ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 62/25

Job Mismatch and Early Career Success

Julie Berry Cullen, Gordon B. Dahl, Richard De Thorpe

www.rfberlin.com SEPTEMBER 2025

Job Mismatch and Early Career Success

Authors

Julie Berry Cullen, Gordon B. Dahl, Richard De Thorpe

Reference

JEL Codes: J24

Keywords: Job mismatch, skill acquisition, retention, promotion

Recommended Citation: Julie Berry Cullen, Gordon B. Dahl, Richard De Thorpe (2025): Job Mismatch and Early Career Success. RFBerlin Discussion Paper No. 62/25

Access

Disclaimer

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

Job Mismatch and Early Career Success

Julie Berry Cullen[*], Gordon B. Dahl[†], Richard De Thorpe[‡§]

August 29, 2025

Abstract

How does being over- or underqualified at the beginning of a worker's career affect skill acquisition, retention, and promotion? Despite the importance of mismatch for the labor market, self-selection into jobs has made estimating these effects difficult. We overcome endogeneity concerns in the context of the US Air Force, which allocates new enlistees to over 130 different jobs based, in part, on test scores. Using these test scores, we create simulated job assignments based on factors outside of an individual's control: the available slots in upcoming training programs and the quality of other recruits entering at the same time. These factors create quasi-random variation in job assignment and hence how cognitively demanding an individual's job is relative to their own ability. We find that being overqualified for a job causes higher attrition, both during technical training and afterward when individuals are working in their assigned jobs. It also results in more behavioral problems, worse performance evaluations, and lower scores on general knowledge tests about the military taken by all workers. On the other hand, overqualification results in better performance relative to others in the same job: job-specific test scores rise both during technical training and while on the job, and these individuals are more likely to be promoted. Combined, these patterns suggest that overqualified individuals are less motivated, but still outperform others in their same job. Underqualification results in a polar opposite set of findings, suggesting these individuals are motivated to put forth more effort, but still struggle to compete when judged relative to others. Consistent with differential incentives, individuals who are overqualified are in jobs which are less valuable in terms of outside earnings potential, while the reverse is true for those who are underqualified.

JEL code: J24; Keywords: Job mismatch, skill acquisition, retention, promotion

[*]UC San Diego, NBER, CESifo

[†]UC San Diego, NBER, CESifo, CEPR, IZA, RFBerlin

[‡]Princeton University

[§]We thank Letian Yin for valuable research assistance and Jeff Campbell, Alexander Chesney, and David Lee for extensive comments. We also thank seminar participants for valuable feedback at ANU, Chicago Fed, Chilean Economic Society, e61 Institute, NYU, Princeton, Texas A&M, Tinbergen Institute, UCI, UCSD, and UNSW.

1 Introduction

A classic question in labor economics is how the quality of a match between a worker and their job affects performance.[1] One key dimension of mismatch is a worker's own cognitive ability relative to the cognitive demands of the job. A worker's relative cognitive ability could affect skill acquisition and productivity, and hence career progression. Those who are overqualified may easily master the necessary job-specific knowledge and skills, while those who are underqualified may struggle to learn. However, mismatch could also affect outcomes through its impact on effort and engagement. For someone who is overqualified, it could be demotivating to be assigned to a job which is not challenging, or alternatively inspiring to be able to easily excel. For someone who is underqualified, being placed in a job which is difficult may cause them to give up, or rise to the challenge and invest more effort because of the opportunities it provides.

Despite the theoretical importance of ability mismatch, it has proven difficult to estimate the causal implications for labor market outcomes. The key challenge is that individuals self select into occupations. While it may appear that workers are mismatched based on observable characteristics, their unobserved characteristics could make them well-suited for the occupations they choose. In this paper, we study how being over- or underqualified at the beginning of a worker's career affects job skill acquisition (both during vocational and on-the-job training), job retention, and promotion. We overcome endogeneity concerns by leveraging quasi-random variation in job assignments.

