

In addition, estimation results show that the large gap in educational attainment between African American males and females is mainly explained by the substantial gender differences in non-cognitive skills distribution.[11] Simulation exercises indicate that if black gender disparities in skills mirrored the white ones, then the size of the gap would be the same for both races.

Policy recommendations that intend to improve educational attainment or close the gender gap may depend on the relative importance of one type of skill over the other. In this regard, cognitive skills show a greater impact (conditional on reaching a certain grade) on the probability of transitioning from one schooling level to the next than non-cognitive skills do, especially after finishing high school. However, the substantial disparities in the distribution of non-cognitive skills between males and females make these skills more pertinent to the gender gap size. Results indicate that if young men had the non-cognitive skill distribution of women, the gender gap in educational attainment would be closed. But, this outcome could not be obtained if, instead, cognitive skills distributions were equalized. Finally, an analysis of the changes in the mean of the factor distributions at each transition of schooling career[12] suggests that selection into college is driven by both skills but with a higher emphasis on cognitive ones. However, the considerable gender disparities in non-cognitive skills prevent many males (relative to females) from finishing high school; which is a necessary step to enroll in postsecondary education. To sum up, differences in skills between males and females at early stages in life can fully explain the disparities in educational

[10] For example, the factors are independent of the observables in Heckman, Stixrud and Urzua (2006). However, this paper allows (after imposing additional structure to the model) factors to be correlated with a subset of the students characteristics (i.e. gender and race).

[11] This result also applies to the other racial groups.

[12] Remember that in each schooling transition a given proportion of students leave the education system, which leads to changes in the distribution of skills (i.e. selection process).

attainment.

This work builds on Jacob (2002) which is the first paper that attributes a key role to non-cognitive skills in explaining the gender gap in college enrollment. However, his findings are somewhat weaker than the ones presented in this manuscript. For instance, Jacob shows that these skills only account for 42% of male-female disparities, while in this paper non-cognitive skills fully account for the gender gap. Many reasons could explain the differences. First, the NLSY97 includes a different set of measures of non-cognitive skills than Jacob's dataset (NELS 88:94). Moreover, his sample is only based on high school graduates in which college attendance rates are higher than those documented in other national surveys. While the 1990 Census shows that 51.4% of 19-21 year old women had some postsecondary education, the corresponding number in Jacob's dataset is 67.3%.

The rest of the document is organized as follows: section 2 describes the data and the gender gap in detail. Section 3 shows reduced form results. Section 4 presents a factor model of educational attainment. Section 5 describes the estimation outcomes. Section 6 discusses implications of the model. Section 7 concludes.

2 Data: NLSY97

2.1 Descriptive Statistics

The empirical content of this paper utilizes the National Longitudinal Survey of Youth of 1997 (NLSY97). Accordingly, before characterizing in more detail the gender gap in postsecondary attainment, it is appropriate to briefly describe this database. The NLSY97 is a nationwide representative sample of youths who were 12 to 17 years old when they were first surveyed in 1997. It collects extensive information on family background characteristics, educational experiences and labor market behavior through time, with the aim to document the transition of the survey participants from school to work and into adulthood.

It can be seen from the NLSY97 that gender disparities in educational attainment are present across all ethnic groups. However, its magnitude varies substantially across races. Table 1 indicates that the proportion of white males (before age 25) enrolled in college (52.8%) was considerably smaller than white females (63.1%).[13] However, a wider gap can be shown among African Americans, where the percentage of black men and women enrolled in college is 32.6% and 49.7%, respectively. In addition, notice that black girls are significantly less likely to drop out of high school than black boys; for instance, 28.5% of black males could not obtain a high school diploma,

[13] Percentages are expressed as proportions of total subsamples (e.g. total white males). Notice that some of the summary statistics presented in the previous section were reported as proportion of undergraduate subsamples (e.g. whites enrolled in college).

versus only 18.9% of black females. This last outcome is surprising given the high pregnancy rates of African American adolescents (17%).[14]

Table 1 also illustrates that disparities in college enrollment between black and white men (20.2%) are only 3.1% points higher than the gender gap among African Americans (-17.1%). Moreover, differences in postsecondary attainment between white and black females (13%) are substantially smaller than those between black females and males.[15] The fact that the size of the gender gap among African Americans is almost as big as the racial gap suggests that gender imbalances are relevant.

Educational Attainment as % of Total Demographic Group (Before age 25)						
	White		Black		Hispanic	
	Males	Females	Males	Females	Males	Females
High School Dropout	14.7%	12.2%	28.5%	18.9%	22.9%	19.2%
Only High School Degree	32.5%	24.7%	38.9%	31.4%	37.3%	35.5%
College Enrollment	52.8%	63.1%	32.6%	49.7%	39.8%	45.3%
Observations	2072	1908	1025	1078	843	829

Table 1: Educational attainment before age 25 split by race and gender, as % of total demographic group (e.g. white male). Source: NLSY97

In order to verify whether cross-racial differences in the educational gender gap are present in other national data sources, summary statistics from the National Center of Educational Statistics (NCES) database are also analyzed. The NCES collects information on enrollment, major and graduation rates split by gender and race from each postsecondary institution in the United States. NCES data shows patterns quite similar to the ones described above. For example, the difference in the percentage of undergraduate fall enrollment in degree-granting institutions between white males and females was -11.8%[16] in 2004; while among African Americans it was -28.6% (i.e. approximately 2 out of 3 black students in college were females). Moreover, disparities are even bigger if degrees granted by two or four year institutions are considered (see Appendix A). This fact may suggest that apart from enrolling in higher proportions, females are more likely to persist and/or finish their studies in a shorter period of time than males. An analysis of career paths indicates that

[14] National Vital Statistics System. U.S. Department of Health and Human Resources, CDC (2006). African American females of age 19 or less.

[15] This result is consistent with data from the CPS survey (years 2000-2001), when considering the population of 18 and 19 years old.

[16] Notice that this proportion was calculated based on the sample of college enrollees. More specifically, the percentages of white males and white females enrolled in degree granting institutions conditional on total white enrollment in 2004 were 44.1% and 55.9% respectively. Then, the difference was -11.8%. On the contrary, the proportions presented based on the NLSY97 were obtained considering the unconditional sample.

majors related to health professions and liberal arts and sciences are highly dominated by females; however, males are still a majority in engineering and computer science.[17]

To sum up, the NLSY97 and the NCES statistics provide consistent evidence related to differences in educational attainment between males and females. But this finding masks considerable heterogeneity by race. Given the large gender disparities among African Americans relative to the other ethnic groups, incorporating a racial dimension into the analysis is critical.

The empirical strategy of this paper makes intensive use of family background characteristics and proxies for cognitive and non-cognitive skills; accordingly, the following two subsections provide a detailed description of these variables based on NLSY97 data.

2.2 Family Background Covariates

Family background characteristics have a substantial impact on enrollment differentials based on race. However, these factors are less likely to account for a large proportion of the gender gap given that males and females belong on average to the same type of families. Table 2 shows means and standard deviations of selected family characteristics:[18] mother's education,[19] number of household members under the age of 18, family structure (i.e. dummy variable for broken family at age 12) and parenting style (i.e. dummy variables for authoritarian, authoritative, uninvolved or permissive). Parenting style is included in the analysis with the aim of controlling for any differential effects that parent-son/parent-daughter relationships may have in educational attainment. Psychologist Diana Baumrind (1991) has identified four patterns of parenting styles based upon two aspects of parenting behavior: control and warmth. 1) Authoritarian Parenting: little warmth and highly controlling, 2) Authoritative Parenting: warmth but firm, 3) Permissive Parenting: warmth but undemanding, 4) Uninvolved Parenting: no warmth and undemanding.

A brief analysis of Table 2 shows similar family background characteristics between girls and boys but significant differences across races. In this regard, a test of differences in means cannot reject the null hypothesis of equal means across genders, conditioning on race. For instance, white families present a smaller number of young household members, but there are no statistical differences across genders. In a similar vein, Table 2 also shows that white mothers are more educated than their black and Hispanic counterparts, and that black kids belong in much higher proportions to broken families (at age 12) than whites and Hispanics. Notice that this empirical regularity is

[17] See table A1 to A3 of Appendix A for more detailed information on gender differences across races, degrees granted, and majors.

[18] Family income was not included due to the extensive number of missing values and inconsistencies. For example, it was found that siblings that lived in the same house and with the same parents reported quite different amounts. Results do not change if family income is included in the empirical strategy.

[19] Mother education takes values 0 if high school drop out, 1 if only completed high school, 2 if some college, and 3 if college graduate (i.e. Bachelor's degree).

consistent with the sharp rise (over the past few decades) in out-of-wedlock births and single parent family formation among African Americans.

	Family Background Covariates: Means and Standard Deviations (NLSY97)					
	White		Black		Hispanic	
	Males	Females	Males	Females	Males	Females
Mother Education	1.62	1.62	1.19	1.20	0.90	0.90
	(0.987)	(0.979)	(0.927)	(0.923)	(0.965)	(0.984)
Family Members	2.25	2.25	2.60	2.61	2.61	2.63
Under Age 18	(1.08)	(1.14)	(1.45)	(1.45)	(1.30)	(1.35)
Broken Family	0.412	0.444	0.797	0.786	0.494	0.502
	(0.492)	(0.497)	(0.402)	(0.410)	(0.500)	(0.500)
Mother Uninvolved	0.118	0.137	0.087	0.134	0.107	0.172
	(0.322)	(0.343)	(0.282)	(0.341)	(0.309)	(0.377)
Mother Permissive	0.375	0.352	0.307	0.304	0.352	0.301
	(0.484)	(0.478)	(0.461)	(0.460)	(0.479)	(0.459)
Mother Authoritarian	0.112	0.142	0.128	0.154	0.127	0.173
	(0.315)	(0.349)	(0.333)	(0.361)	(0.332)	(0.378)
Mother Authoritative	0.395	0.369	0.478	0.408	0.414	0.354
	(0.489)	(0.482)	(0.499)	(0.492)	(0.492)	(0.479)

Table 2: Means and standard deviations of family background characteristics split by race and gender. Source: NLSY97

2.3 Cognitive and Non-Cognitive Proxies

An emerging literature in economics, mainly developed by Heckman and his coauthors,[20] has provided substantial evidence about the relevance of cognitive and non-cognitive skills in a wide number of outcomes. For instance, Neal and Johnson (1996) and Cawley, Heckman and Vytlacil (2001) show that cognitive abilities constitute an important predictor of educational attainment and labor outcomes. However, differences in cognitive skills are not the unique source of observed disparities in educational attainment. Heckman and Rubinstein (2001) point out that GED recipients have similar cognitive abilities to high school graduates with 12 years of schooling. They suggest that lower levels of non-cognitive skills likely explain the observed differences in high school completion and labor outcomes. This conjecture was later confirmed by Heckman, Stixrud and

[20] Cawley, Heckman and Vytlacil (2001), Heckman and Rubinstein (2001); Cunha, Heckman and Navarro (2005); Carneiro, Hansen and Heckman (2003); Heckman, Stixrud and Urzua (2006); Heckman, Lochner and Todd (2006); Cunha and Heckman (2008); and Heckman, Humphries, Urzua and Veramendi (2012), among others.

Urzua (2006), where they find out that GED recipients exhibit a less favorable distribution of non-cognitive abilities than high school graduates.

In a similar vein, psychologists have also argued that children with low non-cognitive skill development are more likely to experience academic delays, enter school at risk of increasing behavior problems, be vulnerable to peer rejection, to dropout of school and to be involved in criminal and risky activities.[21] Therefore, given the relevance of these skills in educational attainment, it is appropriate to analyze their possible contribution to the college gender gap.

Psychologists have come to a consensus regarding an organizational framework that categorizes personality traits into a small number of domains, denoted as the Big Five. This taxonomy refers to five broad domains of personality that are used to describe human personality (extraversion, agreeableness, conscientiousness, neuroticism, and openness). It has been shown that conscientiousness[22] and agreeableness[23] play an important role in educational outcomes.[24] Therefore, including direct measures of these traits in the empirical strategy would be ideal. However, the difficulty of finding accurate measures of personality traits in large-scale surveys makes necessary the use of proxies. Based on the literature of personality psychology and economics, five proxies[25] (constructed until the age of 14) have been identified in the NLSY97 showing strong correlations with the aforementioned traits: GPA at grade 8, precocious sex, retention and suspensions in school between grades 1 and 8, and involvement in fights with the intention to hurt someone. Course grades, school retention, and reckless risk-taking have (mainly) been associated with conscientiousness. For instance, in a recent literature review, Almlund, Duckworth, Heckman, and Kautz (2011) indicate that conscientiousness is the most robust Big Five predictor of course grades. Moreover, they show evidence indicating that associations between grades and conscientiousness are almost as large as those between cognitive ability and grades. Similarly, Martins (2010) shows a strong link between non-cognitive skills levels and retention at school.[26] Gullone and Moore (2000) also investigate the links between personality traits and risk behavior. Their findings suggest that conscientiousness and reckless risk-taking are negatively correlated. More importantly, they show that adolescent females score higher in conscientiousness than male adolescents. Finally, Heckman, Pinto and Wang (2008)

[21] Domitrovich (2008), Nagin et al. (2001), Shaw et al. (2001), Payton et al. (2000), Brody et al. (2003), Ladd et al. (1999), Caspi et al. (1995) and White et al. (1990).

[22] Tendency to be organized, responsible, and hardworking.

[23] Tendency to act in a cooperative, unselfish manner. On the contrary, anti-social behavior and aggression denote low agreeableness.

[24] See Almlund, Duckworth, Heckman, and Kautz (2011) for a literaure review of the power of personality traits as predictors of academic success.

[25] These proxies are considered as indirect and noisy measures of non-cognitive and cognitive skills.

[26] Jacob (2002) also considers retention at school as a proxy for non-cognitive skills. In addition, Cooley, Navarro, and Takahashi (2013) also show the presence of a correlation between kindergarten behavioral ratings and retention at school.

also indicate that non-cognitive skills are highly correlated with risky behavior (e.g. involvement in precocious sex), and Armour et al. (2007) show that experiencing sexual debut earlier than one's peers is linked to a higher probability of engaging in risky activities (i.e. delinquency). To sum up, the literature seems to indicate that GPA at grade 8, retention at school, and precocious sex constitute reasonable proxies for non-cognitive skills (i.e. conscientiousness).[27]

Evidence also illustrates that agreeableness is an important predictor of schooling outcomes. For example, Duncan and Magnuson (2010) indicate that antisocial behavior (i.e. low agreeableness) predicts high-school completion. Consistent with this finding, Fergusson and Horwood (1998) show that teacher ratings of conduct problems at age 8 are negatively related to high-school completion at age 18. Therefore, involvement in fights and suspensions in school between grades 1 and 8 were included into the analysis given their correlation to agreeableness.

