

# Incentives to perform in a low-stakes setting and high-stakes outcomes

Gonzalo Castex and Evgenia Dechter\*

March 2021

## Abstract

We study the subjective incentives to perform well in low-stakes environments and evaluate the limitations of predictions obtained in such settings. We examine how these subjective incentives vary with personality traits, work-effort attitudes and a range of individual characteristics. The robust across specifications result is the gender gap in incentives; women exert more effort and are more motivated in the low-stakes setting than men. Test-specific effort and motivation are strong predictors of the low-stakes test scores; however, they do not predict outcomes in high-stakes settings, including high-stakes tests scores and wages. Our findings point to important implications, which relate to the interpretation of test scores and rankings achieved in low-stakes environments, with a particular emphasis on gender differences in performance.

Keywords: high stakes; low stakes; incentives; performance; gender gap.

JEL codes: J24; J16; I21; I24; D91

---

\*Gonzalo Castex, School of Economics, University of New South Wales, Australia. Email: g.castexhernandez@unsw.edu.au. Evgenia Dechter, School of Economics, University of New South Wales, Australia. Email: e.dechter@unsw.edu.au.

# 1 Introduction

It is a common practice for institutions to administer tests designed to assess achievement, progress and skills. Some of these assessments are administered in low-stakes environments since the arrangements do not involve individual-level performance-based incentives. Examples span from schools, where students undergo periodic assessments, to surveys, which administer achievement tests for research purposes.<sup>1</sup>

Economic theory predicts that individuals will minimize costly effort in activities that do not award achievement.<sup>2,3</sup> However, most participants in tests in low-stakes settings (with no or low non-progressive awards) receive scores in an expected range. This may be due to unobserved benefits respondents gain from achieving high scores in such tests, or benefits from achieving outcomes to the best of their ability (for example, benefits from learning about own skills and interests). A large number of studies show that scores achieved in such tests are positively correlated with a range of measures of socio-economic success.

There is a growing interest in studying how individuals respond to incentives. This research provides tools to help interpreting results of tests administered in a range of environments with varying incentives and competitiveness, including school achievement tests, aptitude tests (for example, for college admissions), and job screening tests. These studies report differences in response to incentives across groups, in particular differences between men and women. Most studies that examine the response to incentives are conducted using controlled experiment techniques where individuals are asked to perform tasks in varying settings. For example, Gneezy et al. (2003), and Gneezy and Rustichini (2004), among others, show that men's performance improve in

---

<sup>1</sup>Examples of such tests at schools include the well known Programme for International Student Assessment (PISA); National Assessment of Educational Progress (NAEP) in the United States; National Assessment Program - Literacy and Numeracy (NAPLAN) in Australia, among many other schemes. These tests are not designed to reward individual achievement, although there are potential benefits for country-level performance (PISA) or school level performance (NAEP and NAPLAN).

<sup>2</sup>One framework that summarizes such behavior is the "rational cheater" model of motivation. It posits that agents are self-interested and continuously search for ways to increase their welfare; they will shirk whenever they perceive that the marginal benefits of such behavior exceed its marginal costs.

<sup>3</sup>A number of studies find a positive relationship between test scores and incentives, see for example Gneezy and Rustichini (2000), Daly and Lavy (2009), and Segal (2012).

competition, whereas women’s performance does not change or declines slightly. Using a field experiment, Levitt et al. (2016) examine how varying reward schemes affects performance of school students in Chicago. They show that boys are more likely to respond to incentives than girls. Azmat et al. (2016) utilize the variation in the stakes of assessment tests in a private school in Spain and show that boys performance increases with stakes. Using an uncontrolled environment of the GRE exam, Schlosser et al. (2019) utilize high-stakes and low-stakes sections of the exam to study how different demographic groups respond to incentives.<sup>4</sup> They show that the drop in achievement measured as a difference in scores between high-stakes and low-stakes sections of the exam is larger for men than for women. Schlosser et al. (2019) argue that the differences in performance between high and low stakes sections of the exam are partly explained by larger declines in test-effort for men.

We contribute to this literature in a number of ways. We focus on outcomes of a survey-based achievement test, where individuals do not receive performance based incentives. This is a low-stakes setting test and subjective reasons define incentives to perform well. Using a set of unique measures of test-specific subjective incentives, we study the formation of incentives and examine whether they are correlated with personality traits and individual characteristics. We test how the performance in a low-stakes test varies with incentives. Further, we examine whether the subjective incentives to do well in a low-stakes settings can predict high-stakes outcomes. We derive a number of implications about the interpretation of test scores and rankings achieved in low-stakes settings.

The focus of our analysis is the Armed Services Vocational Aptitude Battery (ASVAB), administered by the 1997 National Longitudinal Survey of Youth (NLSY97).<sup>5</sup> The NLSY97 participants were invited to participate in the ASVAB without being offered performance-based incentives.<sup>6</sup> The Armed Forces Qualifications Test (AFQT) is con-

---

<sup>4</sup>Graduate Record Examination General Test (GRE) is a commercially run standardized test that is an admissions requirement for many graduate schools in the United States and other English speaking countries.

<sup>5</sup>The ASVAB is a battery of 10 tests; utilized as a screening and sorting exam in the US Armed Forces.

<sup>6</sup>Respondents in NLSY97 were paid a flat rate of \$75 for test participation and received a detailed report summarizing their results.

structured using four tests in the ASVAB to summarize verbal and quantitative abilities. The AFQT scores are often used in the literature as a measure of cognitive achievement, aptitude and intelligence.<sup>7</sup> We utilize a unique information on subjective incentives to do well on the test, measured by test-specific participation motivation and test-specific effort. We examine the relationships between test-specific effort and motivation, individual characteristics, personality traits and work-effort attitudes, and the AFQT scores. We study the differences in performances in high-stakes and low-stakes tests (SAT and AFQT) and show these differences vary with individual characteristics and subjective incentives. Finally, we analyze the relationship between the subjective incentives to perform well in a low-stakes test and long-term labor market success, measured by wages.

First, we study the formation of test-specific subjective incentives by examining the relationship between test-specific effort and motivation, work-effort attitudes and personality traits, range of individual characteristics and AFQT scores. Test-specific motivation and effort vary substantially across individuals and are higher, on average, for women. We find no relationships between a range of personality traits, work-attitudes and individual characteristics and subjective incentives. The more motivated individuals achieve higher AFQT scores but the relationship is not monotone. Our results suggest that participants who have a higher subjective incentive to perform well are not the more able ones and the least motivated are not the least able ones. Thus, the ranking of ability based on AFQT scores might not reflect the true ability ranking.

Second, we study the relationship between the subjective incentives and outcomes in high-stakes settings. We examine whether the measures of test-specific effort and motivation provide any additional information on individual long-term labor market performance. We find no relationship between test-specific motivation and wages, for men and women. For women, we do not find any relationship between test-specific effort

---

<sup>7</sup>See see for example, among others, Herrnstein and Murray (2010); Heckman (1995); Neal and Johnson (1996). Borghans et al. (2016) lists 50 studies that utilize the AFQT scores as a measure of cognitive skill and report positive relationships between the AFQT scores and various socio-economic outcomes.

and wages. For men, the correlation between test-specific effort and wages follows a different pattern from that observed in the analysis of AFQT scores. Male participants who report test-specific effort in the 20th percentile earn higher wages than those in the higher effort percentiles but achieve lower AFQT scores. This suggests that the test-specific effort and motivation in a low-stakes test environment do not predict success in the labor market.

This part of the analysis also adds to the growing literature on returns to cognitive and noncognitive skills; see for example, Deming (2017), Weinberger (2014), Heineck and Anger (2010), Heckman, Stixrud, and Urzua (2006), Mueller and Plug (2006), Bowles, Gintis, and Osborne (2001), among others. First, we analyze whether the subjective incentives in a low-stakes setting predict labor market outcomes. This noncognitive channel has received relatively little attention in the existing literature. We also contribute by examining the limitations of cognitive skills rankings obtained in survey-based achievement tests.<sup>8</sup>

Third, we compare outcomes in high-stakes and low-stakes tests. We study how the differences between AFQT and SAT (low-stakes and high-stakes tests) scores vary with individual characteristics.<sup>9</sup> We compute the differences in terms of percentiles and show that men perform worse than women in the low-stakes setting. We attribute this result to gender differences in effort and motivation in low-stakes tests compared to high-stakes tests. Limiting the sample to those who report high test-specific effort in ASVAB, we show that the difference between AFQT and SAT scores does not vary by gender. These results are consistent with findings reported in a number of studies

---

<sup>8</sup>One study that examines the rankings of skills in a survey-based achievement test is Segal (2012), who proposes to measure the test-specific motivation using a coding speed test score. The coding speed test is a part of ASVAB in NLSY79; the task is to match words with four-digit numbers. Segal (2012) shows that those who perform better on the coding speed test earn higher wages. However, a number of studies, including Heckman (1995) and Cawley, Conneely, Heckman, and Vytlačil (1997), argue that mental speed and fluid intelligence also contribute to the performance on the coding speed test, which may explain its positive correlation with wages. In contrast, using a more direct set of measures of test-takers' effort and motivation, we find that subjective incentives to perform well in a low-stakes test do not predict long term high-stakes outcomes.

<sup>9</sup>The Scholastic Assessment Test (SAT) is an essential test for those who plan to enroll in colleges and universities. The SAT is a standardized college admissions test in the US; it measures reading comprehension, computational ability, and clarity of expression. The SAT score can play an important role in college admission decisions; therefore, for those who are considering higher education this is a high-stakes test.

that compare gender differences in response to incentives, see for example, Schlosser et al. (2019) and Azmat et al. (2016).