Our context is the United States Air Force (USAF), which is often considered the most competitive branch of the military because of the high skill demands and intensive training required (Department of Defense 2019). There are over 130 different occupational specialties for enlisted personnel, ranging from aerospace mechanic, radar technician, security personnel, to information management specialist. In this sense, the USAF is similar to a large diversified

[1] For early reviews, see Levy and Murnane (1992) and Sattinger (1993).

company. As opposed to most companies, however, individuals have less control over which job they get. Further, the USAF provides credit-bearing technical education that blends classroom and hands-on training for each of the assigned jobs. In this sense it is like a community college, but with a guaranteed job upon completion of training.

We use the Armed Forces Qualifying Test (AFQT), which measures math and verbal ability, as our measure of cognitive ability. To define ability mismatch, we compare an individual's AFQT percentile relative to that of the typical person placed in the same job. Characterizing jobs using the average ability of incumbents aligns with a prominent branch of the prior literature.[2] We calculate how much better (*ability surplus*) or worse (*ability deficit*) an individual is along this dimension and link these measures to rich administrative outcome data.

We address the endogeneity issue by exploiting the unique way recruits are sorted into occupations. Individuals do not apply for specific job postings. Instead, jobs are allocated based on an individual's abilities and preferences, which jobs are currently available, and the pool of competing candidates for those jobs. To isolate the quasi-random component, we simulate job assignments based on a strict mapping from own ability rank to job ability rank within the cohort that starts technical training at the same time. We use this information to construct instruments for ability surplus and ability deficit. Predicted job assignments are plausibly orthogonal to any unobserved factors, such as preferences or personality, that could influence actual job assignments. Importantly, we condition flexibly on an individual's own AFQT score, so that the only identifying variation in the instrument comes from variation in the arrival of jobs and jobseekers.

Our data include extensive performance-related information, following the 2002-2006 entry cohorts for up to 5 years. Our short-run outcomes measure performance in technical training

[2] For example, Fredriksson et al. (2018), McGowan and Andrews (2017), and Perry et al. (2014) similarly define job skill requirements using measures such as the average cognitive scores of workers in an occupation.

for the assigned job, which occurs after basic training and typically lasts for 3-4 months. For those who are overqualified, a 10 percentage point (pp) increase in ability surplus decreases the probability of graduation by 1.5 pp. For context, the mean dropout rate is 5.6% and a 10 pp increase in ability surplus equals roughly one standard deviation. In contrast, while ability deficit reduces on-time graduation, it increases delayed graduation, for no net effect on graduation. Turning to achievement, the probability of graduating and scoring above average on technical training tests rises for individuals who are overqualified (2.8 pp for a 10 pp increase in ability surplus), and drops for individuals who are underqualified (-9.6 pp for a 10 pp increase in ability deficit).

In the medium run, we measure retention and behavior outcomes 2 and 3 years after the beginning of an individual's career. For both outcomes, ability surplus is detrimental and ability deficit is beneficial. Three years out, 14% of individuals have separated for reasons related to poor performance. A 10 pp larger ability surplus increases attrition by 4.9 pp. In contrast, being equivalently underqualified decreases attrition by 3.9 pp. Substantial effects are also found for criminal behavior.

In the longer run, we observe several measures related to promotion 4-5 years after entry. Promotion is a tournament that depends on exam and performance evaluation scores, with only 18% promoted in our sample. The headline result is that those with a 10 pp larger ability surplus are 4.5 pp more likely to be promoted, while those with a similar increase in ability deficit are 6.1 pp less likely. These findings line up with estimates showing that ability surplus has a positive impact and ability deficit a negative impact on job-specific exam scores, similar to the patterns found for short-run achievement. Notably, those with a surplus have a promotion advantage despite underperforming on the other inputs, mirroring the negative effects found in the medium run. They are less likely to persist long enough to be promotion eligible, score lower on general military knowledge exams, and receive worse performance evaluations. In a similar vein, the disadvantage for those with a deficit exists despite improved performance on these inputs.