These proxies were constructed until the age of 14, due to the fact that students cannot make any educational decisions before age 16. In this regard, the implicit identifying assumption requires that these variables are not determined by the previous decision not to attend college. By way of illustration, the fact that a black boy is performing more poorly (e.g. lower GPA) in middle school than a black girl cannot be driven by his decision not to attend college later in life. Evidence on college expectations from the NLSY97 suggests that this assumption is reasonable. Black girls and boys between the age of 15 and 16 show small differences (2.1 percentage points)[28] in their reported expected chances of having a four year college degree by the time they turn 30.[29,30]

Table 3 outlines means and standard deviations of this set of variables. Conditional on race, males are more likely to be suspended from school, to be involved in fights with the intention to hurt someone, to engage in precocious sex and to be retained in at least one grade. Moreover, African Americans show the highest differences between males and females in almost all of these variables, for example, 24.9% of black males were suspended from school at age 14 while only 13.4% black females were in that same situation. Furthermore, African American males are considerably more likely to engage in precocious sex than black females. This disparity is consistent with data from the Youth Risk Behavior Surveillance System (YRBSS), a cross-sectional, nationally representative

[27] Heckman, Humphries, Urzua and Veramendi (2012) also include course grades in 9th grade and participation in reckless activity as noisy measures of socio-emotional ability.

[28] More specifically, black girls and boys reported 76.5 and 74.4 percent chance of having a four-year college degree by age 30, respectively. These expectations seem to be quite overly optimistic. However, whites and hispanics also reported large magnitudes. Finally, Carneiro, Heckman and Masterov (2005) also find overoptimistic expectations in educational attainment when using data from the CNLSY.

[29] The NLSY97 only asked the following question (in round one) to those who were 15 years old or older: "What is the percent chance that you will have a four-year college degree by the time you turn 30?"

[30] As it has been pointed out by Carneiro et al (2005), "*Expectation formation models are very complex, and difficult to test empirically, given that they often lead to multiple equilibria.*" Nevertheless, the reported evidence seems to support the claim that proxy measures are not likely to be biased by differences in expectations across genders.

	Non-cognitive Proxies: Means and Standard Deviations					
	White		Black		Hispanic	
	Males	Females	Males	Females	Males	Females
Suspensions at age 14	0.128	0.067	0.249	0.134	0.135	0.089
	(0.334)	(0.250)	(0.432)	(0.341)	(0.341)	(0.285)
Fights	0.210	0.101	0.274	0.174	0.202	0.136
	(0.407)	(0.302)	(0.446)	(0.379)	(0.402)	(0.343)
Precocious Sex	0.165	0.151	0.484	0.232	0.263	0.148
	(0.371)	(0.357)	(0.499)	(0.422)	(0.440)	(0.355)
Retention Grade 1 to 8	0.125	0.082	0.277	0.171	0.172	0.121
	(0.331)	(0.275)	(0.448)	(0.377)	(0.377)	(0.325)
GPA Grade 8 std.	-0.012	0.385	-0.448	0.004	-0.289	0.050
	(1.035)	(0.954)	(0.887)	(0.899)	(0.963)	(0.933)

Table 3: Summary statistics (weighted): means and standard deviations of cognitive and non-cognitive proxies split by race and gender. "std." denotes standardized. Source: NLSY97

survey of students in grades 9-12 established by the Centers for Disease Control and Prevention. Using this database, Cavazos-Rehg et al. (2009) show that by the 14th birthday, the likelihood of sexual debut is 42% for African American males while it is 17% for African American females. Finally, in terms of performance in school, females (conditional on race) obtained higher GPAs at grade 8 than males. To sum up, the NLSY97 data shows that females (conditional on race) do better in all the described proxy measures. Robustness checks indicate that other national databases (as is shown in the following section) can replicate the patterns of Table 3. In terms of differences across races, black males are substantially more likely to be involved in fights, be suspended from school, and to exhibit lower academic performance than white males. However, these differences are smaller when comparing white and Hispanic males, being consistent with the fact that the Hispanic-white male gap in educational attainment is smaller than the black-white gap.

Finally, measures related to cognitive skills come from the Armed Forces Vocational Aptitude Battery (ASVAB) test scores (i.e. mathematical knowledge, arithmetic reasoning, word knowledge, paragraph comprehension, assembling objects and general sciences).[31] ASVAB scores have been widely used in the economics literature as proxies for these skills.[32] Table 4 indicates that white

[31] Coding Speed and Numerical Operations tests were not included because they were administered in a different format (i.e. non-adaptive, all respondents answer the same items in the same order).

[32] See references in footnote 20.

	Cognitive Proxies: Means and Standard Deviations					
	White		Black		Hispanic	
ASVAB Tests	Male	Female	Male	Female	Male	Female
General Sciences	0.483	0.300	-0.528	-0.518	-0.195	-0.357
	(0.977)	(0.850)	(0.889)	(0.842)	(0.930)	(0.868)
Arithmetic Reasoning	0.367	0.314	-0.567	-0.492	-0.140	-0.179
	(0.966)	(0.840)	(0.991)	(0.906)	(0.948)	(0.913)
Math Knowledge	0.225	0.372	-0.537	-0.309	-0.185	-0.138
	(0.962)	(0.920)	(0.948)	(0.938)	(0.965)	(0.923)
Assembling Objects	0.205	0.346	-0.578	-0.428	-0.028	0.014
	(1.102)	(0.916)	(0.861)	(0.892)	(0.958)	(0.928)
Word Knowledge	0.358	0.351	-0.531	-0.414	-0.223	-0.243
	(0.951)	(0.888)	(0.959)	(0.929)	(0.930)	(0.870)
Paragraph Comprehension	0.184	0.428	-0.605	-0.271	-0.230	-0.071
	(0.983)	(0.884)	(0.925)	(0.929)	(0.978)	(0.886)

Table 4: Summary statistics (standardized and weighted): mean and standard deviation of cognitive proxies open by race and gender. Source: NLSY97.

and Hispanic males performed better than their female counterparts in general sciences, arithmetic reasoning and word knowledge, while the opposite was true in math knowledge, assembling objects and paragraph comprehension. However, the picture is different among African Americans, where females performed better than males in all categories.[33]

The following section shows reduced form evidence that highlights the importance (if any) of family background characteristics, non-cognitive and cognitive skills to explain the gender gap in

[33] There may be some concerns that ASVAB scores are racial or gender biased. For example, the literature on "stereotype threat" claims that measured test scores for minorities understate their true ability [Steele et al. (1998), and Rodgers et al. (1996)]. However, other strand of the literature indicates that this is not likely to be a problem. For example, Carneiro, Heckman and Masterov (2005) claim that "*the stereotype literature substitutes wishful thinking for substantial evidence. There is no evidence that it accounts for an important fraction of minority-white test score gaps or that test scores are not good measures of productivity*". Moreover, Kass, Mitchell, Grafton, and Wing (1983) factor analyzed a sample of more than 98,000 ASVAB scores from army applicants. Basically, they found few meaningful differences in factor loadings across race/ethnic group or gender. In a similar vein, Ree and Carretta (1995) studied ASVAB scores from a portion of the NLSY79 sample. They conclude that predictiveness of AFQT should be consistent across racial and gender groups. Finally, Neal and Johnson (1996) summarize the results of a National Academy of Sciences study (for the Department of Defense) that found that AFQT predicts performance in tasks required for military occupations about equally well across races. In this regard, they indicate that AFQT score provides an unbiased measure of pre-market job preparation. However, whether these results can be generalized to jobs outside of the military is unknown.

college enrollment.[34]

3 Reduced Form Evidence

3.1 Family Background Characteristics

In order to study the role of family covariates in the gender gap, two OLS specifications are initially analyzed. The first regression includes as a dependent variable an indicator for college enrollment before age 25,[35,36] and as independent variables gender, race and their interactions (this specification will work as the benchmark case). The second regression adds the following covariates: number of family members under age 18, mother's education, parenting style and an indicator that denotes whether a given kid belonged to a broken family at age 12.[37,38] Column 1 of Table 5 shows that the size of the gender gap is around -10% for whites and Hispanics, and -17% for African Americans. Notice that the gap between black males and black females (17.5%) is larger than between white and black females (12.3%). In addition, this specification also exhibits the presence of the well known disparities in educational attainment across races. The inclusion of family background covariates generates quite interesting results (column 3); while ethnic differences in college enrollment are fully explained,[39] gender disproportions remain fairly constant for all racial groups. In this sense, the persistence of the gender gap after controlling for family covariates is coherent with the notion that males and females come (on average) from the same type of families; therefore large effects are not strongly anticipated.

[34] The empirical strategy of this paper does not directly study the role of incarceration rates in the gender gap. McDaniel el al. (2011) and Sharpe and Darity (2009) indicate that incarceration only explains a small part of the black gender gap. In a similar vein, Fumia (2013) shows, using the NLSY97, that incarceration cannot directly explain racial gaps in educational attainment. These results are consistent with the idea that those at risk of incarceration are also less likely to attend post-secondary education even in the absence of incarceration. Duckworth and Urzua (2009) show that the number of arrests between 14 and 17 years old is correlated with conscientiouness and agreeableness. Therefore, the current empirical strategy accounts for incarceration indirectly (i.e. as a manifestation of cognitive and non-cognitive skills). Finally, the prison population is tracked by the NLSY97 interviewers, and in particular NLS program staff take special measures so respondents can be interviewed while in jail.

[35] OLS regression results are presented because it is straightforward to interpret the coefficients. Logit specifications present similar outcomes.

[36] The dependent variable (i.e. college enrollment) in Tables 5 to 8 is an indicator for whether an individual enrolled in college before age 25. This variable was constructed based on all individuals who completed at least 8 years of education. This implies dropping around 0.5% of the observations in the sample.

[37] Measures of broken family at age 2 or 6 do not change the results.

[38] Family income was not included due to the extensive number of missing values and inconsistencies. For example, it has been found that siblings living in the same house and with the same parents reported quite different amounts. In addition, OLS regressions that include income show that the main results are similar.

[39] This result is consistent with the findings of Cameron and Heckman (2001).

OLS Regressions				
Dependent Variable: College Enrollment				
Variables	Coef.	Std. Err.	Coef.	Std. Err.
	(1)	(2)	(3)	(4)
Constant	0.651***	0.012	0.557***	0.023
Male	-0.101***	0.016	-0.110***	0.015
Black	-0.123***	0.021	0.012	0.020
Black x Male	-0.074**	0.030	-0.078**	0.028
Hispanic	-0.164***	0.024	-0.027	0.023
Hispanic x Male	0.011	0.033	0.007	0.031
Broken Family	-	-	-0.182***	0.013
Mother Education	-	-	0.155***	0.006
Fam. Mem. under Age 18	-	-	-0.016***	0.005
Mother Uninvolved	-	-	-0.145***	0.021
Mother Permissive	-	-	-0.054***	0.014
Mother Authoritarian	-	-	-0.081***	0.019
R^2	0.034		0.199	
Observations	6643		6643	

Table 5: OLS regressions. *, ** and *** indicate significance at 10%, 5% and 1% level, respectively. Robust standard errors are reported. The four (mutually exclusive) patterns of parenting styles are: authoritarian, permissive, uninvolved and authoritative (see definition in subsection 2.2), where the omitted category is authoritative style. The sample was restricted to those who completed at least eighth grade. This resulted in the loss of only 0.5% of the observations. Source: NLSY97.

However, it is still possible that family characteristics may have large differential effects based on gender. For example, a broken family may have more detrimental effects on boys than on girls due to the lack of male role models at home. In order to test for this possibility, similar specifications to those presented in Table 5 are reported, but in this case interactions between gender and family covariates are included (in order to keep the size of the table tractable, results considering only the African American sample are shown).[40] Table 6 shows that in each of the three specifications black males are statistically less likely to attend college than black females, in addition the size of male's coefficient remains approximately stable across specifications. Results in column 5 indicate that family covariates seem not to have differential effects on males given the lack of statistical significance of the interactions. Moreover, a joint test of significance cannot reject the null hypothesis that the OLS coefficients on the interacted variables are equal to zero. Therefore, when considering together the results from Tables 5 and 6 the evidence seems to indicate that broad measures of family background cannot fully explain the observed gender disparities in college enrollment, but can fully explain the racial gap. These results suggest that the channels affecting the gender gap in college enrollment are unlikely to be the same as those that explain the racial gap. Given this, the following subsection analyzes whether differences in skills between males and females can explain the gender gap.

3.2 Cognitive and Non-Cognitive Skills

Substantial gender disparities in behavior and school performance (i.e. proxy but noisy measures for non-cognitive and cognitive skills) can be found at the different schooling levels. For instance, boys show a greater prevalence of behavior problems than girls even at quite early stages of life. Lavigne et al. (1996) find, based on a large sample study of preschool children, that while 6.6% of preschool females presented some kind of behavioral problem, the figure for males was 10%. In a similar vein, evidence from the National Prekindergarten Study[41] indicates that preschool boys are 4.5 times more likely to be expelled than girls [Gilliam, (2005)]. Moreover, many studies have shown an overrepresentation of boys suffering attention deficit disorders; depending on the type of setting (i.e. community or clinical) boy/girl ratios go from 3:1 to 5:1 respectively.[42] These gender disparities should not be disregarded given that behavior problems have large negative effects on schooling attainment, regardless of income and maternal education [Currie and Stabile (2006)]. This problem may be magnified for African American children who are 1.92 times more likely to

[40] Similar results can be found for the other racial groups.

[41] A comprehensive data collection effort across each of the nation's 52 state-funded prekindergarten programs operating in the 40 states that fund prekindergarten.