We conclude that the test-specific motivation and effort that predict scores in low-stakes environments do not predict outcomes high-stakes settings or longer term outcomes such as wages. We also show that women are more likely than men to exert more effort in low-stakes settings. An important implication of these findings relates to the interpretation of test scores and rankings achieved in low-stakes environments, which include academic and research-oriented settings. Achievements in low-stakes settings are better measures of true ability for those who exert higher effort; the results will understate rankings of those individuals or groups that exert low effort. This implication also holds in reverse, high performance in a high-stakes environment does not prescribe high performance in a low-stakes setting, especially for men.<sup>10</sup> A number of previous studies provide further support for these implications. For example, Freund and Rock (1992), Di Chiacchio et al. (2016) and Schlosser et al. (2019) show male students exert less effort in exams with low stakes; Willingham and Cole (1997) and Rothstein (2004) show that college admission standardized tests tend to underpredict performance for women and overpredict performance for men compared to their performance in terms of GPA. Our findings emphasize that the design of assessments in academic, research-oriented and other environments, may affect the resulting ranking and interpretation of outcomes.

The remainder of the paper is organized as follows. Section 2 describes the data and summarizes key variables. In Section 3 we discuss the estimation methods and report the findings. In Section 4 we discuss the findings, outline the main implications, and conclude.

---

<sup>10</sup>For example, achievement on a high-stakes job screening test might not be a good predictor of productivity on a job with a flat compensation scheme.

## 2 Data and summary statistics

The data are from NLSY97, a nationally representative sample of 8984 individuals who were 12-16 years old in 1997. We employ both cross-sectional and supplemental samples (excluding the military supplement) and use the base year weights to achieve representativeness of the population.

Respondents of NLSY97, in 1997-1998, were invited to participate in ASVAB, a sequence of tests that cover basic math, verbal, and operational skills. Using age adjusted scores, we construct normalized Armed Forces Qualifications Test (AFQT) scores, mean zero and standard deviation one, using results from Arithmetic Reasoning, Numerical Operations, Word Knowledge and Paragraph Comprehension tests. There are 7,001 respondents with valid AFQT scores. The AFQT is a low-stakes test since respondents were paid a flat rate of \$75 and there was no performance based compensation. The Armed Forces Qualifications Test (AFQT) scores are often used as a measure of cognitive skill and a large number of studies report positive relationships between the AFQT scores and various socio-economic outcomes; see for example Borghans et al. (2016) who provide an extensive survey of this literature.

To measure subjective incentives we use a set of variables that provide information on individual's ASVAB test-specific effort and motivation. First, "test-specific effort" varies between 1 to 5, and is measured as a response to question about how much the respondent agrees with the statement "I tried to do my best on the test". We construct five dummy variables to reflect the five optional choices, category 1 being the lowest. Second, "test-specific motivation" is defined as a set of dummy variables. Respondents were asked to provide first and second reasons for participation in ASVAB, "What are the two most important reasons you took the test today? (Choose two answers)" optional answers are as follows: (1) Because it's an important study; (2) To see what it's like to take a test on a computer; (3) To see how well I could do on the test; (4) To learn more about my interests; (5) Family member wanted me to take it; (6) To get the money; (7) I had nothing else to do today. We combine categories (3) and (4)

in estimations.<sup>11</sup> We also define a continuous motivation variable using ranking that ranges from 2 to 10, as described in Appendix Table A.1. Third, an additional measure of test-specific motivation is “importance of money” which reports yes/no (1/0) the response to the following question, “If you were not offered any money, would you still have taken the test?”. There are 6,854 respondents for whom all three measures of subjective incentives are available.

Summary statistics of subjective incentives measures are reported in Table 1, for men and women. Appendix Figure A.1 shows distributions of the test-specific effort (“I tried to do my best on the test”) and shows percentage of respondents in each category. Around 20% of men and 14% of women report effort in 1-3 categories. Appendix Figure A.2 depicts distributions of AFQT scores by test-specific effort, we compare respondents with 1-4 effort levels and respondents reporting level 5. Those who exert higher effort achieve higher scores. Appendix Figure A.3 depicts distributions of AFQT scores for each motivation level (motivation levels defined in Appendix Table A.1) and compares them to the scores of the most motivated respondents. Appendix Figure A.3 shows that the distribution of the test scores of the less motivated respondents is skewed to the left but it is not strikingly different from the distribution of the very motivated respondents. The distribution of scores of the least motivated is similar to that of the most motivated respondents.

NLSY97 collects a range of individual characteristics. Personality traits and work-effort attitudes are from the 2008 Personality Scale supplement. The 2008 Personality Scale supplement includes 9 items that describe personality traits and 3 items that describe work-effort attitudes. For each item in this questionnaire individuals rank how much they agree with each statement; for example, “Using a scale from 1 to 7, where 1 means disagree strongly and 7 means agree strongly, please rate how well each pair of traits applies to you, even if one characteristic applies more strongly than the other: Dependable, self-disciplined”.<sup>12</sup> All personality scale variables are

---

<sup>11</sup>Respondents who choose these two categories are very similar in most key variables. Estimation that use all seven categories without combining categories (3) and (4) produce very similar results.

<sup>12</sup>The NLSY97 administered a similar personality scale supplement in 2002. However, the 2002 question-



utilized without modifications (excluding invalid entries). Attitudes towards work effort and work standards are measured by responses to the following statements: “I do what is required, but rarely anything more.”, “I have high standards and work toward them.”, “I make every effort to do more than required.”, where respondents rank how these statements match their attitudes on a 1 to 7 scale. Table 1 summarizes the information on personality traits and work-effort attitudes for men and women. Availability of personality traits and work-attitudes further reduces the sample size to 5,533 individuals.

Individual and family background characteristics used in the analysis include race, metro status, education, parental education levels (we use the higher level of education to represent parental education), intact family indicator (equals one if both parents were living with the child in 1997, when participants are 12-16 years old) and family income (relevant to the period when participants were 16-19 years old, excluding those years when they were not living with their parents). There are 4,589 individuals for whom AFQT, subjective incentives, and individual and family characteristics variables are available. The subsample of those who also have personality traits and work-attitudes consists of 3,872 observations. To analyze labor market outcomes, we construct real hourly wages for 19-29 years old. For this analysis we use individuals not enrolled in school or military service, who work at least 10 hours per week and earn real hourly wages within the range of 1 to 200 dollars (in 2007 prices, deflated using the CPI). There are 17,490 observations available for this analysis, 9,195 for men and 8,295 for women, each individual may be included in the sample more than once.

Table 2 presents summary statistics of the key variables by the primary reason to participate in ASVAB. Background variables, education at the age of 24, wages at the

---

naire covers only a small subsample of NLSY97 respondents whereas the 2008 questionnaire has a wider coverage. Moreover, the 2008 questionnaire has more questions and covers more aspects of personality. We use the 2008 questionnaire and rely on the assumption of stability of personality traits. There is a substantial body of evidence documenting stability in personality traits. For example, Cobb-Clark and Schurer (2012) show that the “big-five” personality traits are stable over time; Epstein (1979) finds that personality traits are stable across multiple situations. Studies in genetics and neuroscience show further that there is a biological basis for the existence of stable traits, see for example Bouchard Jr. and Loehlin (2001), Canli (2006) and DeYoung et al. (2010)

age of 24 and AFQT scores vary across the motivation categories. For example, the highest test scores are achieved by respondents who sat the test “To get the money”, followed by those who answered “To see how well I could do on the test” and “To learn more about my interests”. The lowest test outcomes are achieved by those who responded “To see what it’s like to take a test on a computer” and “I had nothing else to do today”. There are also some differences in gender and race composition across the different categories. For example, there are more males in the “To get the money” and “I had nothing else to do today” categories; whereas proportion of females is higher in “To see how well I could do on the test” and “To learn more about my interests” categories. The proportion of Blacks is relatively high in categories “To see what it’s like to take a test on a computer”, “Because it’s an important study” and “I had nothing else to do today”. The differences in gender and race composition are not particularly large but important and therefore accounted for in the regression analysis.

To compare low-stakes and high-stakes tests results, we use the AFQT scores and Scholastic Assessment Test (SAT) scores. SAT scores are available for a subset of respondents in NLSY79, there are 970 individuals with SAT scores who also have AFQT scores, subjective incentives variables, personality traits and individual and family characteristics. For respondents who report multiple scores, we choose the highest one. We use normalized SAT scores, with mean zero and standard deviation one. Appendix Figure A.4 reports the AFQT and SAT scores distributions for men and women (for the subsample with available SAT scores). On average, men do better than women in SAT scores. The AFQT distribution does not vary much by gender. This result is consistent with other studies, see for example, Speer (2017), who also shows that men and women achieve similar scores on most math and verbal tests in the ASVAB.

### 3 Results

We study the role of subjective incentives in performance in low-stake tests and high-stakes settings. First, we examine how subjective incentives, measured by test-specific effort and motivation, vary with individual characteristics and personality traits. Second, we estimate the relationship between the AFQT scores, test-specific effort and motivation, work-effort attitudes and personality traits. Third, we examine the relationship between test-specific effort and motivation, wages and scores in a high-stakes test.

#### 3.1 The variation in test-specific effort and motivation

We analyze the correlations between the subjective incentives (test-specific effort and motivation), individual characteristics and work-effort attitudes and personality traits. We estimate the following specification,

$$\text{Subjective incentive } var_i = \lambda_0 + X_i \lambda_1 + \lambda_{2j} \sum_{j=1}^9 \text{trait}_{ij} + \lambda_{3j} \sum_{j=1}^3 \text{attitude}_{ij} + \epsilon_i, \quad (1)$$

where *Subjective incentive var<sub>i</sub>* is the measure of test-specific effort or motivation of individual *i*. Motivation is defined in Appendix Table A.1. Vector *X<sub>i</sub>* includes gender, race, parental education, intact family indicator, family income and metropolitan status. Variables in *trait<sub>ij</sub>* include the nine personality traits. The three work-effort attitudes measures are given by *attitude<sub>ij</sub>*.