Across the time horizons, OLS estimates differ sharply from our baseline IV estimates. This suggests that even at the beginning of individuals' careers, there is a large amount of unobservable selection. The qualitative findings based on IV are robust to including saturated controls for own ability, adding controls for cohort peer quality, using quintiles as the functional form for mismatch, and using a broader definition of cohorts to construct the instruments. Examining heterogeneity, we find ability deficits lead to worse outcomes for female and Black individuals relative to the overall population.

Combined, these results paint a rich portrait of how mismatch affects individuals at the beginning of their working lives. A benefit of being overqualified is greater mastery of job-specific skills relative to others in the same occupation. However, being placed in a less demanding job appears to be demotivating, as evidenced by higher attrition, more behavioral problems, worse performance evaluations, and lower general military knowledge. The polar opposite is true for underqualified individuals. Being put in more challenging jobs appears to motivate underqualified individuals to put forth more effort, but they struggle to compete when it comes to acquiring job skills.

While we cannot directly observe motivation, we do have a measure for the incentive to invest in one's assigned job related to earnings potential. Linking jobs in the military to civilian earnings using Standard Occupational Codes, we estimate that a 10 pp increase in ability surplus is predicted to reduce future outside earnings by 10%, while a similar increase in ability deficit is associated with a 13% increase. In other words, relative to individuals of similar ability, those who are overqualified are placed in jobs which have lower post-service earnings potential, while the reverse is true for those who are underqualified. These earnings penalties and rewards provide incentives consistent with the differential investments we observe.

Our study bridges two strands of research. The first is in the educational setting, where mismatch is typically measured by the difference in students' abilities (e.g., SAT scores) and the selectivity of the universities they attend (e.g., average SAT scores). Some of the earliest studies evaluate the consequences of affirmative action, which favors students with worse academic preparation who are thus overmatched to high quality institutions.[3] Other work exploits different factors that ration access to more and less selective colleges, such as admissions cutoffs, zeroing in on individuals at the margin (e.g., Anelli 2020; Hoekstra 2009; Mountjoy 2024; Zimmerman 2014, 2019). Most recently, researchers examine the interaction between academic ability and college quality across the ability distribution. Using different methods and measures of student ability and college quality, prominent studies in this vein reach different conclusions about which types of students are helped or hurt by mismatch (e.g., Black et al. 2023; Dillon and Smith 2020; Mountjoy and Hickman 2021).

Since our setting is about allocations to occupations, it is most akin to the educational research using quasi-experimental methods to study the role of college majors (Andrews et al. 2017; Bleemer and Mehta 2022; Hastings et al. 2013; Kirkeboen et al. 2016).[4] Because this research uses major-specific admissions cutoffs, those lucky enough to get into their preferred major will have lower ability than the typical student, and hence have an ability deficit by our definition. Nonetheless, this literature generally finds that individuals barely getting into their preferred major enjoy higher earnings than if they had not.

The second literature examines ability mismatch in the labor market. The literature on job mismatch began by looking at whether individuals are under- or overeducated for their jobs.[5] Subsequent work has made strides in how to measure mismatch along dimensions other than education, such as the skills needed for different jobs as reported by employees,

[3]See Arcidiacono and Lovenheim (2016) for a review, and Bleemer (2022) for a recent contribution.

[4]Interestingly, there is some evidence that students who are overmatched to more demanding universities shift to less demanding majors (Arcidiacono et al. 2016), and that disadvantaged students are more likely to be undermatched to their majors (Campbell et al. 2022). There is also work showing that choice of major may be as important as choice of institution (Eide et al. 2016).