[42] Diagnostic guidelines that were released in 2000 estimated the prevalence of ADHD to be between 4% and 12% of school-aged children [Schneider et al. (2006)]

OLS Regressions (Black Sample)						
Dependent Variable: College Enrollment						
Variable	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.528***	0.018	0.451***	0.044	0.465***	0.062
Male	-0.175***	0.025	-0.181***	0.024	-0.210**	0.084
Broken Family	-	-	-0.088***	0.029	-0.092**	0.039
Mother Education	-	-	0.179***	0.013	0.193***	0.017
Family Mem. under Age 18	-	-	-0.021**	0.008	-0.023*	0.012
Mother Uninvolved	-	-	-0.066*	0.038	-0.120**	0.051
Mother Permissive	-	-	-0.009	0.027	-0.047	0.038
Mother Authoritarian	-	-	-0.033	0.038	-0.073	0.050
Broken Family x Male	-	-	-	-	0.010	0.057
Mother Education x Male	-	-	-	-	-0.026	0.025
Family Mem. under 18 x Male	-	-	-	-	0.003	0.016
Mother Uninvolved x Male	-	-	-	-	0.123	0.078
Mother Permissive x Male	-	-	-	-	0.072	0.054
Mother Authoritarian x Male	-	-	-	-	0.083	0.078
R^2	0.031		0.176		0.179	
Observations	1732		1732		1732	

Table 6: OLS regressions based on the African American sample. *, ** and *** indicate significance at 10%, 5% and 1% level, respectively. Robust standard errors are reported. The sample was restricted to those who completed at least eighth grade. This resulted in the loss of only 0.5% of the observations. Source: NLSY97.

be labeled as emotionally disturbed.[43]

According to the NCES, special education courses are dominated by males.[44] More specifically, 12% of the students in kindergarten, first, or third grade received this type of education service; with boys are more likely than girls (16% and 8% respectively). Consistent with this empirical regularity, the proportion of male students who were delayed in kindergarten through grade eight during the year 2007 was 11.7%, while for females the proportion was only 7.6%.[45]

Evidence related to academic performance in elementary school[46] indicates that fourth grade females almost close the historical gap in math exams. For example, they performed (in large cities) as well as boys in the math National Assessment of Educational Progress (NAEP) tests of the year 2009. However, girls' advantage in reading exams is still robust (especially among African Americans).[47] In addition, 38% of boys and 31% of girls in fourth grade could not achieve the basic level in the reading NAEP exam, while an identical proportion of males and females (19%) were below the basic level in math.[48]

The percentage of public school male students in kindergarten through 12th grade who were suspended was 9.2% in the year 2000, while among females it was only 3.9% (NCES). These proportions become much higher for African American men and women with a 17.4% and 9.1% respectively. In the NLSY97, 42% of boys agreed with the statement *"When I was in school, I used to break rules quite regularly"* while only 24% of females did so. Higher rates of suspensions for boys are not likely to be irrelevant in terms of the achievement gap given that prior research has confirmed that students who have been suspended from school are at higher risk of other poor school outcomes, including dropping out of school. For example, Segal (2011) finds that those kids who misbehaved in the 8th grade are almost three times more likely to drop out of high school and almost three times less likely to graduate from college. The negative correlation between 8th grade misbehavior and educational attainment remains even after controlling for test scores and family characteristics.[49]

Gender disparities in behavior and school performance seem to persist in high school; for instance a large study of Minnesota adolescents [Harris, Blum and Resnick (1991)] found that a higher percentage of teenage boys reported frequent antisocial acts (e.g. vandalism) as compared

[43] Racial Inequity in Special Education (2002)

[44] Timing and Duration of Student Participation in Special Education in the Primary Grades, March 2007, NCES 2007–043

[45] The Condition of Education 2009, Indicator 18, Grade Retention. NCES.

[46] See Cornwell et al. (2012) for an analysis of the gender gap in test score performance in elementary school.

[47] Table B1 in appendix B shows mean scores and standard deviations of math and reading NAEP exams for fourth grade students open by race.

[48] Data extracted form the NCES website on June 2010, http://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx Additional evidence for 8th grade students can be found in appendix B of this paper, Table B2.

[49] See Bertrand et al. (2011) for an analysis of the gender differences in non-cognitive skills.

to teenage girls (i.e. 10% and 6%, respectively). In addition, another set of studies in psychology finds that females are less often engaged in problem behaviors and are likely to terminate their involvement in such behaviors more quickly than boys [Ensminger (1990), Petersen, Richmond and Leffert (1993), Lerner and Steinberg (2004)]. In terms of academic performance in high school, boys are more likely to drop out;[50] and among those who graduate, females performed better than males with mean GPA of 3.05 and 2.83 respectively in 2000.[51] In addition, girls are more likely than boys to enroll in college preparatory courses,[52] and to participate in all types of after school activities except for athletics. For example, in 2001, 19.2% of high school females seniors and 11.8% of males reported participating in academic clubs.[53]

Therefore, these preliminary facts from national samples suggest that gender differences in cognitive and non-cognitive/socio-emotional skills (measured as differences in behavior and academic performance during pre, elementary and high school) may explain the observed disparities in educational attainment.

Table 7 highlights the likely importance of non-cognitive skills in educational attainment. First, college enrollment before age 25[54] was regressed on gender, race, and their interactions (this specification will work as the benchmark case). Then, proxies for non-cognitive and cognitive skills were included (i.e. suspensions from school,[55] retention at school between grade 1 and 8, GPA at grade eight, involvement in fights and precocious sex).[56] Finally, family covariates were added. In order to alleviate possible problems of endogeneity, these variables were constructed at early points in life (i.e. up to age 14). Results in column 3 show that after including skill proxies, the gender gap is fully explained across all races. In addition, notice that racial disparities between African Americans and whites are also fully explained, but Hispanics are still less likely to enroll in college. However, column 5 indicates that after including family covariates (i.e. mother education), Hispanics are no longer less likely to attend college. To sum up, these regressions suggest that proxy measures for cognitive and non-cognitive skills are relevant factors that can explain the gender gap within and across races.

[50] Although males comprise 51 percent of the population between 16 to 24 years old, they make up 58 percent of the dropouts in this age group. Source: U.S. Census Bureau, School Enrollment–Social and Economic Characteristics of Students: October 2005.

[51] The High School Transcript Study: A Decade of Change (2001). NCES.

[52] In 2005, 38% and 35% of high school graduates females and males respectively, completed college preparatory basic courses. NCES (2008).

[53] Trends in Educational Equity of Girls & Women. NCES (2004).

[54] OLS regression results are presented because it is straightforward to interpret the coefficients. Logit models provide similar results.

[55] This variable measures the number of academic years in which a student was suspended from school.

[56] Subsection 2.3 provides a discussion of the identifying assumptions and an explanation of why these proxies have been chosen.

OLS Regressions						
Dependent Variable: College Enrollment						
Variable	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.651***	0.011	0.616***	0.011	0.430***	0.015
Male	-0.105***	0.016	-0.009	0.014	-0.022	0.014
Black	-0.138***	0.021	-0.021	0.019	0.017	0.019
Black x Male	-0.076***	0.029	-0.018	0.026	-0.016	0.026
Hispanic	-0.176***	0.023	-0.100***	0.021	-0.016	0.022
Hispanic x Male	0.047	0.032	0.048*	0.030	0.027	0.030
Grade Retention	-	-	-0.164***	0.017	-0.133***	0.017
GPA grade 8 std.	-	-	0.195***	0.006	0.174***	0.006
Suspensions	-	-	-0.035***	0.006	-0.030***	0.006
Fights	-	-	-0.028*	0.017	-0.024	0.016
Precocious Sex	-	-	-0.124***	0.016	-0.107***	0.016
R^2	0.036		0.277		0.329	
Family Covariates	No		No		Yes	
Observations	7061		7061		6560	

Table 7: OLS regressions. *, ** and *** indicate significance at 10%, 5% and 1% level, respectively. Robust standard errors are reported. The sample was restricted to those who completed at least eighth grade. This resulted in the loss of only 0.5% of the observations. See Table C1 in appendix C for similar regressions as in tables 5 and 7, where the only difference is that the sample is kept constant across specifications (results do not change). Source: NLSY97.

It may be possible to argue that these results are in fact just driven by the cognitive component of these variables. However, Table 8 shows that similar regressions that include ASVAB[57] test scores (i.e. cognitive measures that have been extensively used in the literature) can explain the racial gap,[58] but cannot explain the gender gap. To sum up, results from Tables 7 and 8 suggest that non-cognitive skills do play a key role in explaining the gender gap in college enrollment.

OLS Regressions						
Dependent Variable: College Enrollment						
Variable	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.663***	0.012	0.554***	0.011	0.476***	0.021
Male	-0.104***	0.017	-0.070***	0.016	-0.076***	0.015
Black	-0.127***	0.022	0.073***	0.021	0.126***	0.021
Black x Male	-0.070***	0.033	-0.054*	0.029	-0.056**	0.028
Hispanic	-0.168***	0.026	-0.039	0.025	0.013	0.025
Hispanic x Male	0.029	0.037	0.012	0.035	0.015	0.034
General Sciences	-	-	0.025***	0.012	0.018	0.011
Arithmetic Reasoning	-	-	0.026***	0.012	0.024**	0.012
Mathematical Knowledge	-	-	0.117***	0.012	0.103***	0.011
Assembling Objects	-	-	0.020**	0.009	0.011	0.008
Word Knowledge	-	-	0.003	0.012	-0.015	0.011
Paragraph Comprehension	-	-	0.080***	0.009	0.069***	0.012
Family Covariates	No		No		Yes	
R^2	0.033		0.237		0.300	
Observations	5709		5709		5709	

Table 8: OLS regressions. *, ** and *** indicate significance at 10%, 5% and 1% level, respectively. Robust standard errors are reported. The sample was restricted to those who completed at least eighth grade. This resulted in the loss of only 0.5% of the observations. See Table C1 in appendix C for similar regressions as in tables 5, 7 and 8, where the only difference is that the sample is kept constant across specifications (results do not change). Source: NLSY97.

[57] Armed Forces Vocational Aptitude Battery (ASVAB) subtests: Mathematical Knowledge, Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, Assembling Objects and General Sciences.

[58] Cameron and Heckman (2001) also show that AFQT contributes to explain the racial gap in educational attainment.

4 Factor Model of Educational Attainment

In this section a sequential model of educational attainment is developed. This is coupled with a measurement system for cognitive and non-cognitive factors that provides several advantages relative to conventional OLS regression outcomes. For instance, possible problems of measurement error can be addressed. Moreover, in this model the effects of non-cognitive skills are distinguished from cognitive ones (since, for example, covariates such as GPA are most likely a function of both cognitive and non-cognitive skills) to isolate the effects of these skills at each stage of the schooling career. In addition, this helps the recovery of the disparities in the relative distribution of cognitive and non-cognitive skills between males and females, and provides further evidence about the importance of skills in the gender gap size. The model follows the spirit of the factor model presented in Cameron and Heckman (2001), Heckman, Stixrud and Urzua (2006), and Heckman, Humphries, Urzua and Veramendi (2012). However, the factors, in this paper, are allowed by construction (i.e. after imposing additional structure to the model) to be correlated with a subset of the agent characteristics (i.e. gender and race).[59]

The key points of this strategy are based on two main ideas. First, incorporate into the analysis the fact that schooling attainment is the consequence of previous educational decisions [Cameron and Heckman (1998)]. Second, the presence of two latent factors associated with cognitive and non-cognitive skills is assumed. These are known by each individual and fixed by the time agents start to make their schooling decisions.[60] In addition, these latent (for the econometrician) skills are assumed to be mutually independent where a measurement system will be used for their identification. The independence assumption may sound strong a priori; however, as it is described below, it provides (in the worst case scenario) a lower bound for the effect of non-cognitive skills. Finally, the identification strategy follows Carneiro, Hansen, and Heckman (2003).

The remaining parts of this section are organized as follows: first, a description of the measurement system for the identification of latent factors (i.e. cognitive and non-cognitive skills) is outlined; second, a sequential model of educational attainment is presented; and third, the complete likelihood is shown.

4.1 Measurement System

The following empirical strategy is purely focused on the African American and white subsamples; therefore, the intention is to recover the relative distribution of skills among white females,

[59] Equations (2), (3) and (6) on pages 21 and 24 show how the factors are allowed to be correlated with race and gender.

[60] Even though, factors are not allowed to change over time, they are able to fully explain the gender gap. Heckman, Stixrud and Urzua (2006), Urzua (2008), and Heckman, Humphries, Urzua and Veramendi (2012) impose a similar assumption.

white males, black females, and black males, and their effects on educational attainment. Cognitive and non-cognitive skills of white females are assumed to be:

$$\begin{pmatrix} \theta^C \\ \theta^{NC} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \tag{1}$$

Factors have no scale associated with them, implying that it is not possible to know whether 1 or 100 is a substantial amount of the factor or not, and given that they are not observed, there is no way to know it. Therefore, normalizations of the means and the variances (without loss of generality) set the location and scale of the factors.

In order to recover the mean and standard deviation of the skills (normal) distributions for each subgroup of the population; cognitive (C) and non-cognitive (NC) skills are expressed relative to white females' skills:

$$\begin{aligned} F_i^C &= \theta_i^C + \alpha_1^C male_i + \alpha_2^C male_i * \theta_i^C + \alpha_3^C black_i + \\ &\quad \alpha_4^C black_i * \theta_i^C + \alpha_5^C black_i * male_i + \alpha_6^C black_i * male_i * \theta_i^C \end{aligned} \tag{2}$$

$$\begin{aligned} F_i^{NC} &= \theta_i^{NC} + \alpha_1^{NC} male_i + \alpha_2^{NC} male_i * \theta_i^{NC} + \alpha_3^{NC} black_i + \\ &\quad \alpha_4^{NC} black_i * \theta_i^{NC} + \alpha_5^{NC} black_i * male_i + \alpha_6^{NC} black_i * male_i * \theta_i^{NC} \end{aligned} \tag{3}$$

Therefore, α_1^u and $(1 + \alpha_2^u)$ (where $u = C$ or NC) provide the mean and standard deviation of white males skills relative to white females, α_3^u and $(1 + \alpha_4^u)$ work similarly for black females, and finally $(\alpha_1^u + \alpha_3^u + \alpha_5^u)$ and $(1 + \alpha_2^u + \alpha_4^u + \alpha_6^u)$ for black males. Notice that the coefficient on θ_i^u, where i represents a person, is set to be equal one for identification purposes (i.e. sets the scale). The assumption that the factors are normally distributed contributes to providing a parsimonious model that gives a simple interpretation of the α^u coefficients, and therefore it is easy to compare differences in skills among demographic subgroups.[61]

A linear measurement system is assumed in order to identify the cognitive factor. In this sense, the Armed Forces Vocational Aptitude Battery (ASVAB) tests will be considered as noisy measures of cognitive skills that will help to identify the factor loadings:

$$\left\{ \begin{array}{c} G_{i1} = \gamma_{11}^C + \gamma_{21}^C F_i^C + \boldsymbol{\gamma}_{31}^C \mathbf{X}_i + \varepsilon_{i1}^G \\ \ldots \\ G_{ij} = \gamma_{1j}^C + \gamma_{2j}^C F_i^C + \boldsymbol{\gamma}_{3j}^C \mathbf{X}_i + \varepsilon_{ij}^G \end{array} \right\} \tag{4}$$