Table 3 reports the results. Columns (1) and (2) report results for test-specific effort and columns (3) and (4) report results for test-specific motivation. The most consistent finding across the different specifications is that men put less effort and are less motivated about achieving their best in the test. There are no correlations (or no consistent correlations) across specifications for other included variables.

### 3.2 Test-specific effort and motivation, personality traits, work-effort attitudes and AFQT scores

To assess the relationship between subjective incentives, measured by test-specific motivation and effort, personality traits, work-effort attitudes and test scores we estimate the following specification:

$$AFQT_i = \beta_0 + X_i\beta_1 + \beta_{2j} \sum_{j=1}^4 effort_{ij} + \beta_{3j} \sum_{j=1}^{11} motivation_{ij} + \beta_{4j} \sum_{j=1}^9 trait_{ij} + \beta_{5j} \sum_{j=1}^3 attitude_{ij} + \epsilon_i, \quad (2)$$

where  $AFQT_i$  is the age-adjusted normalized AFQT score of individual  $i$ ,  $effort_{ij}$  is a vector of four dummy variables measuring the test-specific effort input in the ASVAB tests (the lowest effort category is omitted),  $motivation_{ij}$  is the test-specific motivation, measured as a vector of ten dummy variables summarizing the reasons to participate in the test and the importance of money,  $j$  categorizes the measures. The omitted primary and secondary test participation motivation reason category is the combination of “To see how well I could do on the test” and “To learn more about my interests”.

Table 4 reports estimation results of equation (2). Columns (1)-(3) report results for men, columns (4)-(6) report results for women. Columns (1) and (4) report results including only the test-specific motivation and effort variables; these variables are strongly correlated with the AFQT score. The relationship between the test-specific effort and test score is not monotone. Respondents with the highest level of effort receive higher scores, for men and women; however, men who report effort levels of 2 or 3 score lower than those with the lowest reported effort; scores of those with effort level 4 are statistically indistinguishable from scores of those with the lowest effort level.

Most motivation variables have negative coefficients (as expected, given that the omitted category is the combination of “To see how well I could do on the test” and “To learn more about my interests”). Respondents with participation reason being

“To get the money” receive the highest scores. Those who have stated that they would have participated in the ASVAB even without the participation award (“importance of money”) achieve higher AFQT scores.

Columns (2) and (5) report the results of estimations which include the personality traits. High degrees of extraversion, anxiety and organization are associated with lower AFQT scores, for men and women. More creative respondents score higher. For men, being more critical or quarrelsome is associated with higher AFQT scores.

Columns (3) and (6) report the results of estimations which also include the work-effort attitudes. Respondents who report more positive attitudes towards work effort and work standards receive higher AFQT scores. Higher level in “I make every effort to do more than what is expected of me” is negatively correlated with the AFQT score. One interpretation for this result is that the more able individuals do not need to exert more effort at work to perform to a higher standard. Another interpretation is that the more able individuals occupy more demanding jobs.

### 3.3 Subjective incentives, AFQT scores and wages

We examine whether the measures of test-specific effort and motivation convey any additional information on individual long-term performance, measured by log real hourly wages, beyond the range of personality traits, work-effort attitudes, and individual characteristics.

We estimate the following specification:

$$\begin{aligned} \log W_{it} = & \gamma_0 + X_{it}^w \gamma_1 + \gamma_2 AFQT_i \\ & + \gamma_{3j} \sum_{j=1}^4 effort_{ij} + \gamma_{4j} \sum_{j=1}^{11} motivation_{ij} + \gamma_{5j} \sum_{j=1}^9 trait_{ij} + \gamma_{6j} \sum_{j=1}^3 attitude_{ij} + \xi_{it}, \end{aligned} \tag{3}$$

where  $W_{it}$  is the real hourly wage of individual  $i$  in year  $t$ ,  $X_{it}^w$  includes years of schooling, race, metropolitan status and age. Other controls are as defined in Equations

(1) and (2). Equation (3) is estimated using a pooled sample of 21-29 years old, clustering observations at the individual level. We focus on the coefficients of AFQT scores, test-specific motivation and effort, personality traits and work-effort attitudes.

We report two sets of results, including and excluding the schooling variable. As noted in other studies, since some schooling is determined by prior cognitive skills, and since schooling leads to higher wages, we should place some weight on the estimates that do not control for schooling when trying to examine the effect of AFQT scores on wages (see for example, Neal and Johnson 1996; Zax and Rees 2002, among others). By excluding the schooling variable from the estimation we focus on the estimates of net effects of cognitive skills, test-specific effort and motivation, personality traits and work-effort attitudes on wages (the direct effects of these channels additionally to their effects through schooling).

Tables 5 and 6 present estimation results of equation (3), for men and women, respectively. Columns (1) and (5) report estimation results including only controls in  $X_{it}^w$  and depicts the positive correlation between wage and AFQT score, the positive return to AFQT is a result commonly reported in the literature. An increase by one standard deviation in the AFQT score is associated with a 5.6% (2.2% when schooling variable is included) increase in the hourly wage for men and 10.7% (3.5% in the estimation with schooling) for women.<sup>13</sup>

Results in Table 5 show no statistically significant relationship between test-specific motivation and wages. One exception is the importance of money, those who state that they would have participated even without a compensation earn lower wages. In contrast, this variable is positively correlated with the AFQT score (see Table 4). The relationship between the test-specific effort and wages is not monotone, participants who report effort in the 20th percentile earn higher wages than those in the higher percentiles but achieve lower AFQT scores. Table 6 reports the results for women. There are no statistically significant relationships between test-specific motivation, effort and

---

<sup>13</sup>This result is consistent with findings in a large number of studies. For example, Deming (2017), Castex and Dechter (2014) and a survey study by Borghans et al. (2016).

wages.

Columns (3) and (7) in Tables 5 and 6 report estimation results that include personality traits. Coefficients of several personality traits are statistically significant. For men, being more critical and dependable, less reserved and sympathetic, are associated with higher wages. With the exception of being more critical, these personality traits are not associated with higher AFQT scores for men. For women, being more dependable, less sympathetic and more organized are associated with higher wages. These personality traits are not associated with higher AFQT scores for women. There is no overlap in personality traits that positively affect both wage rates and AFQT scores (see Table 4).

Columns (4) and (8) in Tables 5 and 6 report estimation results that include work-effort attitudes. Men and women with more positive attitudes towards work effort and higher work standards earn higher wages. More favorable work-effort attitudes are also positively correlated with the AFQT scores (see Table 4).

### 3.4 Differences in ranking between high-stakes and low-stakes test results

We further examine the role of subjective incentives by comparing rankings in academic assessment tests performed in low-stakes environments and those obtained in high-stakes environments. This analysis compares the percentile rankings of SAT and AFQT scores.<sup>14,15</sup>

We measure individual change in performance using the following regression model,

$$\%AFQT_i - \%SAT_i = \delta_0 + X_{it}\delta_1 + \delta_{2j} \sum_{j=1}^9 trait_{ij} + \delta_{3j} \sum_{j=1}^3 attitude_{ij} + \xi_{it}, \quad (4)$$

---

<sup>14</sup>We follow Neal and William (1996), among others, and assume that both SAT and AFQT measure the “pre-college factors”; where these factors reflect all forces that have shaped one up until the time the test score is measured. This includes innate ability, family background, including parental, school, peer, and environmental influences.

<sup>15</sup>the two tests are comparable in structure. The SAT has four sections: Reading, Writing and Language and Math. Our analysis uses the total SAT score and therefore comparable to the AFQT score.

where  $\%AFQT_i$  and  $\%SAT_i$  are scores expressed in 1 to 100 percentiles. We control for individual characteristics to allow for average differences in performance. We report results for the entire sample of SAT takers as well as by AFQT effort level (“I tried to do my best on the test”), distinguishing between low effort (ASVAB effort < 5) and high effort groups (ASVAB effort = 5).

Table 7 summarizes the results. Columns (1) and (4) report results for the entire sample, with and without personality traits controls; men rank lower in the AFQT scores compared to the SAT scores. The performance of men increases by 2.7 percentile points in SAT compared to AFQT. This pattern is more pronounced for the low effort group (columns 2 and 5), within this group, men performance is higher by 3.4 to 3.7 percentile points. However, for the high effort group there is no statistically significant difference between the AFQT and SAT ranking between men and women.

## 4 Discussion and conclusion

A range of subjective reasons may define incentives to perform well on a low-stakes test. Using unique measures of subjective incentives, we study how they vary with personality traits, work-effort attitudes, a range of individual characteristics and family background variables. We study how these incentives affect the low-stakes test performance outcomes. Further, we examine whether the incentives that affect individual performance in the low-stakes test can predict outcomes in high-stakes settings. We draw a set of implications about the relationship between ranking of skills obtained in a low-stakes setting and performance in a high-stakes environment.