[5]See Duncan and Hoffman (1981) and Hersch (1991) for two early influential papers.

employers, or experts.[6] Correlational studies generally find negative impacts from being overskilled, while there is less evidence on underskilling (see McGuinness et al. (2018) for a review). Two noteworthy papers in recent years study the effect of mismatch starting at career entry.[7] The first, Fredriksson et al. (2018), argues that little information is available for inexperienced workers, so there should be little selection bias for this group. Consistent with this hypothesis, the authors find empirical evidence that starting wages are unrelated to mismatch for inexperienced workers, but that separations and wage growth are highly sensitive. Using the history of jobs, Guvenen et al. (2020) find that initial mismatch increases occupational switching and depresses both wages and wage growth.[8] Relative to this literature, our contribution is to leverage quasi-experimental variation in job assignments to identify the causal impacts of mismatch in the labor market. To our knowledge, we are the first to do so.

The remainder of the paper proceeds as follows. We first describe our setting and data and then lay out our empirical approach. We then discuss our results, robustness and heterogeneity, and outside labor market incentives. The final section offers some concluding remarks.

2 Setting and Data

Our setting allows us to follow individuals from the start of their Air Force careers for up to 5 years, which is near the end of their first contracts (which usually have 4- or 6-year terms). They begin with 6 weeks of basic training, after which they move on to technical training for their specific jobs, which typically lasts 3-4 months. They then perform their

[6]While jobs are typically characterized by average reports, such as in O*NET, Deming and Kahn (2018) use job postings to document that skill requirements for the same job vary across employers.

[7]There is a related literature on the scarring effects of entering the labor market during a recession, where one of the mechanisms could be increased mismatch (e.g., Kahn 2010; Liu et al. 2016; Oreopoulos et al. 2012; Schwandt and von Wachter 2019).

[8]Other work uses structural models to study the dynamic evolution of sorting and the implications for labor market equilibria (e.g., Lindenlaub 2017; Lise and Postel-Vinay 2020).

jobs, advancing through the ranks and moving up the pay ladder during the duration of their contracts, before deciding whether to re-enlist or separate. We track how job assignment, and the resulting amount of ability mismatch, affects success both during vocational training and on the job.

In this section, we provide details on the enlistment, training, and job assignment process, as well as the administrative data we use. Our data cover those who enter during fiscal years 2002 through 2006. As background, this period involved moderate reductions in the Air Force, as needs and resources shifted to ground troops in Afghanistan and Iraq (Lytell et al. 2015).

2.1 Assignment to Jobs in the Air Force

An individual's job assignment process begins when they make a scheduled visit to a military entrance processing station. During the visit, the individual takes cognitive, physical, and medical exams and provides demographic and other background information. The cognitive exam is the Armed Services Vocational Aptitude Battery (ASVAB), which has nine subtests. The well-known AFQT score is a nationally-normed composite of the two math and two verbal sections. Viewed to be a measure of an individual's cognitive ability, the AFQT is much like a college entrance exam such as the SAT. Individuals need to score at least at the 31^{st} percentile to qualify for service.

Jobs, also known as Air Force Specialty Codes, are divided into four career fields: mechanical, administrative, general, and electronics. There are 134 different jobs. Career counseling is provided regarding the set of jobs with upcoming openings that an individual qualifies for based on their exam scores. The individual composes a preference list that includes around 10 jobs and a career field. After signing the initial contract, the individual then waits to be called to basic training, typically 5-6 months later.

When called to basic training, the individual learns either that they have been assigned a specific job or that they have been guaranteed a career field with the specific job being assigned later. Allocating a share to career fields, rather than specific jobs, preserves flexibility in filling upcoming slots for vocational training.[9] Around one-third of individuals were guaranteed a career field during our time period.[10] A few weeks into basic training, these individuals make a preference list over jobs still available to the cohort. An algorithm then matches individuals to available jobs based on qualifications and preferences. After graduating basic training, individuals then head off to begin technical training for their assigned jobs where they take a prescribed series of courses.

Unfortunately, we do not observe guaranteed job status or preference lists. However, those with a guaranteed career field or specific job are effectively in competition for the same set of jobs. Thus, we characterize "job markets" based on an individual's career field and timing of entry, given the importance of these to sorting individuals throughout the process. Cohort-specific shocks to job markets will form the basis for our instrumental variables approach, as described in detail below.