[61] Heckman, Stixrud and Urzua (2006) rejects the normality assumption. However, this does not imply that normality of the factors will not hold for a different type of model. For example, Cunha, Heckman, and Schennach (2010) claim that imposing normality, instead of mixture of normals, has only minor effects on the estimates of the technology of skill formation. See Cunha, Heckman, and Schennach (2010) web appendix A.11, page 37. http://www.econometricsociety.org/ecta/Supmat/6551_proofs.pdf

where G_{ij} with $j = 1, ..., 6$ represents the result of agent i in test j (i.e. paragraph comprehension, word knowledge, mathematical knowledge, arithmetic reasoning, general sciences and assembling objects)[62], $F_i^C$the cognitive factor and $\mathbf{X}_i$ a vector of covariates (i.e. family background characteristics and age at the time the test was taken). The fact that not all individuals took the test at the same age may introduce endogeneity problems. For example, Hansen, Heckman, and Mullen (2004) show that an additional year of schooling could increase AFQT scores. However, this problem is likely to be less important in the context of this paper, given that a test of difference in means cannot reject the null hypothesis of equality of highest school grade ever completed between genders at the time the test was taken.[63]

A relatively similar system is considered in order to identify the non-cognitive factor:

$$\left\{ \begin{array}{c} H_{i1} = \gamma_{11}^{CNC} + \gamma_{21}^{CNC} F_i^C + \gamma_{31}^{CNC} F_i^{NC} + \boldsymbol{\gamma}_{41}^{CNC} \mathbf{X}_i + \varepsilon_{i1}^H \\ ... \\ H_{im} = \gamma_{1m}^{CNC} + \gamma_{2m}^{CNC} F_i^C + \gamma_{3m}^{CNC} F_i^{NC} + \boldsymbol{\gamma}_{4m}^{CNC} \mathbf{X}_i + \varepsilon_{im}^H \end{array} \right\} \tag{5}$$

where H_{im} with $m = 1, ..., 5$ represents GPA at grade 8, school retention between grade 1 to 8, suspensions from school (until age 14), involvement in fights with the intention to hurt (until age 14) and precocious sex.[64] In order to control for possible misreport bias, the measure precocious sex includes a dummy for male. $F_i^C$and F_i^{NC} denote the cognitive and non-cognitive factors and $\mathbf{X}_i$ represents a vector of family covariates. It is important to notice that there are no exclusive measures for non-cognitive skills. This implies a literal interpretation of the "non-cognitive" term, where F_i^{NC} will capture all the information in the $H's$ that cannot be explained by cognitive skills.[65] In this regard, the independence assumption of the factors leads (in the worst case scenario) to a lower bound of the effect of non-cognitive skills. Finally, a last normalization is required in order to completely identify the model; the sign of the factors effect needs to be established; hence, the coefficients γ_{21}^C and γ_{31}^{CNC} were set such that more of the factors are "good".[66]

The joint probability of the observed data is assumed to be independent over equations once

[62] See footnote 33 for a discussion on whether ASVAB scores are likely to be gender or racially biased.

[63] Females reported on average 8.91 years of education at the day of the test, while males reported 8.83 years. Therefore, this implies small differences in schooling when taking the exam. In order to provide robustness checks, controls for the highest school grade at the time of the test were also included in other specifications and results did not change.

[64] Precocious sex denotes sexual debut before age 15.

[65] Given that the aim of this paper is to explain the gender gap, the fact that this model is not accounting for constraints is not likely to be a concern (boys and girls on average face the same constraints given that they come from same type of families and attend same type of schools).

[66] An alternative normalization would be to set γ_{21}^C and γ_{31}^{CNC} equal to 1. This normalization would take care of the scale and sign of the effect of the factors. Notice that the normalization implemented in the paper is equally valid as this one.

conditioning on F_i^C, F_i^{NC} and $\mathbf{X}_i$. This implies that the errors (i.e. ε's) are mutually independent. Moreover, the error distributions are assumed normal with mean zero in the case of continuous variables and logit distributed for the binary ones.[67] Thus, this methodology can be characterized as a particular type of matching where the match variables creating conditional independence are not observed by the econometrician.

4.2 Sequential Model of Educational Attainment

Agents make sequential decisions in order to define their final schooling level based on a set of family covariates and latent endowments.[68] In each period males and females have to decide whether to continue their studies, where their choice set is determined by their previous decisions. More specifically, students initially choose whether to finish grade 10. If the student drops out[69] no further decisions are made (i.e. dropping out constitutes an absorbing state)[70,71] if the student finishes grade 10, then he/she has to make an additional decision, that is whether to finish grade 11. This process continues until the last year of college education or until they decide to stop their schooling career.

The latent utility of agent i from making educational choice s is defined as follows:

$$V_{is} = \beta_{0s} + \beta_{1s} F_i^C + \beta_{2s} F_i^{NC} + \boldsymbol{\beta}_{3s} \mathbf{Z}_i + \boldsymbol{\beta}_{4s} \mathbf{X}_i + \varepsilon_{is} \tag{6}$$

where $\mathbf{X}_i$, $F_i^C$and F_i^{NC} have the same definition as above, $\varepsilon_{i,s}$ is the error term which is logit distributed and independent of the regressors, and $\mathbf{Z}_i = \{male, black, black{*}male\}$.[72] The intention is to analyze the statistical significance and the sign of the coefficients on $\mathbf{Z}_i$ (at each educational

[67] Basically, the different configuration of the error terms depending on the type of variable is due to computational issues. Assuming logit distributions speeds up considerably the time of estimation.

[68] As in Cameron and Heckman (2001) wages are not considered in the empirical strategy.

[69] If a student is not enrolled in school for two consecutive periods, then he/she is considered as a drop out. This definition avoids considering as a drop out a student who left school for one period due to a quite specific reason (e.g. health problem).

[70] This assumption has been extensively used in the literature, see for example Arcidiacono (2004) and Stange (2012). Basically, it helps to simplify the estimation of the model. Among those who returned to school, 53% of them were females. This implies that there are not systematic differences on this regard between genders.

[71] The sample used in the sequential model includes all black and white individuals of the NLSY97 that reported their highest completed level of education by age 25, who finished at least eighth grade, and from whom their full schooling history is available. If an individual dropped out school for two consecutive periods then it is assumed that he/she cannot return. It is important to mention that the regression results from Section 3 can also be replicated with this configuration of the data (see Appendix C, Table C2). Notice that the configuration of the sample used in the reduced form strategy (Section 3) is similar to this one but with the main difference that to determine whether an individual has enrolled in college or not it does not matter if he/she dropped out from school for more than two consecutive periods.

[72] $\boldsymbol{\beta}_{3s}$ and $\boldsymbol{\beta}_{4s}$represent vectors of parameters.

level) once the latent factors are incorporated. In this regard, notice that $\mathbf{Z}_i$ and (F_i^C, F_i^{NC}) are allowed to be correlated by construction.

A key feature of this model is the inclusion of the factors controls for the dynamic selection process that occurs during the transitions from one grade to the next. Basically, selection occurs as low ability students leave school in early stages; and hence drop out from the sample. Therefore, it is expected that the distribution of skills shifts to the right with later grades.[73]

The binary outcome variable is defined as:

$$D_{is} = \left\{ \begin{array}{l} 1 \text{ if } V_{is} \geq 0 \\ 0 \text{ otherwise} \end{array} \right\} \tag{7}$$

Therefore, the probability of finishing school level s can be expressed as a logit model:

$$\begin{aligned} \Pr(D_{i,s} &= 1|F_i^C, F_i^{NC}, \mathbf{Z}_i, \mathbf{X}_{i,} D_{i,s-1} = 1) = \Pr(V_{is} \geq 0 \mid F_i^C, F_i^{NC}, \mathbf{Z}_i, \mathbf{X}_{i,}, D_{i,s-1} = 1) \quad (8) \\ &= \frac{\exp\{\beta_{0s} + \beta_{1s} F_i^C + \beta_{2s} F_i^{NC} + \boldsymbol{\beta}_{3s} \mathbf{Z}_i + \boldsymbol{\beta}_{4s} \mathbf{X}_i\}}{\sum \exp\{\beta_{0s} + \beta_{1s} F_i^C + \beta_{2s} F_i^{NC} + \boldsymbol{\beta}_{3s} \mathbf{Z}_i + \boldsymbol{\beta}_{4s} \mathbf{X}_i\}} \end{aligned}$$

where $D_{i,s-1}$ is the past decision taken by agent i. Therefore, the probability of any sequence of life cycle schooling histories can be written as:

$$\prod_{s=1}^{S} [\Pr(D_{i,s}|F_i^C, F_i^{NC}, \mathbf{Z}_i, \mathbf{X}_{i,} D_{i,s-1} = 1)] \tag{9}$$

Notice that any dependence between $D_{i,s}$ and $D_{i,s-1}$ for the same person conditional on $\mathbf{Z}_i$ and $\mathbf{X}_i$, arises from F_i^C and F_i^{NC} (the only exception occurs when $D_{i,s-1} = 0$). Moreover, the schooling decisions are assumed to be independent from the measurement equations once F_i^C, F_i^{NC}, $\mathbf{X}_i$ and $\mathbf{Z}_i$ are included in the estimation. By imposing this assumption, it is possible to obtain a more tractable likelihood function (as it is shown in the following subsection).[74]

4.3 Likelihood

In general, the likelihood could be written as follows:

$$\prod_{i=1}^{N} \Pr(\mathbf{D}_i, \mathbf{G}_i, \mathbf{H}_i|\mathbf{Z}, \mathbf{X}, \mathbf{D}_{-1})$$

[73] Conditional on the initial schooling decision, θ (which is embedded in the $F's$) and the covariates are not independent. In order to understand why; consider the following example: it is expected that youths from very poor families tend to continue schooling only if they show high levels of cognitive and/or non-cognitive skills.

[74] Cunha, Heckman, and Navarro (2007) provide conditions under which these type of models could adequately represent dynamic discrete choice models. Basically, there are two key requirements. First, local comparisons between the rewards of adjacent states locate the global optimum. Second, it is necessary to impose separability between observables and unobservables.

However, given the assumptions of the model,[75] this joint probability is independent over equations once we condition on the unobserved factors. Due to the fact that the factors are not directly observed, it is necessary to integrate them out. Therefore, the complete likelihood can be written as follows:[76]

$$\prod_{i=1}^{N} \int \int \left[\prod_{s=1}^{S} \Pr(D_{i,s}|F_i^C, F_i^{NC}, \mathbf{Z}_i, \mathbf{X}_i, D_{i,s-1}) \right] * \left[\prod_{j=1}^{J} \Pr(G_{i,j}|\mathbf{X}_i, F_i^C) \right] * \tag{10}$$
$$* \left[\prod_{m=1}^{M} \Pr(H_{i,m}|\mathbf{X}_i, F_i^C, F_i^{NC}) \right] f(\theta^C) f(\theta^{NC}) d\theta^C d\theta^{NC}$$

5 Results

The results presented in this section are focused on the African American and white subsamples of the NLSY97.[77] The remaining parts of this subsection are organized as follows: first, results of the educational attainment model are presented; and second, the means and standard deviations of the estimated factors distributions are compared between genders.

5.1 Educational Attainment

Table 9 shows estimation results of the educational attainment model where agents make sequential decisions from grade 10 to (at most) the last year of college.[78] Three main conclusions can be extracted from this table. First, males are no longer less likely to finish high school or enroll in postsecondary education after controlling for the latent factors. Indeed, men are shown to have "higher preferences" for educational attainment.[79,80] This finding is consistent with the empirical regularity that women still spend more time at home,[81] and the fact that the expected benefits for

[75] Two assumptions are key in order to write the likelihood in a simpler way: a) $\varepsilon's$ are independent over equations once conditioning on X, F^C, F^{NC}, and b) cogntive and non-cognitive factors are independent.

[76] Following, Aguirregabiria and Mira (2007), the distributions of the latent factors are considered as a discretized version of Normal$(0,1)$ distributions with $T = 21$ points of support. Alternative numerical integration methods such as the trapezoidal rule or composite Simpson's rule provide similar results.

[77] The estimation outcomes do not include family covariates due to their lack of effect on the gender gap size. Table D1 of appendix D shows that the gender gap results are similar when family covariates are included.

[78] Most students are 16 years old in grade 10, which is the age when they start to make their own schooling decisions. If a student did not complete grade 9 (at any point of his/her life), then it is considered as if he did not complete grade 10.

[79] It is important to emphasize that given boys and girls come on average from the same type of families, then family background covariates are not likely to play a significant role explaining the gap (as it was shown in the previous section).

[80] "Preferences" for schooling refers to the coefficient β_{3s} in equation 6. Notice that β_{3s} may also account for other residual explanations.

[81] While 96% and 89% of college educated and (just) high school graduate males participate in the labor force, 84% and 74% respectively of females do so. These proportions purely comprise white and black subsamples between 24

education attainment continue to be higher for males than females [see Becker et al. (2010)]. In a similar vein, Hubbard (2011) has shown that the college premium for women is not higher than the premium for men once topcoding biases in the CPS survey are corrected.

Second, African Americans show "higher preferences" for college. This result should not be surprising given that wage premium for college educated blacks, conditional on skills, is a distinctive characteristic of the U.S. labor market [Arcidiacono et al. (2010)]. Similarly, Neal (2006) indicates that college educated blacks and whites have comparable wages at the time of initial entry into the labor market, which implies the presence of a substantial black wage premium given the racial differences in average AFQT scores. Finally, Cameron and Heckman (2001) also find this empirical regularity using the male sample of the NSLY79; however, they are not able to distinguish the role of cognitive and non-cognitive skills.

Third, both latent factors are statistically significantly different from zero at each stage of schooling career; however, the relative importance of one skill over the other varies across transitions. Cognitive skills (conditional on reaching a certain grade) have a larger impact on educational completion at any level than non-cognitive ones, especially after finishing high school. For example, Figures 1 and 2 show that the probability of finishing grade 12 for white females (conditional on being enrolled in it) is more responsive to different values of non-cognitive skills than the probability of completing the fourth year of college (conditional on being enrolled in it). More precisely, Figure 1 indicates that young people with quite low levels of cognitive skills still show high probabilities of finishing high school if their levels of non-cognitive skills are high. By contrast, Figure 2 shows that the probability of finishing the fourth year of college is considerably smaller (irrespective of non-cognitive levels) if cognitive skills are very low (see for example, coordinates -3 (cog), 5 (non-cog) in each graph). This finding is consistent with Heckman, Humphries, Urzua and Veramendi (2012) where they show that cognitive skills play a larger role than socio-emotional skills in the decision whether to enroll in college. However, as it is shown in the following section, the substantial disparities in the distribution of non-cognitive skills between males and females make these skills more relevant in terms of explaining the gender gap size across races.