Incentives to do well on a low-stakes test vary with individual characteristics, but the most consistent and robust is the gender difference. Women exert more effort in a low-stakes test and are more motivated to do well than men; but there is no consistent variation across other observable individual characteristics. We further explore the gender differences in incentives by comparing results in low-stakes and high-stakes tests, AFQT vs. SAT. The AFQT scores do not vary by gender; however, men achieve



higher SAT scores than women. We examine the percentile differences between AFQT and SAT scores; male performance is higher in SAT compared to AFQT. The difference is larger for the low ASVAB effort group, implying that differences in effort in the low-stakes exam distort the ability ranking. These results suggest that performance in a low-stakes test is not a good predictor of performance in a high-stakes, especially when the focus is gender differences in outcomes. These results are consistent with findings reported in Azmat et al. (2016) who show that female students outperform male students when the stakes are low but not when stakes are high. The results are also consistent with Schlosser et al. (2019), who study differences in performance in the high-stakes environment (the GRE exam) and low-stakes environment (voluntary experimental section of the GRE). They show that men exhibit a larger difference in performance between the high-stakes and low-stakes exams than women. Azmat et al. (2016) attribute some of the gender difference to the behavioural response to an increase in stakes, suggesting that women performance declines under pressure compared to men. Schlosser et al. (2019) argue that the difference between high-stakes and low-stakes outcomes are partly explained by the fact that men exert lower effort in low-stakes tests compared to women. Our findings are more consistent with Schlosser et al. (2019). Using unique measures of subjective incentives we show that women exert more effort than men in the low-stakes environment.

We find that test-specific effort and motivation are strong predictors of the AFQT scores, including in specifications that control for long-term work-effort attitudes, personality traits, and a range of individual characteristics. Higher test-specific effort and motivation are associated with higher test scores, however the relationship is not monotone. The results suggest that higher cognitive ability is not necessarily positively correlated with subjective incentives to perform well in a low-stakes setting. This finding raises further concerns about the observable rankings of the low-stakes tests scores, since they might not reflect the true ability ranking but a combination of subjective incentives to do well and ability.

This part of the analysis also adds to the growing literature on the relationship

between noncognitive skills and achievement tests outcomes. First, we extensively examine the role of subjective incentives, a noncognitive channel which received relatively little attention in the existing literature. Second, we analyze the relationships between subjective incentives, personality traits, attitudes, individual characteristics, and performance on achievement test. We find no relationship between subjective incentives and personality traits. We also report that high degrees of extraversion, anxiety and organization are associated with lower AFQT scores, for men and women. More creative respondents score higher. In addition, for men, being more critical or quarrelsome is associated with higher AFQT scores. This part of the analysis is consistent with other studies. For example, in economics, Dohmen, Falk, Huffman, and Sunde (2010), Benjamin, Brown, and Shapiro (2013), Cubel, Nuevo, Chiquero, and Fernandez (2016), and Borghans et al. (2016) show that more neurotic subjects perform worse, and that more conscientious individuals perform better. In psychology, a large number of studies show that openness to experience is positively correlated and neuroticism is negatively correlated with test outcomes (see for example the meta-analysis by Ackerman and Heggestad 1997), which is consistent with our findings.

We examine whether the measures of subjective incentives to perform well in a low-stakes setting provide any additional information on individual long-term labor market performance. We find no relationship between test-specific motivation and wages, for men and women. For women, we do not find any relationship between test-specific effort and wages. For men, the correlation between test-specific effort and wages follows a different pattern from the correlation between test-specific effort and AFQT scores. Male participants who report effort in the 20th percentile earn higher wages than those in the higher percentiles but achieve lower AFQT scores. This suggests that subjective incentives to do well in a low-stakes test environment do not predict success in the labor market.

This part of the analysis also adds to the growing literature on the relationship between noncognitive channels and labor market outcomes. First, we find that higher subjective incentives to do well in a low-stakes setting do not predict higher wages.

Second, the analysis of the relationship between other personality traits and wages is consistent with previous studies. We find that more critical and dependable, less reserved and sympathetic men earn higher wages. With the exception of being more critical, these personality traits are not associated with higher AFQT scores for men. For women, being more dependable, less sympathetic and more organized are associated with higher wages. These personality traits are not associated with higher AFQT scores for women. Wage regression results are consistent with findings in the literature. For example, Heckman et al. (2006) and Fletcher (2013), report that higher conscientiousness, lower agreeableness and lower neuroticism show a robust positive relationship with earnings.

#### **4.1 Implications for academic assessment in low-stakes environment**

Assessment tests in low-stakes environments are a common practice at schools around the world. For example, Programme for International Student Assessment (PISA), designed to measure 15-year-olds' abilities across countries; National Assessment of Educational Progress (NAEP) in the United States is a continuing assessment scheme administered in grades 4, 8, and 12; National Assessment Program - Literacy and Numeracy (NAPLAN), administered in all schools in Australia in grades 3,5,7 and 9; all these tests are designed to measure achievement in major academic fields. There are no specified rewards for individual achievement in such tests, although there are potential benefits for country-level performance (PISA) or school level performance (NAEP and NAPLAN). The objective of such tests is to assess student performance in the context of gender, race/ethnicity, public or private school, teacher experience, and other factors. However, as we show in our study, the ranking achieved in such test environment might be biased due to the variation of subjective incentives to do well on the test within school (for example, gender differences), across schools, and across countries. Previous studies provide support for these implications. For example, Freund and

Rock (1992) using NAEP exam and Di Chiacchio et al. (2016) using the Italian Pisa Science exam, find that male students are more likely to attempt to complete these exams exerting as little effort as possible. Schlosser et al. (2019) show that that test score differences between males and females are larger in the SAT compared to the NAEP. The reverse implication of our findings is that high performance in a high-stakes environment does not predict high performance in a low-stakes setting, especially for men. This implication is consistent with findings in Willingham and Cole (1997) and Rothstein (2004), who show that college standardized tests underpredict performance for women and overpredict performance for men. Another example is a high-stakes job screening test, which might not be a good predictor of productivity on a job with a flat compensation scheme. Our findings combined with findings in the exiting literature suggest that policy makers should direct more attention to designing and interpreting assessments in academic and other environments.

## References

- Ackerman, P. L. and E. D. Heggstad (1997). Intelligence, personality, and interests: evidence for overlapping traits. *Psychological Bulletin* 121(2), 219–45.
- Azmat, G., C. Calsamiglia, and N. Iriberry (2016). Gender Differences in Response to Big Stakes. *Journal of the European Economic Association* 14(6), 1372–1400.
- Benjamin, D. J., S. A. Brown, and J. M. Shapiro (2013). Who is behavioral? Cognitive ability and anomalous preferences. *Journal of the European Economic Association* 11(6), 1231–1255.
- Borghans, L., B. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences* 113(47), 13354–13359.
- Bouchard Jr., T. J. and J. C. Loehlin (2001). Genes, evolution, and personality. *Behavior genetics* 31(3), 243–273.
- Bowles, S., H. Gintis, and M. Osborne (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature* 39(4), 1137–1176.
- Canli, T. (2006). *Biology of personality and individual differences*. The Guilford Press.
- Castex, G. and E. Dechter (2014). The Changing Roles of Education and Ability in Wage Determination. *Journal of Labor Economics* 32(4).
- Cawley, J., K. Conneely, J. Heckman, and E. Vytlačil (1997). Cognitive ability, wages, and meritocracy. In *Intelligence, genes, and success*, pp. 179–192. Springer.
- Cobb-Clark, D. A. and S. Schurer (2012). The stability of big-five personality traits. *Economics Letters* 115(1), 11 – 15.
- Cubel, M., A. Nuevo, S. S. Chiquero, and M. V. Fernandez (2016). Do personality traits affect productivity? Evidence from the laboratory. *Economic Journal* 126(592), 654–681.

- Daly, H. and V. Lavy (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *The American Economic Review* 99(4), 1384–1414.
- Deming, D. J. (2017, 06). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics* 132(4), 1593–1640.
- DeYoung, C. G., J. B. Hirsh, M. S. Shane, X. Papademetris, N. Rajeevan, and J. R. Gray (2010). Testing predictions from personality neuroscience. Brain structure and the big five. *Psychological science : a journal of the American Psychological Society / APS* 21(6), 820–828.
- Di Chiacchio, C., S. De Stasio, and C. Fiorilli (2016). Examining how motivation toward science contributes to omitting behaviours in the Italian PISA 2006 sample. *Learning and Individual Differences* 50, 56–63.
- Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010). Are Risk Aversion and Impatience Related to Cognitive Ability? *The American Economic Review* 100(3), 1238–1260.
- Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of personality and social psychology* 37(7), 1097–1126.
- Fletcher, J. M. (2013). The effects of personality traits on adult labor market outcomes: Evidence from siblings. *Journal of Economic Behavior Organization* 89, 122–135.
- Freund, D. S. and D. A. Rock (1992). A preliminary investigation of pattern-marking in 1990 NAEP data. *Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 347 189.*
- Gneezy, U., M. Niederle, and A. Rustichini (2003, 08). Performance in Competitive Environments: Gender Differences\*. *The Quarterly Journal of Economics* 118(3), 1049–1074.

- Gneezy, U. and A. Rustichini (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics* 115(3), 791–810.
- Gneezy, U. and A. Rustichini (2004, May). Gender and Competition at a Young Age. *American Economic Review* 94(2), 377–381.
- Heckman, J., J. Stixrud, and S. Urzua (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics* 24(3), 411–482.
- Heckman, J. J. (1995). Lessons from the bell curve. *Journal of Political Economy* 103(5), 1091–1120.
- Heineck, G. and S. Anger (2010). The returns to cognitive abilities and personality traits in Germany. *Labour Economics* 17(3), 535 – 546.
- Herrnstein, R. J. and C. Murray (2010). *Bell curve: Intelligence and class structure in American life*. Simon and Schuster.
- Mueller, G. and E. Plug (2006). Estimating the Effect of Personality on Male and Female Earnings. *ILR Review* 60(1), 3–22.
- Neal, D. A. and W. R. Johnson (1996). The role of premarket factors in black-white wage differences. *Journal of Political Economy* 104(5), 869–895.
- Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics* 121(1), 297–317.
- Schlosser, A., Z. Neeman, and Y. Attali (2019, 05). Differential Performance in High Versus Low Stakes Tests: Evidence from the Gre Test\*. *The Economic Journal* 129(623), 2916–2948.
- Segal, C. (2012). Working when no one is watching: Motivation, test scores, and economic success. *Management Science* 58(8), 1438–1457.