2.2 Data

We use rich administrative data to study a wide array of outcomes. A key advantage of our dataset is that it contains direct measures of skill accumulation and performance, in addition to worker retention. Other studies measure the productivity of worker-job matches through wages (e.g., Fredriksson et al. 2018; Guvenen et al. 2020). In our setting, wages are not a good proxy for productivity since earnings are a deterministic function of rank and step. Another advantage of our data is that we observe performance-related outcomes for up to 5 years after job market entry. Online Appendix A contains additional details on data sources, sample construction, and variable definitions; in this section we provide a quick overview of the outcomes we study over different horizons.

[9] This flexibility is important for at least two reasons: there is some attrition during basic training and some individuals need to repeat vocational training.

[10] We thank Alexander Chesney for providing this statistic based on internal USAF sources.

Short-run outcomes are observed during technical training, which is formal job-specific coursework that immediately follows graduation from basic training. Training typically lasts 3-4 months, and lasts more than 6 months only 11% of the time. A first measure of success is defined as graduating on time, which 80% of individuals achieve. This measure is consequential, as individuals who have to repeat training or who drop out impose substantial costs (Manacapilli et al. 2012). For those who ever graduate (94%), which includes those who graduate with delay (14%), we also observe final technical training grades.

Medium-run outcomes capture retention and misbehavior 2-3 years after entry. Given that initial contracts are either 4 or 6 years, these are observed halfway through the first term. Hence, for our medium-run outcomes, individuals will have completed their technical training and started working in their jobs, which are spread across diverse locations. Attrition occurs when there is a separation for an infraction, a failure to meet physical standards, or other forms of poor performance. The fractions retained 2 and 3 years after entry are 91% and 86%, respectively. Criminal incidents are rare, with only 3% and 4% having any incidents within 2 and 3 years.

Long-run outcomes are based on the rate at which individuals accumulate job skills by 4-5 years after entry. To be eligible for promotion to non-commissioned officer, individuals have to advance steadily through the lower ranks and complete sufficient coursework and on-the-job training. Promotion is competitive, and the most highly rated candidates within each job (including not-yet-promoted candidates from prior cohorts) are selected to fill available vacancies. Important contributors to the rankings are scores from an exam that tests job-specific knowledge, an exam that assesses general knowledge about the Air Force, and recent performance evaluations. For those who enter early enough for long-run outcomes to be observed, 58% are eligible for promotion and 18% are promoted on their first attempt.

To avoid sample selection issues, outcomes are defined to be unconditional on persisting in the Air Force. For example, if an individual drops out of technical training early on in their career, they are coded as being ineligible for promotion 4-5 years after entry. Likewise, for test scores, we define the outcome variables to be scoring above the average, coding non-test takers (e.g., those who drop out) as being below the average. The availability of outcomes across cohorts depends on whether they can feasibly have been observed for the required length of time.

To help with the interpretation of our results, we complement the outcomes observed while serving in the military with information about earnings for the same types of jobs in the civilian labor market. Specifically, we use mean earnings for the Standard Occupational Classification (SOC) codes which correspond to individuals' jobs. These capture the outside labor market opportunities individuals have once they leave the military.[11]

3 Empirical Strategy

Our goal is to estimate how ability mismatch in a job affects outcomes over different horizons. In this section, we first explain how we measure the degree to which an individual is over- or underqualified for their job in terms of cognitive ability. We then lay out our model for how outcomes depend on ability mismatch. Finally, we describe our instrumental variables strategy to deal with potential endogeneity.