5.2 Skills Distributions

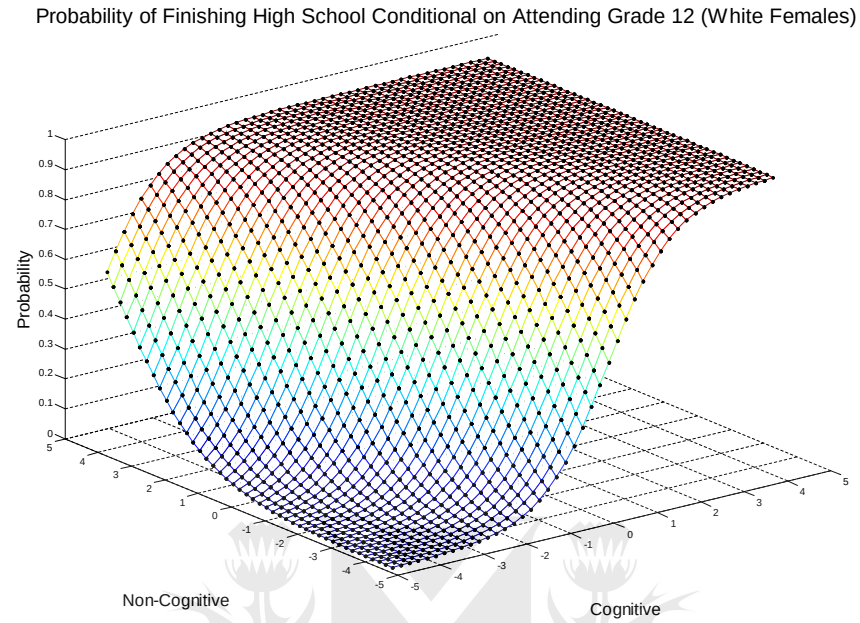
Table 10 presents the means and standard deviations of the estimated distributions of skills for both genders and racial groups.[82] White and black males exhibit lower average skill levels than their female counterparts. However, gender differences in non-cognitive skills are substantially larger than in cognitive ones. In terms of variances, white males exhibit greater dispersion than white females

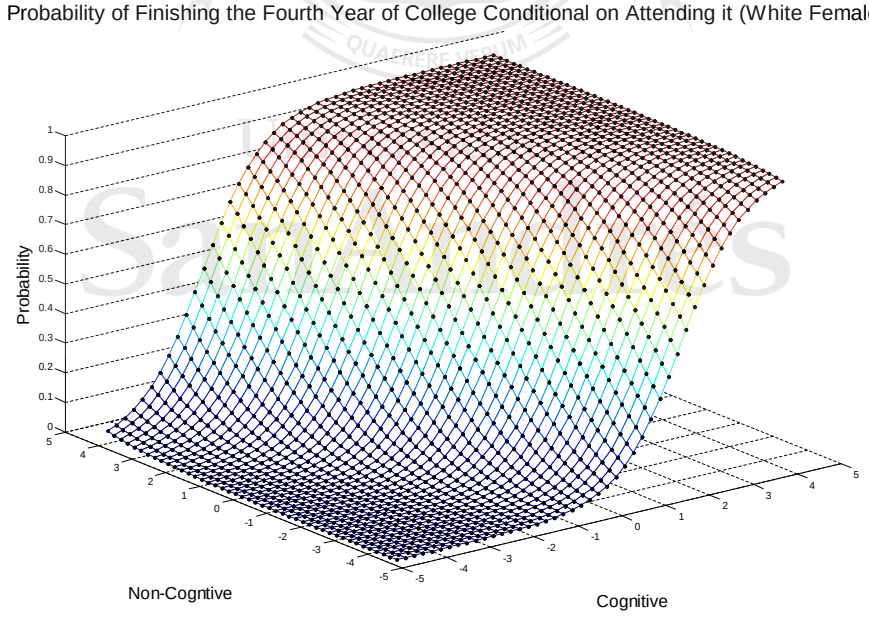
and 50 years old during the period 2000 - 2009. Source: IPUMS-CPS.

[82] The values of the parameters α_1^u to α_6^u (where $u = C$ or NC) and their statistical significance can be found in Table D3 of appendix D. In addition, appendix E shows the values of the measurement system coefficients (i.e. $\gamma's$, eqs. (4) and (5)).

Educational Progression. Full Sample																
	Finish Grade 10		Finish Grade 11		Finish High School		College 1		College 2		College 3		College 4		Finish	
	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
Constant	3.539***	0.123	3.423***	0.122	2.934***	0.106	0.582***	0.067	1.294***	0.088	0.788***	0.101	0.876***	0.134	0.699***	0.151
Male	0.792***	0.157	0.626***	0.169	0.813***	0.161	0.151	0.105	0.222	0.147	0.232	0.160	0.018	0.182	0.253	0.172
Black	1.337***	0.207	0.899***	0.220	0.907***	0.202	1.723***	0.169	0.852***	0.218	1.202***	0.258	0.909***	0.299	0.237	0.280
Black x Male	-0.106	0.247	-0.170	0.263	0.006	0.254	-0.005	0.205	0.554*	0.290	0.430	0.335	0.583	0.399	0.492	0.411
F^C	1.357***	0.080	1.099***	0.084	1.151***	0.082	1.688***	0.077	1.272***	0.103	1.444***	0.126	1.458***	0.159	0.636***	0.135
F^{NC}	0.829***	0.086	0.760***	0.093	0.773***	0.093	0.761***	0.077	0.667***	0.108	0.908***	0.138	0.731***	0.169	0.475***	0.166

Table 9: Logit coefficients of the sequential schooling model. Black and white sample. *, ** and *** indicate significance at 10%, 5% and 1% level, respectively. Source: NLSY97





Figures 1 - 2: White females, probability of finishing 12th grade and 4th year of college (conditional on being enrolled) for different levels of cognitive and non-cognitive skills. Source: NLSY97

	Factors			
	Cognitive		Non-cognitive	
	Mean	Std. Dev.	Mean	Std. Dev.
White Female	0	1	0	1
White Male	-0.067	1.092	-1.003	1.353
Black Female	-1.203	0.957	-0.256	0.849
Black Male	-1.336	0.909	-1.672	1.125

Table 10: Estimated means and standard deviations of cognitive and non-cognitive (normal distributed) factors split by race and gender. Source: NLSY97

on both skill types. It follows that the (right) intersection of their cognitive distributions occurs at the 93rd percentile; while for the non-cognitive distributions it occurs at the 99th percentile. Similarly, African American males exhibit a higher variance in non-cognitive skills than their female counterparts. This evidence is consistent with Hedges et al. (1995), Arden et al. (2006) and Deary (2007) et al. who show that males exhibit higher variance in cognitive skills than females.

In Table 10 it is also apparent that disparities in skills between black males and females are larger than those between white males and females. For instance, while the difference in non-cognitive skills between whites is around one standard deviation (i.e. 1.003) of the white female distribution, for blacks the difference (based on the same scale as whites) is 1.416.[83] In this sense, simulation exercises (as it is shown in the following section) indicate that if disparities in skills between black males and females mirrored (in size) the disparities between white men and women, then the size of the gender gap in college enrollment would be the same for both races.

To sum up, results from Tables 9 and 10 suggest that large differences in non-cognitive skills between males and females play a key role in shaping the current gender gaps in educational attainment. Moreover, a possible explanation for the larger gender gap among blacks could be due to the finding that disparities in skills between African American boys and girls are larger than between their white counterparts. In light of this, the following section analyzes these points in detail.

6 Implications of the Model

The sequential model makes it possible to analyze the full schooling career profile of males and females from multiple perspectives. In order to provide a baseline picture, panel A of Table 11 shows the total sample proportion of girls and boys (split by racial group) that complete the different

[83] Being -1.416 statistically significant higher than -1.003. Table D3 of Appendix D shows the statistical significance of the different parameters used to calculate the mean and standard deviation of the different demographic groups.

educational levels.[84] Women (conditional on race) constitute the majority in every schooling year, and their overrepresentation is increasing every year. The fact that a substantial proportion of black males drop out of high school indicates that gender disparities in college enrollment are, in part, a consequence of a significant number of boys not even completing the necessary steps to attend college.

It has been shown in Table 9 that males show "higher preferences" for educational attainment than females after controlling for skills differences. This implies that the gender gap in college enrollment would be even larger if boys had the same preferences as girls. In order to quantify this, panel B of Table 11 displays the educational attainment of men after imposing female preferences on them. Results indicate that the percentage of black and white men enrolled in college would be only 22% and 39% respectively. In this case, gender disparities for whites would increase from 10% to 15%, and for African Americans from 17% to 22%.

The relative importance of one skill over the other is relevant in terms of policy recommendations. For instance, if a policy intends to close gender disparities in educational attainment, then establishing the importance of each skill matters, given that non-cognitive skills are more malleable than cognitive ones [Cunha et al. (2005)]. Panel C of Table 11 shows the proportion of males that would finish each educational level if one of the factors is increased by one standard deviation.[85] The results indicate that such an increase in cognitive skills has a higher impact on the probability of finishing a given grade than a similar increase in non-cognitive skills. In addition, cognitive skills become more relevant for college enrollment than non-cognitive ones. For instance, a one standard deviation increase in males' non-cognitive skills would improve the proportion of white males enrolled in college from 0.44 to 0.60, and from 0.27 to 0.41 for black males. However, a similar increase in cognitive skills would lead to proportions of 0.70 and 0.51 of white and black males respectively.

These results do not imply that non-cognitive skills are not important explanations for the gender disparities in educational attainment. On the contrary, the fact that boys and girls have larger differences in these skills than in cognitive skills turns out to be a more relevant explanation of the gender gap. In order to show this result, panel D of Table 11 presents the unconditional probabilities of finishing each schooling level for white and black males after imposing female skills distributions on them. For instance, if black males had black females' non-cognitive distribution,

[84] The proportion of students enrolled in college is smaller than the one presented in table 1. See footnotes 69, 70, and 71 for an explanation on this point. However, notice, that the size of the gender gap does not change when using this sample. Table C2 in Appendix C shows that reduced form results (presented in Section 3) do not change when using the sample used in the factor model of educational attainment. Finally, Figure D1 of appendix D shows how well the model fits the data.

[85] For example, calculating the proportion of white males that would have finished each educational level if their cognitive skills were increased by one standard deviation, while keeping their non-cognitive skills in their original level; and vice versa.

82% would have graduated from high school and 45% would have enrolled in college. However, the percentage of black men finishing grade 12 and attending postsecondary education would only be 65% and 31% respectively, if instead they had the cognitive distribution of black women.[86]

To sum up, the results indicate that the observed gap in college enrollment is not only a consequence of a significant number of boys deciding not to enroll in college after finishing high school; a significant proportion leave the system before graduating from high school, in part due to low levels of non-cognitive skills. Moreover, large differences in non-cognitive skills between black males and females appear to play a key role in explaining the size of the gap for this racial group.

6.1 Cross-Racial Differences

As was described earlier, gender disparities in college enrollment are much larger among blacks than among any other racial group. In this sense, it is appropriate to investigate whether this empirical regularity can be explained by the fact that African Americans exhibit higher gender differences in average skill levels (see Table 10) than whites. Three simulation exercises have been performed with this aim. First, college enrollment was simulated under the assumption that average differences in cognitive skills between black males and females are similar to white gender differences. More specifically, this implies turning off the coefficient α_5^C from equation 2. The second simulation repeats this same procedure but this time with non-cognitive skills (i.e. turning off $\alpha_5^{NC}$from equation 3). Finally, the last simulation turns off both coefficients α_5^{NC} and α_5^C. Table 12 shows the differences between the white and black gender gap in each of the described scenarios (i.e. white gap - black gap). Results indicate that more than 70% of the "additional" gap observed among African Americans can be explained by larger gender differences in non-cognitive skills than whites, while less than 30% of the "additional" gap is explained by higher differences in cognitive skills. Therefore, non-cognitive skills seem to be crucial in explaining the greater gender disparity among African Americans than among whites, in terms of educational attainment.

6.2 Oaxaca's Decompositions

An alternative approach to aid understanding of the gender gap in educational attainment is to perform a sequence of Oaxaca's decompositions for each level of schooling career. These decompositions will show in more detail how differences in skill levels and preferences contribute to the gaps. For example, results in panel B1 of Table 13 indicate that if gender differences in preferences were eliminated, then equalizing the cognitive and non-cognitive levels of African American males and females would close a gap of 0.221. This would result in the sign of the gender gap reversing (i.e. more males than females would attend college). In a similar vein, this

[86] Notice that the proportions of black females that finished high school and enrolled in college are 74% and 44%, respectively.

	Educational Attainment							
	Grade 10	Grade 11	Grade 12	College 1	College 2	College 3	College 4	Finish
Panel A: Baseline Model								
Males and Females in Each Grade as Proportion of Total Demographic Subsample								
White Female	0.93	0.88	0.82	0.54	0.45	0.37	0.32	0.25
White Male	0.91	0.85	0.78	0.44	0.36	0.28	0.23	0.18
Black Female	0.91	0.83	0.74	0.44	0.33	0.24	0.18	0.13
Black Male	0.85	0.73	0.63	0.27	0.21	0.14	0.10	0.08
Percentage of Males in Each Grade Conditional on Race								
White Male	49.4%	49.0%	48.9%	45%	44.3%	43%	42%	42.2%
Black Male	48.2%	46.6%	45.9%	38.1%	38.4%	36.5%	35.8%	37.5%
Panel B: Males with Female Preferences for Educ. Attainment								
Males in Each Grade as Proportion of Total Demographic Subsample								
White Male	0.85	0.76	0.67	0.39	0.32	0.24	0.20	0.16
Black Male	0.77	0.63	0.49	0.22	0.15	0.09	0.06	0.04
Panel C: One Standard Deviation Increase in Skills								
Males in Each Grade as Proportion of Total Demographic Subsample								
White Male Cognitive	0.97	0.94	0.91	0.70	0.63	0.54	0.48	0.40
White Male Non-cognitive	0.96	0.93	0.89	0.60	0.53	0.45	0.39	0.33
Black Male Cognitive	0.94	0.87	0.80	0.51	0.43	0.34	0.28	0.22
Black Male Non-cognitive	0.92	0.85	0.77	0.41	0.34	0.26	0.21	0.16
Panel D: Males with Female Skills Distributions								
Males in Each Grade as Proportion of Total Demographic Subsample								
White Male Cognitive	0.92	0.86	0.80	0.45	0.37	0.29	0.23	0.18
White Male Non-cognitive	0.96	0.92	0.88	0.56	0.48	0.40	0.34	0.28
Black Male Cognitive	0.86	0.75	0.65	0.31	0.24	0.17	0.13	0.10
Black Male Non-cognitive	0.94	0.88	0.82	0.45	0.38	0.30	0.23	0.18

Table 11: Estimated educational attainment split by gender and race. The percentage of males in each grade conditional on race (see bottom rows of Panel A) refers to the number of males of race r (i.e. black or white) in grade g divided by the total number of race r youths (males plus females) in that grade. Source: NLSY97.