- Speer, J. D. (2017). The gender gap in college major: Revisiting the role of pre-college factors. *Labour Economics* 44, 69–88.
- Weinberger, C. J. (2014). The Increasing Complementarity between Cognitive and Social Skills. *The Review of Economics and Statistics* 96(5), 849–861.
- Willingham, W. W. and N. S. Cole (1997). Gender and fair assessment., Lawrence Erlbaum Associates Publishers.
- Zax, J. and D. Rees (2002). IQ, Academic Performance, Environment, and Earnings. *Review of Economics and Statistics* 84, 600–616.



**Table 1: Summary statistics: subjective incentives, personality traits and work-effort attitudes**

	all		men		women	
	mean	sd	mean	sd	mean	sd
<b>Subjective incentives:</b>						
<b>Test-effort:</b>						
"I tried to do my best on the test."	4.06	1.08	4.01	1.11	4.11	1.06
<b>Motivation:</b>						
Test-takers motivation	7.21	2.34	6.87	2.43	7.55	2.21
Importance of money: "If you were not offered any money, would you still have taken the test?" (yes=1 / no=0)	0.65	0.48	0.58	0.49	0.72	0.45
N	6854		3452		3402	
<b>Personality traits:</b>						
extraverted, enthusiastic	5.30	1.36	5.23	1.38	5.37	1.34
anxious, easily upset	3.50	1.81	3.30	1.77	3.69	1.82
critical, quarrelsome	3.58	1.65	3.71	1.63	3.45	1.67
dependable, self-disciplined	6.11	1.04	6.05	1.09	6.17	1.00
reserved, quiet	3.90	1.95	4.10	1.95	3.71	1.93
sympathetic, warm	5.54	1.33	5.19	1.39	5.89	1.16
disorganized, careless	2.75	1.69	2.85	1.67	2.66	1.70
calm, emotionally stable	5.49	1.37	5.62	1.33	5.35	1.39
conventional, uncreative	2.78	1.61	2.75	1.59	2.80	1.64
<b>Work-effort attitudes:</b>						
"I do what is required, but rarely anything more."	2.38	1.68	2.49	1.71	2.26	1.63
"I have high standards and work toward them."	6.18	1.05	6.11	1.10	6.24	1.00
"I make every effort to do more than what is expected of me."	5.79	1.20	5.71	1.26	5.87	1.13
N	5533		2744		2789	

Note: Test-specific effort ("I tried to do my best on the test") values are between 1-5. The "importance of money" variable values are {0,1}. Test-takers motivation values are between 2-10. The construction of the motivation variable is described in Appendix Table A.1. Personality traits and work-effort attitudes are ranked on a 1-7 scale (from strongly disagree to strongly agree).

**Table 2: Summary statistics, by test-specific participation motivation**

	All	Because it's an important study	..what it's like to take test on comp.	To see how well I could do	To learn more about my interests	Family member wanted me to take it	To get the money	I had nothing else to do today
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AFQT	171.1 (29.56)	164.7 (32.11)	152.4 (30.50)	173.5 (28.28)	173.5 (27.25)	166.4 (29.31)	177.2 (27.14)	161.3 (34.51)
Age in 97	14.3 (1.48)	14.0 (1.47)	13.9 (1.44)	14.2 (1.48)	14.4 (1.47)	14.3 (1.48)	14.5 (1.44)	14.4 (1.50)
Male, {0,1}	0.51 (0.50)	0.50 (0.50)	0.52 (0.50)	0.42 (0.49)	0.44 (0.50)	0.53 (0.50)	0.60 (0.49)	0.59 (0.49)
Black, {0,1}	0.12 (0.33)	0.15 (0.36)	0.17 (0.37)	0.14 (0.35)	0.11 (0.31)	0.12 (0.32)	0.09 (0.28)	0.15 (0.36)
Hispanic, {0,1}	0.10 (0.30)	0.12 (0.32)	0.15 (0.36)	0.11 (0.32)	0.10 (0.30)	0.09 (0.29)	0.08 (0.26)	0.12 (0.33)
Father education	13.2 (3.01)	12.8 (2.99)	12.3 (3.02)	12.9 (2.84)	13.4 (3.08)	13.3 (2.95)	13.7 (2.99)	12.4 (3.10)
Mother education	13.2 (2.90)	13.0 (3.40)	12.5 (2.51)	13.1 (3.14)	13.4 (2.65)	13.4 (2.71)	13.6 (2.59)	12.7 (2.67)
Intact family, {0,1}	0.69 (0.46)	0.65 (0.48)	0.65 (0.48)	0.69 (0.46)	0.71 (0.46)	0.72 (0.45)	0.71 (0.46)	0.61 (0.49)
Metro status, {0,1}	0.57 (0.49)	0.56 (0.50)	0.56 (0.50)	0.57 (0.50)	0.58 (0.49)	0.60 (0.49)	0.58 (0.49)	0.50 (0.50)
ln Family income	10.8 (1.14)	10.7 (1.18)	10.7 (1.10)	10.8 (1.10)	10.9 (1.14)	10.9 (1.31)	10.9 (1.07)	10.6 (1.22)
N	4589	766	273	1084	846	302	1163	155
Education at 24	13.4 (2.54)	13.1 (2.57)	12.2 (2.38)	13.3 (2.49)	13.6 (2.45)	13.4 (2.56)	13.7 (2.57)	12.8 (2.48)
N	3223	531	176	794	603	217	776	126
Hourly wage at 24	13.9 (7.38)	13.1 (6.95)	13.3 (8.55)	13.8 (7.15)	14.0 (7.32)	14.1 (6.86)	14.7 (7.25)	13.5 (10.20)
N	3223	531	176	794	603	217	776	126

Note: All statistics are weighted by the cross-sectional weights. Test-specific motivation is measured by the first reason to participate in ASVAB. Mean and SD presented; SD in parantheses.

**Table 3: Test-specific effort and motivation, N=3872**

	effort		motivation	
	(1)	(2)	(3)	(4)
Male	-0.101*** (0.035)	-0.099*** (0.038)	-0.650*** (0.080)	-0.614*** (0.087)
Black	-0.144*** (0.049)	-0.149*** (0.050)	0.243** (0.102)	0.199* (0.106)
Hispanic	-0.047 (0.049)	-0.044 (0.049)	0.179 (0.115)	0.145 (0.115)
Parental education	0.004 (0.007)	0.002 (0.007)	-0.023 (0.016)	-0.022 (0.016)
Metro status	0.051 (0.036)	0.048 (0.036)	0.053 (0.082)	0.061 (0.082)
Intact family	0.086** (0.040)	0.084** (0.040)	0.099 (0.089)	0.084 (0.088)
Log Family income	-0.006 (0.017)	-0.007 (0.017)	0.003 (0.037)	-0.006 (0.037)
<b>Personality traits and attitudes:</b>				
extraverted, enthusiastic		0.006 (0.016)		0.077** (0.037)
anxious, easily upset		-0.017 (0.012)		-0.014 (0.027)
critical, quarrelsome		0.003 (0.012)		-0.014 (0.027)
dependable, self-disciplined		0.031 (0.021)		-0.012 (0.047)
reserved, quiet		0.013 (0.010)		0.039 (0.024)
sympathetic, warm		0.008 (0.015)		0.024 (0.036)
disorganized, careless		0.01 (0.012)		-0.016 (0.027)
calm, emotionally stable		-0.017 (0.017)		0.015 (0.038)
conventional, uncreative		-0.011 (0.012)		0.015 (0.026)
I do what is required, rarely more		-0.02 (0.014)		-0.036 (0.030)
High standards at work		0.008 (0.023)		-0.031 (0.054)
I make every effort to do more		-0.027 (0.020)		0.136*** (0.046)
const.	4.119*** (0.181)	4.140*** (0.266)	7.652*** (0.415)	6.620*** (0.614)
R2	0.007	0.01	0.022	0.033

Note: All statistics are weighted using the cross-sectional weights. "Effort" takes the values between 1-5; "motivation" takes the values between 2-10. Coefficients presented, standard errors in parentheses. Significance levels are indicated as follows: +15%; \* 10%; \*\* 5%; \*\*\* 1%.