3.1 Defining Ability Surplus and Deficit

Our measure of an individual's cognitive ability is their score on the AFQT. To define relative ability, we compare the individual's ability level to the average of others in the same job. This builds on the empirical approach taken by many in the literature (as surveyed by McGuinness et al. 2018), where the average numeracy and literacy (or sometimes education levels) observed in realized matches are used to characterize the general ability level of

[11] We link individuals' jobs to year 2000 SOC codes starting with the Defense Manpower Data Center mapping of military jobs to 2018 SOC codes (DMDC 2024), and then using BLS crosswalks to convert to 2000 SOC codes (BLS 2010, 2018). We then link data on mean earnings in 2006 (near the middle of our study period) from the BLS's Occupational Employment and Wage Statistics (BLS 2006).

different jobs. By defining relative ability in this way, we are comparing individuals to those undergoing the same training and working in the same job.

Specifically, relative ability for individual i assigned to job j in cohort c is defined as:

$$relative\,ability_{ijc} = AFQT_i - \overline{AFQT}_{j,-c} \tag{1}$$

The first term on the right hand side is the individual's own cognitive ability percentile, while the second is the average percentile score across all others in the same job across all cohorts other than cohort c, where a cohort is defined as week-by-career field. We calculate an individual's job AFQT using this leave-out mean to ensure the way jobs are characterized is unrelated to the cohort's actual assignments.

The distributions of both components of relative ability are shown in Figure 1. The left-hand graph is a histogram of individual AFQT percentiles. Since the AFQT is nationally normed, the distribution would be uniform if individuals in the Air Force were representative of the entire population. However, the left tail of the distribution is truncated because of the minimum entry requirement of 31, combined with the fact that in practice most jobs require a score above 40. There is declining representation at higher levels of cognitive ability. The right-hand graph plots the distribution of individuals' job AFQTs. There is a wide range of job AFQTs, with most of the density between the 50th and 80th AFQT percentiles. This distribution has spikes, in part because some of the 134 jobs are more common than others.

In the left panel of Figure 2, we plot the distribution of relative ability. The spread in own AFQT relative to job AFQT is substantial; the 10-90 range in this difference is -16 to +18 percentiles. This variation comes from two sources: differences in individual ability holding the job constant and differences in job ability level holding individual ability constant. Our identification strategy is designed to exploit the second source of variation. The right panel isolates this component by plotting the residual variation in relative ability after flexibly controlling for an individual's own AFQT. A substantial amount of variation remains, with the 10-90 range being -9 to +10 percentiles.

Figure 3 helps to explain how individuals of the same ability level can end up in jobs with different cognitive demands. It plots the mean AFQT scores of individuals and available jobs by weekly cohort.[12] Those graduating basic training at different points in time face a stronger or weaker job market based on (i) the level of competition from others in their cohort and (ii) the cognitive demands of the set of available jobs. For example, compared to someone with the same AFQT, an individual competing with more able individuals or for less demanding job options is more likely to be placed in a less demanding job and hence be overqualified.

To characterize the degree of over- and underqualification, we define ability surplus and deficit for individual i in job j and cohort c as:

$$\begin{aligned} ability\ surplus_{ijc} &= \begin{cases} relative\ ability_{ijc} & relative\ ability_{ijc} \geq 0 \\ 0 & \text{otherwise} \end{cases} \\ ability\ deficit_{ijc} &= \begin{cases} |relative\ ability_{ijc}| & relative\ ability_{ijc} < 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{2}$$

Summary statistics related to our ability measures are shown in the top six rows of Table 1. For those with an ability surplus (47% of individuals), the average surplus is 11 percentile points. Likewise, the average conditional ability deficit is 10 percentile points.

3.2 Regression Model

We model how individual outcomes depend on ability surplus and deficit as:

[12] To see how these two sources of variation combine to create across-cohort variation in relative ability, see Online Appendix Figure A3.

$$Y_{ijc} = \beta_1 ability\ surplus_{ijc} + \beta_2 ability\ deficit_{ijc} + f(AFQT_i) + X_i\Gamma + \alpha_C + e_{ijc} \quad (3)$$

where i, j, and c index individuals, jobs, and entry cohorts (defined as week-by-career field). The function $f(\cdot)$ is a flexible control for own cognitive ability; we use a third-order polynomial in our baseline specification. The regression also includes predetermined individual characteristics (X_i) observed at the time of entry, including gender, race/ethnicity, age, education, family status, health, enlistment waivers and options, and geographic Census division of the military entrance processing station (see Table 1). It also includes year-by-career field fixed effects (α_C) to capture differences in work environments over time, such as changing deployment patterns.