College Enrollment Gender Gap Across Races	
	White Gap - Black Gap
Actual Difference	-0.07
Simulation Cognitive Skills ($\alpha_5^C = 0$)	-0.05
Simulation Non-cognitive Skills ($\alpha_5^{NC} = 0$)	-0.02
Simulation Both Skills ($\alpha_5^{NC} = \alpha_5^C = 0$)	0

Table 12: Differences between white and black gender gaps *in* counterfactual scenarios. $\alpha_5^u = 0$ (where $u = C$ or NC) implies that the mean differences in cognitive or non-cognitive skills between black males and females are similar to white gender differences. Source: NLSY97.

Table also illustrates that males have "higher preferences" for educational attainment than females. For instance in the absence of differences in skills between genders, the proportion of white males finishing the first year of college would be 4.7% higher than the proportion of white females.

Panels A2 and B2 of Table 13 indicate that most of the gap attributable to skills differences is mainly explained by disparities in non-cognitive skills. If preferences for educational attainment were the same across genders, then more than 80% of the gap attributable to skills differences would be explained by differences in non-cognitive skills (for both races).[87] Therefore, Table 13 provides a neat picture of the relevance that non-cognitive skills have in explaining the gender gap within and across racial groups.

6.3 Dynamic Selection Process: Factors Distribution

Finally, Table 14 shows the evolution of cognitive and non-cognitive skills as a consequence of the dynamic selection process that occurs at each schooling level (i.e. low skill students leave school at early stages and hence drop out from the sample). The intention is to analyze if the selection process is mainly driven by a particular type of skill. Results indicate a substantial increase in the mean of the cognitive skills distribution between the end of high school and the last year of college. In addition, white males show a higher mean of cognitive abilities than females by the end of schooling career (despite them starting behind). In relation to non-cognitive skills, black males show important shifts between grade 9 and the end of college. The size of this change is substantially larger than the one experienced by white or black females (i.e. more than double). To sum up, these results indicate that the dynamic selection process is operating intensively on both factors, though the selection process in terms of cognitive skills is more aggressive after high school.

[87] Appendix F shows Oaxaca decompositions using males covariates as base.

	Oaxaca's Decompositions: Unconditional Probability (Base: Females Covariates)							
	Grade 10	Grade 11	Finish HS	College 1	College 2	College 3	College 4	Finish
Panel A1: Whites								
Actual Gap	-0.023	-0.035	-0.036	-0.097	-0.093	-0.091	-0.087	-0.067
Gap due to Preferences	0.033	0.05	0.077	0.047	0.047	0.045	0.034	0.037
Gap due to Skills Diff.	-0.056	-0.085	-0.113	-0.144	-0.140	-0.136	-0.121	-0.104
Panel A2: Contributions to *Gap due to Skills Differences*								
Cognitive	13%	12%	12%	15%	13%	10%	9%	7%
Non-cognitive	87%	88%	88%	86%	87%	90%	91%	93%
Panel B1: Blacks								
Gender Gap	-0.062	-0.106	-0.113	-0.169	-0.125	-0.104	-0.082	-0.051
Gap due to Preferences	0.039	0.057	0.095	0.052	0.084	0.089	0.083	0.088
Gap due to Skills Diff.	-0.101	-0.163	-0.208	-0.221	-0.209	-0.193	-0.165	-0.139
Panel B2: Contributions to *Gap due to Skills Differences*								
Cognitive	6%	7%	8%	17%	19%	20%	22%	22%
Non-cognitive	94%	93%	92%	83%	81%	80%	77%	77%

Table 13: Oaxaca's decompositions for unconditional probabilities of finishing different stages of schooling career. Black and white samples were considered separately. Female covariates are used as base. Source: NLSY97.

Dynamic Selection Process: Factor Means				
	Grade 9	Grade 12	College 1	College 4
	Cognitive Skills			
White Female	0	0.182	0.499	0.796
White Male	-0.067	0.170	0.603	0.953
Black Female	-1.203	-0.966	-0.622	-0.213
Black Male	-1.336	-1.039	-0.615	-0.238
	Non-Cognitive Skills			
White Female	0	0.117	0.245	0.424
White Male	-1.003	-0.772	-0.512	-0.187
Black Female	-0.256	-0.131	-0.026	0.168
Black Male	-1.672	-1.375	-1.135	-0.782

Table 14: Dynamic selection process mean of cognitive and non-cognitive skills at selected grades split by gender and race. Source: NLSY97.

To conclude, the results presented in Sections 5 and 6 have shown that the skills needed for success in the schooling career are multiple in nature. Basically, cognitive and non-cognitive skills are both important to explain disparities in educational attainment. However, non-cognitive skills play a key role explaining cross-racial differences in the educational gender gap. In particular, black males exhibit substantially lower levels of non-cognitive skills than black females. Therefore, policies that attempt to foster this type of skill on specific groups of the population (e.g. black males) could lead to significant reductions in current gender inequalities.

7 Conclusions

The sizable gender gap in college enrollment, especially among African Americans, constitutes an empirical regularity that may have serious implications for family formation, parenting, and male labor outcomes. Regressions results indicate that family background covariates cannot fully explain the gender gap. However, the estimation of a sequential model for educational attainment indicates that disparities in cognitive and non-cognitive skills more than explain the gender differences. Indeed, males are shown to have "higher preferences" for schooling than females after controlling for the latent factors.

Cognitive skills exhibit a higher effect (conditional on reaching certain grade) on transitions from one schooling level to the next than non-cognitive ones, especially for college enrollment. However, the substantial disparities in the distribution of non-cognitive skills between males and females make these skills crucial factors in explaining the gender gap size within races. Moreover,

the observed gap in college enrollment is not only a consequence of a significant number of boys deciding not to enroll in college after finishing high school; a large proportion drop out of high school due to low levels of non-cognitive skills. The large gap in educational attainment between African American males and females is mainly explained by the substantial gender differences in non-cognitive skills. Simulation exercises show that if black gender disparities in skills mirrored white ones, then the size of the gap would be the same for both races.

To conclude, a small but growing body of early childhood intervention studies has shown that non-cognitive skills can be improved over long periods of time in response to interventions [Almlund, Duckworth, Heckman and Kautz (2011)]. In this regard, the findings of this manuscript in conjunction with the fact that non-cognitive skills seem to be more malleable than cognitive ones [Cunha, Heckman and Schennach (2010)], suggest that initiatives which aim to foster the non-cognitive skills of males, and in particular black males, could effectively reduce current inequalities in educational attainment. In this sense, the promotion of non-cognitive skills for particular subgroups of the population is an important avenue for policy, and one which requires far greater attention.

References

[1] Acemoglu, D. and Autor, D. (2011), Skills, tasks, and technologies: Implications for employment and earnings, in O. Ashenfelter and D. Card (eds.), Handbook of Labor Economics 4b, Elsevier B.V., Amsterdam.

[2] Aguirregabiria, Victor; and Mira, Pedro (2007). "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75(1), pp. 1-53.

[3] Almlund Mathilde, Duckworth Angela, Heckman James and Tim Kautz (2011). "Personality Psychology and Economics". *Handbook of the Economics of Education*, Volume 4, Chapter 1.

[4] Arcidiacono, Peter (2004). "Ability Sorting and the Returns to College Major," Journal of Econometrics, Vol. 121, Nos. 1-2, 343-375

[5] Arcidiacono, Peter; Bayer, Patrick; and Hizmo, Aurel (2010). "Beyond Signaling and Human Capital: Education and the Revelation of Ability." *AEJ: Applied Economics*, Vol. 2, No. 4 (October 2010), 76-104 .

[6] Arden, Rosalind and Plomin, Robert (2006). "Sex differences in variance of intelligence across childhood." *Personality and Individual Differences,* 41, 39–48

[7] Armour, Stacy and Dana Haynie. "Sexual Debut and Delinquency." *Journal of Youth and Adolescence*, 36:141-152.

[8] Autor, D., and Wasserman, M. (2013). "Wayward Sons: The Emerging Gender Gap in Labor Markets and Education," Third Way, March

[9] Averett, Susan L., Mark L. Burton (1996). "College Attendance and the College Wage Premium: Differences by Gender" *Economics of Education Review*, 15(1): 37–49.

[10] Baumrind, Diana (1991). Effective parenting during the early adolescent transition. In P.A. Cowan & E. M. Hetherington (Eds.), Advances in family research (Vol. 2). Hillsdale, NJ: Erlbaum.

[11] Becker Gary, Hubbard William, Murphy Kevin (2010). "Explaining The Worldwide Boom in Higher Education of Women" *Journal of Human Capital* 4(3):203-241.

[12] Bertrand, M. (2010). "New Perspectives on Gender" *Handbook of Labor Economics,* Volume 4b, Chapter 17, Elsevier Ltd.

[13] Bertrand, M. and Pan, Jessica (2011). "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior" NBER Working Paper No. 17541

[14] Brody, L.M., Nagin, D.E., Tremblay, R.E., Bates, J.E., Brame, B., Dodge, K.A. (2003). "Developing trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study." *Developmental Psychology*, Vol. 39(2), Pages 222-245.

[15] Cameron Stephen, Heckman James (2001). "The Dynamics of Educational Attainment for Black, Hispanic, and White Males." *Journal of Political Economy*, vol. 109(3), pages 455-499, June.

[16] Cameron Stephen, Heckman James (1998). "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males" *Journal of Political Economy*, vol. 106(2), pages 262-333.

[17] Carneiro Pedro, Hansen Karsten, Heckman James (2003). "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." Working Paper Series 2003:9, IFAU - Institute for Labour Market Policy Evaluation.

[18] Carneiro, P., Heckman, J., and Masterov, D. (2005). "Understanding the Sources of Ethnic and Racial Wage Gaps and Their Implications for Policy," in L. B. Nielsen and R. L. Nelson: Handbook of Employment Discrimination Research. Dordrecht: Springer, 2005, pp. 99-136

[19] Carneiro, P., Heckman, J., and Masterov, D. (2005). "Labor Market Discrimination and Racial Differences in Premarket Factors," *Journal of Law and Economics*, Vol. 48, No. 1, pp. 1-39

[20] Caspi Avshalom, Shiner Rebecca (2006). "Personality Development." In: W. Damon, & R. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of Child Psychology: Vol.3*. Social, Emotional, and Personality Development (6th ed.). New York: Wiley.

[21] Caspi, A.; Henry, B.; McGee, R.; Moffit, T; Silva, P. (1995). "Temperamental Origins of Child and Adolescent Behavior Problems: From Age Three to Age Fifteen." *Child Development*, Vol(66), Pages: 55-68.

[22] Cavazos-Rehg PA, Krauss MJ, Spitznagel EL, Schootman M, Bucholz KK, Peipert JF, Sanders-Thompson V, Cottler LB, Bierut LJ (2009). "Age of sexual debut among US adolescents." *Contraception* 80, Pages: 158–162

[23] Cawley John, Heckman James, Vytlacil Edward (2001). "Three Observations on Wages and Measured Cognitive Ability." *Labour Economics* 8(4), 419-442.

[24] Cho, Donghun (2007). "The Role of High School Performance in Explaining Women's Rising College Enrollment." *Economics of Education Review*, Volume 26, Issue 4, Pages 450-462.

[25] Cooley, J., Navarro, S., and Takahashi, Y. (2013). "How The Timing of Grade Retention Affects Outcomes: Identification and Estimation of Time-Varying Treatment Effects" Working Paper The University of Western Ontario

[26] Cornwell, C., Mustard, D. and Van Parys, J. (2012). "Non-cognitive Skills and Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School" Forthcoming *Journal of Human Resources*

[27] Cunha Flavio, Heckman James (2008). "Formulating, Identifying and Estimating the Technology of Cognitive and Non-cognitive Skill Formation." *Journal of Human Resources*, vol. 43(4).

[28] Cunha Flavio, Heckman James, Lance Lochner, Dimitriy V. Masterov (2005). "Interpreting the Evidence on Life Cycle Skill Formation." Discussion Paper Series, IZA DP No. 1675.

[29] Cunha Flavio, Heckman James, Navarro Salvador (2005). "Separating Uncertainty from Heterogeneity in Life Cycle Earnings." *Oxford Economic Papers*, Oxford University Press, vol. 57(2), pages 191-261, April.

[30] Cunha, F., Heckman, J., and Navarro, S. (2007). "The Identification and Economic Content of Ordered Choice Models with Stochastic Thresholds", *International Economic Review*, 48(4)

[31] Cunha, F., Heckman, J., and Schennach, S. (2010). "Estimating the Technology of cognitive and Noncognitive Skill Formation," *Econometrica*, May 1; 78(3): 883–931. doi:10.3982/ECTA6551

[32] Currie Janet, Stabile Mark (2004). "Child Mental Health and Human Capital Accumulation the Case of ADHD." NBER, Working Paper 10435.

[33] Deary, I.; Irwing, P.; Der, G.; Bates, T. (2007). "Brother–sister differences in the g factor in intelligence: Analysis of full, opposite-sex siblings from the NLSY1979". *Intelligence*, vol. 35, no5, pp. 451-456

[34] Digman John (1990). "Personality Structure: Emergence of the Five Factor Model." *Annual Review of Psychology*, 41, 417-440.

[35] Domitrovich, Celene (2008). "Social-Emotional Skills in Early Childhood: Implications for Educational Standards." Penn State Prevention Research Center, Texas Pre-Kindergarten Standards Summit.

[36] Duckworth, A., and S. Urzua (2009). "Determinants of Success in Early Adulthood: Comparing the Effects of Intelligence and Big Five Personality Traits," Unpublished manuscript, University of Pennsylvania, Department of Psychology.

[37] Duckworth, A., and Seligman, M. (2005). "Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents," *Psychological Science* doi: 10.1111/j.1467-9280.2005.01641.x, December, vol. 16 no. 12 939-944

[38] Duncan, G.J., Magnuson, K. (2010). "The Nature and Impact of Early Achievement Skills, Attention and Behavior Problems" Working paper 2010 at the Department of Education, UC Irvine.