**Table 4: AFQT score and noncognitive channels**

	men, N=1921			women, N=1951		
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Subjective incentives:</b>						
<b>Test effort:</b>						
Test-effort = 2	-0.541*** (0.154)	-0.495*** (0.153)	-0.429*** (0.146)	-0.367* (0.200)	-0.302 (0.194)	-0.308 (0.193)
Test-effort = 3	-0.280** (0.110)	-0.225** (0.110)	-0.238** (0.108)	-0.132 (0.121)	-0.097 (0.120)	-0.124 (0.119)
Test-effort = 4	0.003 (0.092)	0.047 (0.091)	0.044 (0.089)	0.061 (0.089)	0.074 (0.088)	0.053 (0.087)
Test-effort = 5	0.312*** (0.091)	0.339*** (0.090)	0.339*** (0.088)	0.236*** (0.088)	0.238*** (0.087)	0.214** (0.087)
<b>Test-motivation:</b>						
Importance of money	0.142*** (0.045)	0.126*** (0.044)	0.110** (0.043)	0.158*** (0.049)	0.150*** (0.048)	0.157*** (0.047)
1: .. it's an important study	-0.175*** (0.056)	-0.178*** (0.057)	-0.175*** (0.056)	-0.125** (0.055)	-0.131** (0.054)	-0.125** (0.053)
1: ..what's like take test on comp.	-0.381*** (0.090)	-0.376*** (0.089)	-0.375*** (0.088)	-0.452*** (0.095)	-0.433*** (0.095)	-0.432*** (0.093)
1: Family member wanted me...	-0.213** (0.084)	-0.199** (0.081)	-0.201*** (0.078)	-0.214** (0.086)	-0.205** (0.087)	-0.177** (0.087)
1: To get the money	0.243*** (0.056)	0.214*** (0.055)	0.193*** (0.054)	0.298*** (0.056)	0.293*** (0.055)	0.283*** (0.054)
1: I had nothing else to do today	-0.173 (0.129)	-0.212* (0.125)	-0.228* (0.120)	-0.194 (0.132)	-0.200 (0.131)	-0.206 (0.134)
2: .. it's an important study	0.029 (0.067)	0.020 (0.066)	-0.001 (0.063)	-0.238*** (0.068)	-0.229*** (0.068)	-0.223*** (0.067)
2: ..what's like take test on comp.	-0.337*** (0.086)	-0.348*** (0.086)	-0.321*** (0.086)	-0.310*** (0.097)	-0.323*** (0.095)	-0.277*** (0.094)
2: Family member wanted me...	-0.083 (0.066)	-0.085 (0.065)	-0.088 (0.063)	-0.014 (0.069)	-0.019 (0.069)	-0.028 (0.068)
2: To get the money	0.194*** (0.053)	0.182*** (0.052)	0.177*** (0.050)	0.182*** (0.052)	0.176*** (0.051)	0.163*** (0.051)
2: I had nothing else to do today	-0.122* (0.070)	-0.120* (0.068)	-0.126* (0.066)	-0.225*** (0.081)	-0.210*** (0.078)	-0.188** (0.078)

Continued next page

Table 4 continued

<b>Personality traits and attitudes:</b>						
extraverted, enthusiastic	-0.042***	-0.039**			-0.041**	-0.041**
	(0.016)	(0.016)			(0.017)	(0.017)
anxious, easily upset	-0.081***	-0.071***			-0.071***	-0.062***
	(0.013)	(0.013)			(0.013)	(0.013)
critical, quarrelsome	0.024*	0.023*			-0.006	-0.007
	(0.013)	(0.012)			(0.013)	(0.013)
dependable, self-disciplined	-0.018	-0.029			0.033	0.025
	(0.020)	(0.021)			(0.024)	(0.024)
reserved, quiet	-0.008	-0.004			0.001	0.004
	(0.010)	(0.010)			(0.011)	(0.011)
sympathetic, warm	-0.022	-0.018			0.003	0.002
	(0.015)	(0.014)			(0.019)	(0.019)
disorganized, careless	0.020*	0.015			0.020	0.025**
	(0.012)	(0.012)			(0.013)	(0.013)
calm, emotionally stable	0.018	0.020			0.008	0.009
	(0.018)	(0.018)			(0.017)	(0.017)
conventional, uncreative	-0.020	-0.010			-0.026**	-0.016
	(0.012)	(0.012)			(0.012)	(0.012)
Do what is required, rarely more		-0.109***				-0.080***
		(0.013)				(0.014)
High standards at work		0.039*				0.082***
		(0.023)				(0.024)
I make every effort to do more		-0.124***				-0.105***
		(0.019)				(0.024)
const.	-2.130***	-1.520***	-0.642**	-2.567***	-2.200***	-1.887***
	(0.225)	(0.290)	(0.303)	(0.232)	(0.332)	(0.361)
R2	0.357	0.382	0.411	0.318	0.340	0.359

Note: All statistics are weighted using the cross-sectional weights. All estimations include controls for parental education, intact family indicator, family income, metro status, Black, Hispanic and constant. "1" and "2" indicate primary and secondary test-takers motivation categories. Omitted test participation reason 1 and reason 2 categories are "To see how well I could do on the test" and "To learn more about my interests". Standard errors clustered at individual level. Significance levels are indicated as follows: +15%; \* 10%; \*\* 5%; \*\*\* 1%.

**Table 5: Wages and cognitive and noncognitive channels, men, N=9195**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AFQT	0.056*** (0.010)	0.054*** (0.011)	0.053*** (0.011)	0.045*** (0.011)	0.022** (0.011)	0.020* (0.011)	0.021* (0.011)	0.014 (0.011)
Education					0.041*** (0.005)	0.041*** (0.005)	0.039*** (0.005)	0.039*** (0.005)
<b>Subjective incentives:</b>								
<b>Test effort:</b>								
Test-effort = 2		0.013 (0.068)	0.010 (0.066)	0.002 (0.067)		0.031 (0.068)	0.025 (0.066)	0.018 (0.066)
Test-effort = 3		0.118*** (0.043)	0.115*** (0.042)	0.114*** (0.042)		0.126*** (0.042)	0.120*** (0.041)	0.118*** (0.041)
Test-effort = 4		0.081** (0.034)	0.077** (0.033)	0.077** (0.033)		0.099*** (0.033)	0.092*** (0.033)	0.091*** (0.033)
Test-effort = 5		0.060* (0.035)	0.058* (0.034)	0.057 (0.034)		0.075** (0.034)	0.071** (0.034)	0.069** (0.034)
<b>Test-motivation:</b>								
Importance of money		-0.060*** (0.022)	-0.060*** (0.021)	-0.056*** (0.021)		-0.056*** (0.022)	-0.056*** (0.021)	-0.053** (0.021)
1: .. it's an important study		-0.003 (0.028)	0.001 (0.027)	0.003 (0.027)		-0.001 (0.028)	0.003 (0.027)	0.004 (0.027)
1: ..what's like take test on comp.		-0.035 (0.033)	-0.020 (0.032)	-0.022 (0.032)		-0.033 (0.034)	-0.019 (0.032)	-0.022 (0.032)
1: Family member wanted me...		0.011 (0.037)	0.007 (0.037)	0.014 (0.037)		0.005 (0.035)	0.003 (0.035)	0.010 (0.035)
1: To get the money		-0.017 (0.028)	-0.012 (0.028)	-0.005 (0.028)		-0.020 (0.028)	-0.014 (0.027)	-0.008 (0.027)
1: I had nothing else to do today		-0.067 (0.041)	-0.057 (0.041)	-0.048 (0.041)		-0.055 (0.041)	-0.047 (0.040)	-0.039 (0.040)
2: .. it's an important study		-0.023 (0.035)	-0.024 (0.034)	-0.020 (0.034)		-0.023 (0.035)	-0.024 (0.034)	-0.020 (0.034)
2: ..what's like take test on comp.		-0.039 (0.036)	-0.029 (0.036)	-0.017 (0.037)		-0.052 (0.036)	-0.040 (0.036)	-0.028 (0.037)
2: Family member wanted me...		0.006 (0.035)	0.003 (0.034)	0.007 (0.033)		0.012 (0.034)	0.010 (0.033)	0.014 (0.032)
2: To get the money		0.040 (0.026)	0.039 (0.025)	0.040 (0.025)		0.035 (0.025)	0.034 (0.025)	0.035 (0.024)
2: I had nothing else to do today		0.014 (0.036)	0.029 (0.036)	0.028 (0.036)		0.025 (0.036)	0.038 (0.036)	0.037 (0.035)

Continued next page

Table 5 continued

<b>Personality traits and attitudes:</b>									
extraverted, enthusiastic				0.004	-0.001			0.004	0.000
				(0.007)	(0.007)			(0.007)	(0.007)
anxious, easily upset				-0.008	-0.007			-0.006	-0.005
				(0.006)	(0.006)			(0.006)	(0.006)
critical, quarrelsome				0.011*	0.013**			0.012**	0.013**
				(0.006)	(0.006)			(0.006)	(0.006)
dependable, self-disciplined				0.049***	0.039***			0.048***	0.039***
				(0.009)	(0.009)			(0.008)	(0.009)
reserved, quiet				-0.014***	-0.013***			-0.012***	-0.011**
				(0.005)	(0.005)			(0.005)	(0.005)
sympathetic, warm				-0.016**	-0.019***			-0.020***	-0.022***
				(0.007)	(0.007)			(0.007)	(0.007)
disorganized, careless				-0.009*	-0.007			-0.009	-0.007
				(0.006)	(0.006)			(0.006)	(0.006)
calm, emotionally stable				0.006	0.003			0.002	-0.001
				(0.008)	(0.008)			(0.008)	(0.008)
conventional, uncreative				0.004	0.008			0.003	0.007
				(0.006)	(0.006)			(0.006)	(0.006)
Do what is required, rarely more					-0.020***				-0.021***
					(0.006)				(0.006)
High standards at work					0.020*				0.017
					(0.011)				(0.011)
I make every effort to do more					0.001				0.002
					(0.009)				(0.009)
const.		1.244***	1.224***	1.032***	1.032***	0.977***	0.939***	0.782***	0.799***
		(0.074)	(0.084)	(0.119)	(0.124)	(0.085)	(0.096)	(0.125)	(0.130)
R2		0.089	0.099	0.119	0.125	0.108	0.117	0.136	0.142

Note: The dependent variable is log real hourly wage. All statistics are weighted using the cross-sectional weights. All estimations include age, metro status, Black, Hispanic, and constant. "1" and "2" indicate primary and secondary test-takers motivation categories. Omitted reason to take the test category is the group of respondents who indicated the reason to be "To see how well I could do on the test" or "To learn more about my interests". Lowest "test-effort" category is omitted. Standard errors clustered at individual level. Significance levels are indicated as follows: \* 10%; \*\* 5%; \*\*\* 1%.

**Table 6: Wages and cognitive and noncognitive channels, channels, women, N=8295**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AFQT	0.107*** (0.010)	0.108*** (0.011)	0.113*** (0.011)	0.111*** (0.011)	0.035*** (0.011)	0.038*** (0.012)	0.046*** (0.012)	0.044*** (0.012)
Education					0.073*** (0.005)	0.073*** (0.005)	0.069*** (0.005)	0.069*** (0.005)
<b>Subjective incentives:</b>								
<b>Test effort:</b>								
Test-effort = 2		0.008 (0.066)	0.002 (0.063)	0.008 (0.060)		0.061 (0.066)	0.057 (0.063)	0.063 (0.060)
Test-effort = 3		0.014 (0.055)	0.024 (0.054)	0.031 (0.053)		0.050 (0.050)	0.059 (0.050)	0.065 (0.049)
Test-effort = 4		0.027 (0.035)	0.028 (0.035)	0.036 (0.035)		0.048 (0.033)	0.050 (0.033)	0.057* (0.033)
Test-effort = 5		0.012 (0.035)	0.009 (0.035)	0.015 (0.035)		0.036 (0.033)	0.035 (0.033)	0.040 (0.033)
<b>Test-motivation:</b>								
Importance of money		-0.025 (0.028)	-0.026 (0.028)	-0.024 (0.027)		-0.012 (0.026)	-0.011 (0.026)	-0.010 (0.026)
1: .. it's an important study		0.032 (0.027)	0.035 (0.027)	0.034 (0.027)		0.028 (0.027)	0.032 (0.027)	0.031 (0.027)
1: ..what's like take test on comp.		0.019 (0.048)	0.021 (0.047)	0.020 (0.047)		0.028 (0.045)	0.027 (0.045)	0.024 (0.045)
1: Family member wanted me...		0.008 (0.041)	0.000 (0.041)	0.006 (0.040)		0.015 (0.040)	0.011 (0.040)	0.016 (0.039)
1: To get the money		0.050 (0.035)	0.058* (0.034)	0.058* (0.034)		0.032 (0.033)	0.041 (0.032)	0.042 (0.032)
1: I had nothing else to do today		0.010 (0.061)	0.024 (0.060)	0.026 (0.062)		0.000 (0.058)	0.012 (0.057)	0.015 (0.058)
2: .. it's an important study		-0.020 (0.038)	-0.020 (0.036)	-0.021 (0.036)		-0.024 (0.035)	-0.025 (0.034)	-0.026 (0.034)
2: ..what's like take test on comp.		0.021 (0.033)	0.023 (0.033)	0.023 (0.033)		0.009 (0.033)	0.008 (0.032)	0.008 (0.032)
2: Family member wanted me...		0.036 (0.043)	0.035 (0.042)	0.033 (0.041)		0.031 (0.043)	0.031 (0.042)	0.029 (0.041)
2: To get the money		-0.012 (0.030)	-0.007 (0.030)	-0.005 (0.029)		-0.015 (0.029)	-0.010 (0.029)	-0.008 (0.028)
2: I had nothing else to do today		-0.071* (0.042)	-0.047 (0.043)	-0.044 (0.042)		-0.047 (0.040)	-0.030 (0.041)	-0.028 (0.041)

Continued next page



Table 6 continued

<b>Personality traits and attitudes:</b>								
extraverted, enthusiastic			0.019**	0.016**			0.009	0.006
			(0.008)	(0.008)			(0.008)	(0.008)
anxious, easily upset			-0.002	-0.001			0.001	0.002
			(0.006)	(0.006)			(0.006)	(0.006)
critical, quarrelsome			0.001	0.002			0.001	0.002
			(0.006)	(0.006)			(0.006)	(0.006)
dependable, self-disciplined			0.026***	0.018*			0.021**	0.014
			(0.010)	(0.010)			(0.010)	(0.011)
reserved, quiet			0.004	0.004			0.002	0.002
			(0.006)	(0.006)			(0.005)	(0.005)
sympathetic, warm			-0.024***	-0.027***			-0.019**	-0.021***
			(0.008)	(0.008)			(0.008)	(0.008)
disorganized, careless			-0.031***	-0.028***			-0.029***	-0.026***
			(0.006)	(0.006)			(0.006)	(0.006)
calm, emotionally stable			0.003	0.002			0.003	0.002
			(0.008)	(0.008)			(0.008)	(0.008)
conventional, uncreative			0.005	0.007			0.004	0.006
			(0.006)	(0.006)			(0.006)	(0.006)
Do what is required, rarely more				-0.005				-0.007
				(0.006)				(0.006)
High standards at work				0.023**				0.015
				(0.011)				(0.011)
I make every effort to do more				0.007				0.008
				(0.012)				(0.011)
const.	0.965***	0.965***	0.938***	0.838***	0.591***	0.557***	0.580***	0.514***
	(0.081)	(0.091)	(0.125)	(0.133)	(0.088)	(0.094)	(0.127)	(0.134)
R2	0.115	0.12	0.14	0.143	0.177	0.18	0.193	0.195

Note: The dependent variable is log real hourly wage. All statistics are weighted using the cross-sectional weights. All estimations include age, metro status, Black, Hispanic, and constant. "1" and "2" indicate primary and secondary test-takers motivation categories. Omitted reason to take the test category is the group of respondents who indicated the reason to be "To see how well I could do on the test" or "To learn more about my interests". Lowest "test-effort" category is omitted. Standard errors clustered at individual level. Significance levels are indicated as follows: \* 10%; \*\* 5%; \*\*\* 1%.

**Table 7: Difference in performance in high-stakes vs. low-stakes tests, AFQT vs. SAT**

	all (1)	low effort (2)	high effort (3)	all (4)	low effort (5)	high effort (6)
Male	-2.685** (1.228)	-3.697** (1.679)	-1.299 (1.818)	-2.702** (1.286)	-3.427** (1.711)	-1.949 (1.924)
Black	1.572 (1.577)	-0.561 (2.049)	4.083* (2.458)	1.362 (1.601)	-0.765 (2.023)	3.494 (2.571)
Hispanic	1.726 (1.944)	1.509 (2.005)	1.810 (4.031)	1.722 (2.007)	1.524 (2.058)	1.593 (4.271)
Parental education	-0.478* (0.247)	-0.554 (0.348)	-0.446 (0.350)	-0.481* (0.249)	-0.590* (0.348)	-0.462 (0.355)
Metro status	-2.381* (1.305)	-0.700 (1.831)	-4.289** (1.864)	-2.043 (1.322)	-0.305 (1.844)	-3.949** (1.913)
Intact family	-2.617* (1.474)	-3.280* (1.925)	-1.794 (2.304)	-2.644* (1.498)	-3.421* (1.960)	-1.964 (2.352)
<b>Personality traits and attitudes:</b>						
extraverted, enthusiastic				0.575 (0.535)	0.587 (0.762)	0.683 (0.740)
anxious, easily upset				0.045 (0.428)	-0.130 (0.558)	0.365 (0.679)
critical, quarrelsome				0.184 (0.425)	-0.291 (0.625)	0.727 (0.557)
dependable, disciplined				-0.101 (0.753)	1.016 (1.134)	-1.070 (0.973)
reserved, quiet				-0.074 (0.394)	-0.335 (0.536)	0.218 (0.571)
sympathetic, warm				-0.087 (0.529)	-0.029 (0.756)	-0.334 (0.793)
disorganized, careless				0.305 (0.411)	0.425 (0.543)	0.301 (0.633)
calm, emotionally stable				0.354 (0.588)	0.421 (0.736)	0.483 (0.969)
conventional, uncreative				-0.244 (0.466)	0.711 (0.597)	-1.225* (0.709)
I do what is required				0.387 (0.576)	1.200 (0.798)	-0.853 (0.868)
High standards at work				-0.073 (0.874)	1.179 (1.193)	-1.972 (1.315)
I make effort to do more				0.593 (0.705)	0.294 (1.046)	0.911 (0.993)
N	970	533	437	970	533	437
R2	0.028	0.031	0.033	0.032	0.05	0.057

Note: The dependent variable is the difference between AFQT score percentile and SAT score percentile. All estimations include a constant. All statistics are weighted using the cross-sectional weights. Significance levels are indicated as follows: \* 10%; \*\* 5%; \*\*\* 1%.