[39] Ensminger Margaret (1990). "Sexual Activity and Problem Behaviors Among Black, Urban Adolescents." *Child Development*. Vol 61 pp. 2032–2046

[40] Ferguson Ronald (2005). "Why America's Black-White School Achievement Gap Persists." *Ethnicity, Social Mobility, and Public Policy.* eds. Glenn C. Loury, Tariq Modood and Steven Teles. New York, NY: Cambridge University Press, 2005.

[41] Fergusson. D.M., and Horwood, L.J., (1998). "Early Conduct Problems and Later Life Opportunities." *J. Child Psychol. Psychiatry*, 39(8):1097–108.

[42] Fortin, M., Oreopoulos, P. and Phipps, S. (2012). "Leaving Boys Behind: Gender Disparities in High Academic Achievement", Working Paper University of British Columbia

[43] Freeman, Catherine (2004). "Trends in Educational Equity of Girls & Women: 2004." U.S. Department of Education, Institute of Education Sciences, (NCES 2005-016).

[44] Frierson, Herny; Pearson Willie; and Wyche, James. (2009). Black American Males in Higher Education: Diminishing Proportions. Diversity in Higher Education. Volume 6. Emerald Group Publishing Limited. First Edition.

[45] Fryer, Roland (2010). "Racial Inequality in the 21st Century: The Declining Significance of Discrimination" NBER Working Paper No. 16256.

[46] Fumia, D. (2013). "Criminal Sanctions and Racial Gaps in Educational Attainment," Ph.D. Dissertation, University of Washington, Daniel J. Evans School of Public Affairs

[47] Goldammer, C. (2012). "Racial Gaps and Noncognitive skills: The Asian Exception," Working Paper University of Chicago

[48] Claudia Goldin, Lawrence F. Katz, Ilyana Kuziemko (2006). "The Homecoming of American College Women: The Reversal of the College Gender Gap." *The Journal of Economic Perspectives*, Vol. 20, No. 4, pp. 133-156.

[49] Gilliam Walter (2005). "Prekindergarteners Left Behind: Expulsion Rates in State Prekindergarten Programs." Foundation for Child Development, http://www.fcd-us.org/PDFs/NationalPreKExpulsionPaper03.02_new.pdf

[50] Gullone Eleonora, Moore Susan (2000). "Adolescent Risk Taking and the Five Factor Model of Personality." *Journal of Adolescence*, 23(4), 393-407.

[51] Harris Linda, Blum Robert, Resnick Michael (1991). "Teen Females in Minnesota: A portrait of Quiet Disturbance." *Women and Therapy*, 11 (3-4), 119-135.

[52] Hauser, Robert (1999). "Should We End Social Promotion? Truth and Consequences." Center for Demography and Ecology, University of Wisconsin-Madison, Working Paper No. 99-06

[53] Hausmann Ricardo, Tyson Laura, Zhidi Saadia (2009). "The Global Gender Gap Report." World Economic Forum.

[54] Heckman James, Rubinstein Yona (2001). "The Importance of Non-cognitive Skills: Lessons from the GED Testing Program." *American Economic Review*, 91, 145-149

[55] Heckman James, Lochner Lance, Todd Petra (2006). "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond." *Handbook of the Economics of Education*, Elsevier.

[56] Heckman James, Stixrud Jora, Urzua Sergio (2006). "The Effects of Cognitive and Non-cognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics*, vol. 24(3), pages 411-482, July.

[57] Heckman, J., and Navarro, S. (2007). "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, Volume 136, Issue 2, February 2007, Pages 341–396

[58] Heckman James, Rodrigo Pinto, Hongbo Wang (2008). "Determinants of Risk Behaviors in Early Adulthood: Interplay of Cognitive Ability, Non-cognitive Skills, and Health." Working Paper, University of Chicago.

[59] Heckman, J. (2011). "The American Family in Black and White: A Post-Racial Strategy for Improving Skills to Promote Equality," NBER, Working Paper 16841

[60] Heckman James, Humphries, John, Urzua Sergio, Veramendi Gregory (2012). "The Effects of Educational Choices on Labor Market, Health, and Social Outcomes." Working Paper, University of Chicago.

[61] Hedges, LV; Nowell, A (1995). "Sex differences in mental test scores, variability, and numbers of high-scoring individuals". *Science*, 269, 41–5

[62] Hoff Sommers, Christina (2000). The War Against Boys: How Misguided Feminism is Harming Our Young Boys. Simon and Shuster

[63] Holt Emily, McGrath Daniel, Herring Willian (2007). "Timing and Duration of Student Participation in Special Education in the Primary Grades." U.S. Department of Education, (NCES 2007–043).

[64] Hubbard, William (2011). "The Phantom Gender Difference in the College Wage Premium." *Journal of Human Resources* 46(3): 568–586

[65] Jacob, Brian (2002). "Where the Boys Aren't: Non-cognitive Skills, Returns to School and the Gender Gap in Higher Education." *Economics of Education Review*, December, 21:6, pp. 589-98.

[66] Jencks Christopher, Philips Meredith (1998). *The Black-Whites Test Score Gap.* Washington, D.C. Brookings Institution Press, 1998.

[67] Judge Timothy, Higgins Chad, Thoresen, Carl, Barrick Murray (1999). "The Big Five Personality Traits, General Mental Ability, and Career Success Across the Life Span." *Personnel Psychology*, 52, 621–652.

[68] Kass, R. A., Mitchell, K. J., Grafton, F. C., & Wing, H. (1983). "Factorial validity of the Armed Services Vocational Aptitude Battery (ASVAB), Forms 8, 9, and 10: 1981 army applicant sample". *Educational and Psychological Measurement*, 43, 1077-1087.

[69] Kerwin Charles , Luoh Ming-Ching (2003). "Gender Differences in Completed Schooling." *The Review of Economics and Statistics*, Vol. 85, No. 3, pp. 559-577

[70] Kewal Ramani, Gilbertson, Fox, Provasnik (2007). "Status and Trends in the Education of Racial an Ethnic Minorities." U.S. Department of Education, Institute of Education Sciences, (NCES 2007-039).

[71] King, Jacqueline (2010). "Gender Equity in Higher Education: 2010." American Council on Education Center for Policy Analysis.

[72] Judith, Kleinfeld (2006). "Five Powerful Strategies for Connecting Boys to Schools." Paper for White House Conference on Helping America's Youth Indianapolis, Indiana

[73] Ladd, G.W., & Burgess, K.B. (1999). "Charting the relationship trajectories of aggressive, withdrawn, and aggressive-withdrawn children during early grade school." *Child Development*, Vol. 70(45), Pages 910-929.

[74] Lavigne,J., Gibbons, R., Christofel, K; Arend, R. et al. (1996). "Prevalence Rates and Correlates of Psychiatric Disorders among Preschool Children." *Journal of the American Academy of Child and Adolescent Psychiatry*, Vol. 35(2), 204-214.

[75] Lerner Richard, Steinberg Laurence (2004). *Handbook of Adolescent Psychology,* (Second Edition). Hoboken, N.J.: John Wiley & Sons.

[76] Losen, Dan and Orfield, Gary (2002). *Racial Inequity in Special Education.* Harvard Education Publishing Group.

[77] Martins, P. (2010). "Can Targeted, Non-Cognitive Skills Programs Improve Achievement? Evidence from EPIS," IZA DP No. 5266

[78] McDaniel, A., DiPrete, T., Buchmann. C., and Shwed, U. (2011). "The Black Gender Gap in Educational Attainment: Historical Trends and Racial Comparisons" *Demography*, 48:889–914

[79] Nagin, D.S., & Tremblay, R.E. (2001). "Parental and early childhood predictors of persistent physical aggression in boys from kindergarten to high school." *Archives of General Psychiatry*, 58(4), 389-394.

[80] Neal Derek, Johnson William (1996). "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy*, 104(5), pp. 869-95

[81] Neal Derek (2004). "The Measured Black-White Wage Gap among Women Is Too Small." *Journal of Political Economy*, 2004, 112(1), pp. S1-28.

[82] Neal, Derek. 2006. "Why Has Black-White Skill Convergence Stopped?". *Handbook of Economics of Education I*, eds. Eric Hanushek and Finis Welch, North Holland.

[83] NCES, Digest of Education Statistics (2005). U.S. Department of Education, National Center for Education Statistics.

[84] NCES, Integrated Postsecondary Education Data System (2003). U.S. Department of Education, National Center for Education Statistics.

[85] Planty, M., Hussar, W., Snyder, T., Kena, G., KewalRamani, A., Kemp, J., Bianco, K., Dinkes, R. (2009). "The Condition of Education 2009" (NCES 2009-081). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

[86] OECD (2010), "PISA 2009 Results: Overcoming Social Background – Equity in Learning Opportunities and Outcomes" (Volume II) http://dx.doi.org/10.1787/9789264091504-en

[87] Owings, W. and Magliaro, S. (1998). "Grade Retention: A History of Failure." ERIC Clearing House. Washington, DC: U.S. Department of Education.

[88] Payton, J.; Wardlaw, D.; Graczyk, P., Blooddworth, M.; Tompsett, C.; Weissberg, R. (2000). "Social and Emotional Learning: a framework for mental health and reducing risk behavior in children and youth." *Journal of School Health*, Vol. 70, Pages 179-185.

[89] Peter Katharin, Horn Laura. (2005). "Gender Differences in Participation and Completion of Undergraduate Education and How They Have Changed Over Time." U.S. Department of Education, (NCES 2005–169).

[90] Petersen Anne, Richmond Julius, Leffert Nancy (1993). "Social Changes Among Youth: The United States Experience." *Journal of Adolescent Health*, Volume 14, Issue 8, Pages 632-637.

[91] Perkins Robert, Brian Kleiner, Stephen Roey (2004). "The High School Transcript Study: A Decade of Change in Curricula and Achievement, 1990-2000." U.S. Department of Education, Institute of Education Sciences, NCES 2004-455.

[92] Planty, Hussar, Snyder, Kena, Kewal Ramani, Kemp, Bianco, Dinkes (2009). "The Condition of Education 2009." U.S. Department of Education, Institute of Education Sciences, (NCES 2009-081).

[93] Poropat, A. (2009). "A meta-analysis of the five-factor model of personality and academic performance," *Psychological Bulletin*, Mar; 135(2):322-38. doi: 10.1037/a0014996.

[94] Rampey, Dion, Donahue (2009). "NAEP 2008 Trends in Academic Progress." U.S. Department of Education, Institute of Education Sciences, (NCES 2009–479).

[95] Ree, M. J., & Carretta, T. R. (1995). "Group differences in aptitude factor structure on the ASVAB." *Educational and Psychological Measurement*, 55(2), 268-277.

[96] Reynolds, John, and Jennifer Pemberton (2001). "Rising College Expectations among Youth in the United States: A Comparison of the 1979 and 1997 NLSY." *The Journal of Human Resources*, Vol. 36, No. 4, Special Issue on Early Results from the National Longitudinal Survey of Youth, 1997 Cohort, pp. 703-726.

[97] Rodgers III, W., and Spriggs W. (1996). "What does the AFQT really measure: Race, wages, schooling and the AFQT score" *Review of Black Political Economy*; 24, 4; ABI/INFORM Global, pg. 13

[98] Sandberg Seija (1996). Hyperactivity and Attention Disorders of Childhood. Cambridge University Press.

[99] Segal Carmit (2011). "Misbehavior, Education, and Labor Market Outcomes." *The Journal of the European Economic Association*, Forthcoming..

[100] Schneider Helen, Eisenberg Daniel (2006). "Who Receives a Diagnosis of Attention-Deficit/Hyperactivity Disorder in the United States Elementary School Population?" *Pediatrics*, Vol 117(4), Pages 601-609.

[101] Sharpe, R., and Darity, W. (2009), Where are the brothers? Alternatives to four-year college for Black males, in Henry T. Frierson, Willie Pearson, James H. Wyche (ed.) Black American Males in Higher Education: Diminishing Proportions (Diversity in Higher Education, Volume 6), Emerald Group Publishing Limited, pp.135-153

[102] Shaw, D., Owens, E., Giovannelli, J. & Winslow, E. (2001). "Infant and toddler pathways leading to early externalizing disorders." *Journal of the American Academy of Child & Adolescent Psychiatry*, Vol. 40, 36-43.

[103] Singh Ilina (2008). "Beyond Polemics: Science and Ethics of ADHD." *Nature Reviews, Neuroscience*, Vol 9, Pages 957–964.

[104] Stange, K. (2012). "An Empirical Investigation of the Option Value of College Enrollment," American Economic Journal: Applied Economics, Vol. 4 (1), January.

[105] Steele, C., and Aronson, J. (1998). Stereotype Threat and the Test Performance of Academically Successful African Americans, in The Black-White Test Score Gap 401 (Christopher Jencks & Meredith Phillips eds. 1998).

[106] Urzua, S (2008). "Racial Labor Market Gaps: The Role of Abilities and Schooling Choices". *Journal of Human Resources*, Vol. 43(4): pp. 919-971, 2008.

[107] Wansbeek, Tom and Erik Meijer. "Measurement Error and Latent Variables." A Companion to Theoretical Econometrics. Baltagi, Badi H. Blackwell Publishing, 2003. Blackwell Reference Online. 29 July 2010 <http://www.blackwellreference.com/subscriber/tocnode?id=g9781405106764_chunk_g978140510676413>

[108] White, J.L., Moffitt, T.E., Earls, F, Robins, L., & Silva, P.A. (1990). "How early can we tell? Predictors of childhood conduct disorder and adolescent delinquency." *Criminology*, Vol. 28, Pages 507-533.

[109] Whitemire, Richard (2010). Why Boys Fail: Saving Our Sons from and Educational System That's Leaving Them Behind. Amacon, Broadway, New York.

8 Appendix A

Gender Gap in Undergraduate Fall Enrollment				
	Race			
Year	White	Black	Hispanic	Asian
2000	-10.7%	-25.5%	-13.8%	-4.9%
2001	-10.8%	-26.2%	-14.4%	-5.6%
2002	-11.2%	-27.2%	-15.3%	-6.1%
2003	-11.8%	-28.1%	-16.9%	-7.3%
2004	-11.8%	-28.6%	-17.1%	-7.5%

Table A1: Difference in the percentage of undergraduate fall enrollment in degree-granting institutions between males and females (conditional on race). For instance, the percentages of white males and white females enrolled in degree granting institutions (conditional on total white enrollment) in 2004 were 44.1% and 55.9% respectively. Then, the difference is -11.8%. The percentages in Tables A2 and A3 follow this same interpretation. Source: NCES

Gender Gap in Associate's degrees. Year 2002 - 2003				
	White	Black	Hispanic	Asian
Gap	-18%	-32%	-21%	-14%
Selected Majors				
Business	-26%	-28%	-26%	-44%
Engineering	25%	10%	22%	37%
Health Professions	-65%	-35%	-30%	-43%
Liberal Arts and Sciences	-47%	-30%	-60%	-55%

Table A2: Difference in the percentage of Associate's degrees and selected majors obtained in the academic year 2002-2003 between males and females conditional on race. See Table A1 for an explanation on how these %s are obtained.