## Appendix

Figure A.1: Distribution of test-specific effort, percent in each category

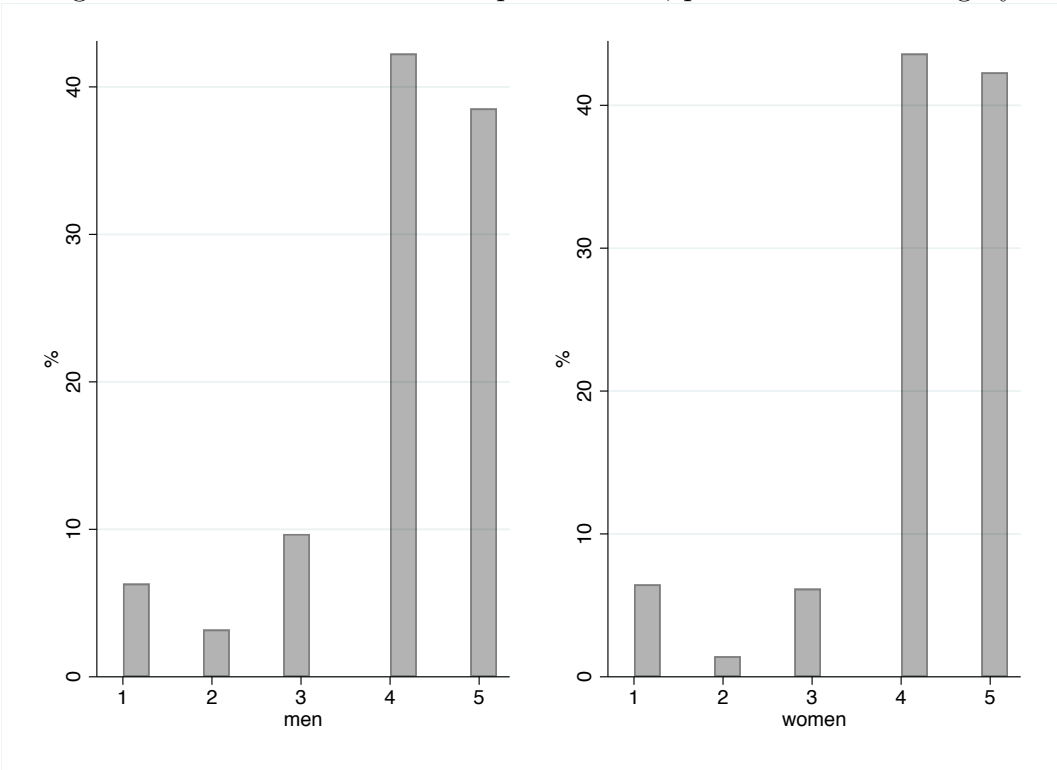


Figure A.2: Distributions of AFQT scores by effort

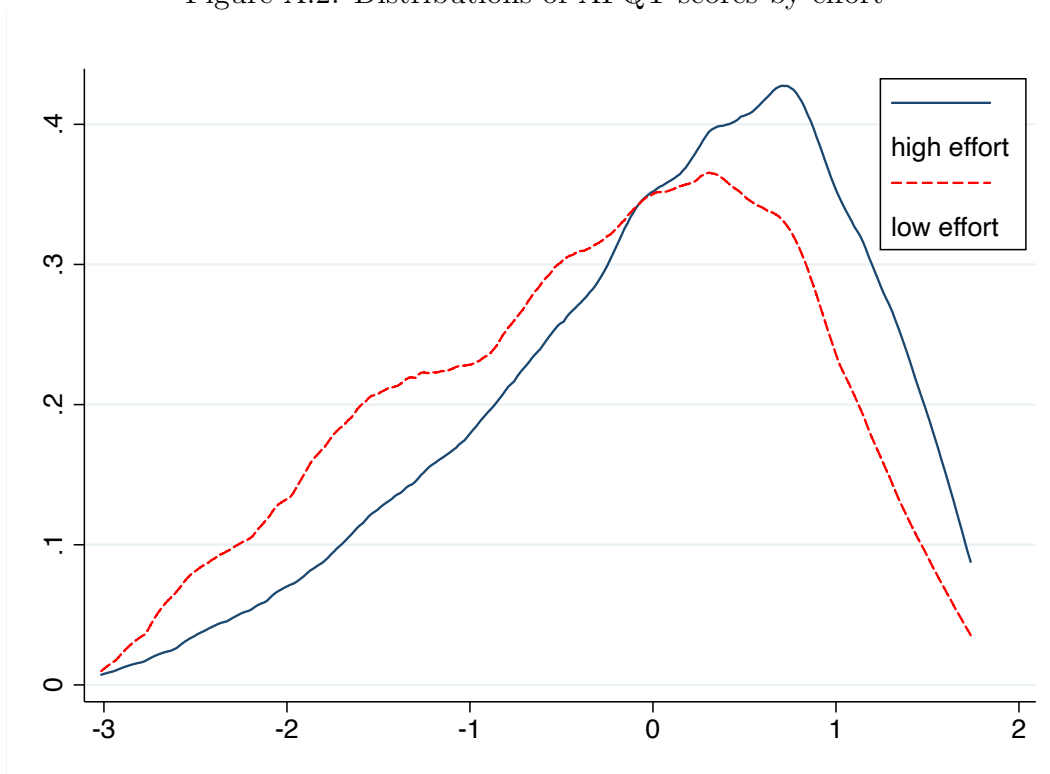


Figure A.3: Distributions of AFQT scores by motivation ranking

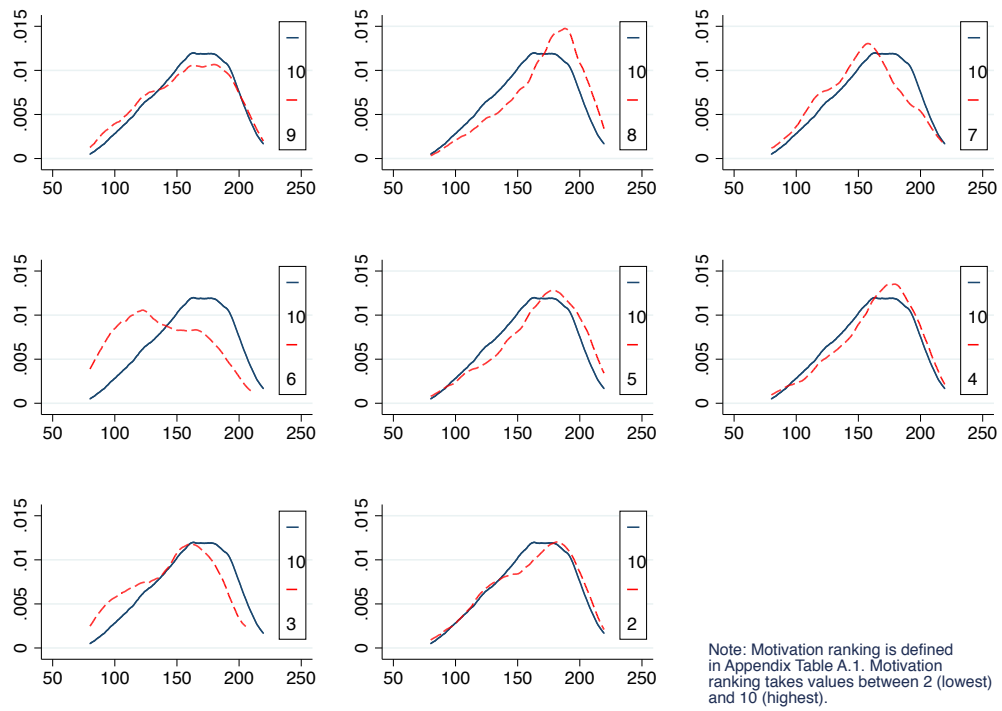
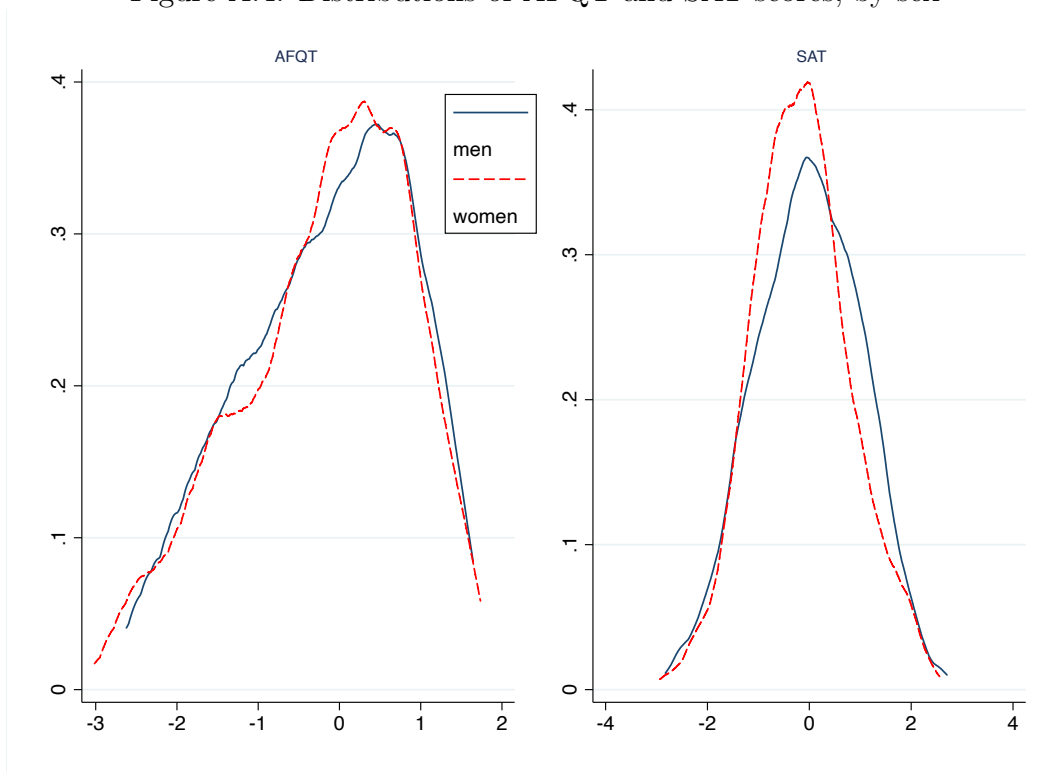


Figure A.4: Distributions of AFQT and SAT scores, by sex



**Appendix Table A.1: Ranking of test-takers motivation**

	R2:						
	Because it's an important study	R2: ..what it's like to take test on comp.	R2: To see how well I could do..	R2: To learn about my interests	R2: Family member wanted me take it	R2: To get the money	R2: I had nothing else to do today
R1: Because it's an important study		6	9	9	6	5	4
R1: ..what it's like to take test on comp.	6		8	8	6	5	4
R1: To see how well I could do..	9	8		10	9	8	7
R1: To learn about my interests	9	8	10		9	8	7
R1: Family member wanted me take it	6	6	9	9		4	3
R1: To get the money	5	5	8	8	4		2
R1: I had nothing else to do today	4	4	7	7	3	2	

Note: Motivation ranking constructed using the test-takers reasons to participate in ASVAB. R1 and R2 denote test-specific participation reasons. Each cell in Table A.1 denotes value for a specific combination of R1 and R2. The motivation variable takes values from 2 to 10.