Gender Gap in Bachelor's degrees. Year 2002 - 2003				
	White	Black	Hispanic	Asian
Gap	-13%	-33%	-21%	-9%
Selected Majors				
Business	7%	-22%	-10%	-31%
Computer and Information Sci.	15%	1%	7%	44%
Health Professions	-29%	-14%	-13%	-29%
Liberal Arts and Sciences	-7%	-5%	-13%	-9%
Psychology	-24%	-12%	-19%	-27%

Table A3: Difference in the percentage of Bachelor's degrees and selected majors obtained in the academic year 2002-2003 between males and females conditional on race. See Table A1 for an explanation on how these %s are obtained. Source: NCES

9 Appendix B

NAEP Reading and Math Average Scores				
	White		Black	
	Math			
	Male	Female	Male	Female
National	249	247	222	223
	(0.3)	(0.2)	(0.4)	(0.5)
Large Cities	250	250	219	220
	(1.4)	(0.9)	(0.8)	(0.7)
	Reading			
National	227	233	200	209
	(0.3)	(0.3)	(0.6)	(0.6)
Large Cities	230	236	198	205
	(1.7)	(1.4)	(1.1)	(1.0)

Table B1: National Assessment of Educational Progress, average scores of fourth grade students in reading and math at national level and large cities, split by race and gender. Source: U.S. Department of Education

NAEP Reading and Math				
Proportion Performing Below Basic Level				
	National Level			
	Math		Reading	
	Male	Female	Male	Female
Grade 4	19%	19%	38%	31%
Grade 8	28%	29%	30%	22%

Table B2: Proportion of students in fourth and eighth grade performing below the basic level in the National Assessment of Educational Progress tests, open by gender. Source: The Nation Report Card. Reading 2009 and Mathematics 2009, National Assessment of Educational Progress at grades four and eight.

10 Appendix C

OLS Regressions (Constant Sample Across Specifications)								
Dependent Variable: College Enrollment								
Variable	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.689***	0.013	0.591***	0.025	0.639***	0.013	0.575***	0.012
Male	-0.107***	0.018	-0.115***	0.016	-0.013	0.016	-0.076***	0.016
Black	-0.132***	0.024	0.009***	0.023	-0.008	0.021	0.065***	0.022
Black x Male	-0.087**	0.034	-0.092***	0.032	-0.022	0.031	-0.066**	0.031
Hispanic	-0.184***	0.027	-0.048*	0.027	-0.112***	0.025	-0.054**	0.027
Hispanic x Male	0.031	0.039	0.018	0.037	0.054	0.036	0.018	0.037
Family Covariates	No		Yes		No		No	
Non-Cognitive Proxies	No		No		Yes		No	
Cognitive Proxies	No		No		No		Yes	
R^2	0.038		0.199		0.279		0.240	
Observations	5109		5109		5109		5109	

Table C1: OLS regressions similar to those presented in Tables 5, 7 and 8 with the only difference being that the sample remains constant across specifications. While Tables 5, 7, and 8 intend to maximize the size of the sample, this table shows that results are similar when the sample is kept constant across specifications. Family covariates, non-cognitive proxies and cognitive proxies are the same as those included in tables 5, 7 and 8 respectively. *, ** and *** indicate significance at 10%, 5% and 1% level, respectively. Source: NLSY97

OLS Regressions (Black and White Sample used in the Estimation of the Sequential Model)						
	Dependent Variable: College Enrollment					
Variable	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.487***	0.022	0.558***	0.011	0.474***	0.011
Male	-0.118***	0.014	-0.011	0.014	-0.074***	0.015
Black	0.022	0.019	-0.017	0.016	0.073***	0.019
Black x Male	-0.061**	0.026	0.007	0.024	-0.045*	0.025
Family Covariates	Yes		No		No	
Non-Cognitive Proxies	No		Yes		No	
Cognitive Proxies	No		No		Yes	
R^2	0.192		0.282		0.245	

Table C2: OLS regressions similar to those presented in Tables 5, 7 and 8 with the only difference being that the sample corresponds to the one used in the sequential model of educational attainment. Family covariates, non-cognitive proxies and cognitive proxies are the same as those included in tables 5, 7 and 8 respectively. *, ** and *** indicate significance at 10%, 5% and 1% level, respectively. Source: NLSY97

11 Appendix D

Table D1 presents the estimation results of the educational attainment model (as in table 9) with the only difference being that the following family background covariates were included in the estimation: mother's education, number of household members with age less than 18, and an indicator for broken family. As was mentioned earlier, the relevant results do not change: the gender gap in college enrollment is no longer present after the inclusion of the factors and both skill types are statistically significant at each stage of the schooling career.

	Educational Progression. Full Sample (With Family Covariates)															
	Finish Grade 10		Finish Grade 11		Finish High School		College 1		College 2		College 3		College 4		Finish	
	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
Constant	3.429***	0.158	3.197***	0.147	2.702***	0.135	-0.010	0.109	0.840***	0.129	0.227	0.157	0.232	0.209	0.362*	0.209
Male	0.940***	0.176	0.817***	0.187	0.961***	0.179	0.202*	0.118	0.329***	0.163	0.269	0.176	0.255	0.218	0.425***	0.197
Black	1.326***	0.230	1.090***	0.241	1.147***	0.228	1.855***	0.185	0.938***	0.235	1.347***	0.279	1.193***	0.347	0.333	0.303
Black x Male	-0.009	0.276	-0.277***	0.289	0.031	0.287	0.173	0.225	0.670*	0.311	0.594	0.361	0.784*	0.451	0.679	0.422
F^C	1.182***	0.079	0.994***	0.079	0.981***	0.076	1.489***	0.072	1.104***	0.093	1.300***	0.115	1.349***	0.157	0.590***	0.123
F^{NC}	0.928***	0.096	0.807***	0.100	0.856***	0.102	0.820***	0.084	0.767***	0.116	0.977***	0.147	1.061***	0.205	0.679***	0.185

Table D1: Model estimation results (logit coefficients). White and black sample . *, ** and *** indicate significance at 10%, 5% and 1% level, respectively. Family covariates were included in the estimation. Source: NLSY97

Table D2 shows the means and standard deviations of factor distributions after including family covariates in the estimation. The results indicate that differences between blacks and whites on cognitive and non-cognitive skills are reduced in relation to the results in table 10; however, racial differences persist.

	Factors: Normal Distributions			
	Cognitive		Non-cognitive	
	Mean	Std. Dev.	Mean	Std. Dev.
White Female	0	1	0	1
White Male	-0.086	1.151	-1.102	1.329
Black Female	-1.035	0.959	-0.091	0.860
Black Male	-1.228	0.969	-1.613	1.071

Table D2: Factor means and standard deviations of black and white males and females, after including family background covariates. Source: NLSY97.

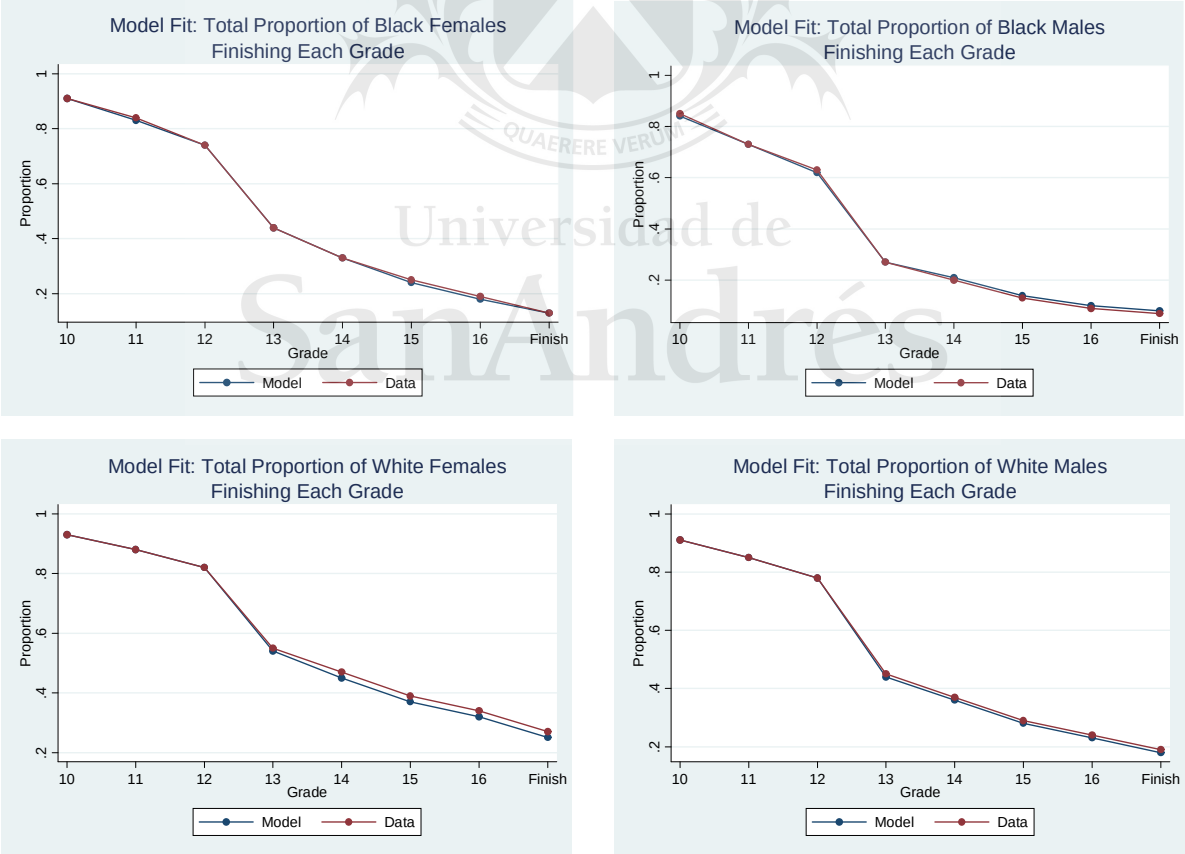


Figure D1: Data and model fit: educational attainment split by race and gender:

Parameters of Factors Distributions (α_s^u)				
	Cognitive		Non-cognitive	
Variable	Coef.	Std.Err.	Coef.	Std.Err.
	(1)	(2)	(3)	(4)
$male_i$	-0.067*	0.035	-1.003***	0.074
$male_i * \theta_i^u$	0.092***	0.027	0.355***	0.076
$black_i$	-1.203***	0.060	-0.256***	0.097
$black_i * \theta_i^u$	-0.043	0.040	-0.151	0.128
$black_i * male_i$	-0.067	0.078	-0.413***	0.133
$black_i * male_i * \theta_i^u$	-0.140***	0.058	-0.079	0.075

Table D3: Coefficients and standard errors of the factor parameters (i.e α_1^u to α_6^u, see equations (2) and (3)). *, ** and *** indicate significance at 10%, 5%, and 1% level, respectively. Source: NLSY97

12 Appendix E

	Measurement System Parameters (Cognitive Variables)					
	Arithmetic Reasoning	Assembling Objects	Word Knowledge	Paragraph Comprehension	Math Knowledge	General Sciences
	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.
Constant	-2.022***	-1.777***	-3.000***	-2.357***	-3.547***	-2.540***
F^C	0.641***	0.567***	0.596***	0.637***	0.625***	0.609***
Age_Asvab	0.156***	0.140***	0.221***	0.177***	0.256***	0.192***

Table E1: Coefficients and standard errors of the cognitive measurement system, see equation (4). "Age_Asvab" denotes the age at the time the exam was taken. Source: NLSY97

	Measurement System Parameters (Cognitive and Non-Cognitive Variables)				
	Fights	Grade Retention	Precocious Sex	GPA Grade Eight	Suspensions
	Coef.	Coef.	Coef.	Coef.	Coef.
Constant	-2.534***	-3.144***	-2.255***	6.189***	0.238***
F^C	-0.354***	-1.208***	-0.699***	0.996***	-0.267***
F^{NC}	-0.804***	-0.323***	-1.042***	0.541***	-0.323***
Male	-	-	-0.933***	-	-
Black	-	-	-	0.866***	-

Table E2: Coefficients and standard errors of the cognitive/non-cognitive measurement system (see equation (5)). Binary variables such as grade retention, fights and precocious sex generate logit coefficients; therefore, they cannot be interpreted directly. Given the substantial differences between males and females in reporting sexual behavior, a dummy for being male was included to control for misreporting bias. Similarly, due to the fact that whites and blacks attend on average different types of schools a dummy for race was included in the measure for GPA at grade eight. *, ** and *** indicate significance at 10%, 5% and 1% level, respectively. Source: NLSY97

13 Appendix F

	Oaxaca's Decompositions: Unconditional Probability (Base: Males Covariates)							
	Grade 10	Grade 11	Finish HS	College 1	College 2	College 3	College 4	Finish
Panel A1: Whites								
Actual Gap	0.023	0.035	0.036	0.097	0.093	0.091	0.087	0.067
Gap due to Preferences	-0.061	-0.082	-0.111	-0.049	-0.044	-0.036	-0.026	-0.027
Gap due to Skills Diff.	0.084	0.117	0.147	0.146	0.137	0.127	0.113	0.094
Panel A2: Contributions to *Gap due to Skills Differences*								
Cognitive	18%	15%	13%	9%	7%	4%	3%	1%
Non-cognitive	82%	86%	88%	91%	93%	96%	97%	99%
Panel B1: Blacks								
Gender Gap	0.062	0.106	0.113	0.169	0.125	0.104	0.082	0.051
Gap due to Preferences	-0.079	-0.097	-0.134	-0.047	-0.06	-0.051	-0.042	-0.039
Gap due to Skills Diff.	0.141	0.203	0.247	0.216	0.185	0.155	0.124	0.090
Panel B2: Contributions to *Gap due to Skills Differences*								
Cognitive	14%	13%	13%	15%	15%	14%	14%	13%
Non-cognitive	86%	87%	87%	84%	85%	86%	86%	87%

Table F1: Oaxaca's decompositions for unconditional probabilities of finishing different stages of schooling career. Black and white samples were considered separately. Male covariates are used as base. Source: NLSY97.