By controlling for own ability, the regression model is designed to isolate variation in ability surplus and deficit arising from job assignments. Conceptually, we are comparing outcomes for similarly-able individuals who end up in more and less cognitively demanding jobs. Our functional form for mismatch allows for differential effects of being over- or underqualified, but imposes that the reference point is the average cognitive ability in a job and that effects are linear. We explore robustness to including quintiles of relative ability instead.

This approach contrasts with other studies in the literature which restrict ability surplus and deficit to enter symmetrically, by taking the absolute value of relative ability (e.g., Fredriksson et al. 2018; Guvenen et al. 2020). An advantage of imposing symmetry is that controls for job ability levels (e.g., job fixed effects) can also be included in the regression. Under symmetry, mismatch is defined as being different from the average, but without a distinction for whether a worker is above or below average. The tradeoff that comes with our choice to not impose symmetry is that the ability level of a job is constrained to operate only through ability surplus and deficit.

A shared problem for observational studies of mismatch is that individuals non-randomly sort into jobs. Other work attempts to minimize selection bias by controlling for as many covariates as possible or by focusing on the beginning of workers' careers before realized productivity sorts individuals into different jobs over time. Even though our dataset contains many of the key variables used to assign individuals to early-career jobs, and even though there is less scope for job selection in the military, it is still possible that unobserved soft skills or individual preferences for jobs creates selection bias. For example, low-ability individuals who are assigned high-ability jobs may have demonstrated grit. For this reason, in the next section we outline our instrumental variables approach designed to isolate exogenous variation in ability surplus and deficit.

3.3 Instrumental Variables Approach

Our instruments for ability surplus and deficit are derived from simulated job assignments. The idea behind the instruments is that an individual's job assignment depends, in part, on factors outside of their control: the available slots in upcoming training programs and the quality of other recruits entering at the same time. These factors create quasi-random variation in job assignment and hence how cognitively demanding the individual's job is.

Simulated job assignments are based on individuals' rankings among others competing for jobs within the same week-by-career field entry cohort (i.e., the relevant job market). The logic is that, ceteris paribus, more able individuals are more likely to be assigned to more cognitively-demanding jobs. Specifically, we sort individuals by AFQT score. We then slot individuals into jobs with the corresponding rank, assigning the highest ability individual to the highest ability job, the second-highest ability individual to the second-highest ability job, and so on. We can readily construct simulated relative ability based on the job assigned by this rule, by replacing the actual job AFQT with the simulated job AFQT in equation (1).

Table 2 provides an example of how our simulation assigns individuals to jobs. The example cohort consists of 78 individuals and 78 jobs in the electronics career field. The example was chosen so that the correlation between actual and simulated job aptitude levels is similar to the full sample.[13] In the table, individuals are ranked in descending order by their AFQT scores (column 2), and in the case of ties, further by their actual job AFQTs (column 3a). It does not matter how we sort in the case of ties, since our process ultimately treats them identically, as explained below.

There is a clear correlation between an individual's own ability and that of their assigned job. However, there are exceptions, such as the high-ability individual (AFQT = 0.95) who is in a relatively low-aptitude job (job AFQT = 0.59) and the low-ability individual (AFQT = 0.48) who is in a middle-aptitude job (AFQT = 0.73). These types of matches could reflect unobserved preferences or traits, which is the type of selection our IV approach is designed to address.

To simulate job assignments, the full set of available jobs are ranked by job AFQT, with the titles shown in column 5b. Individuals are matched to the same-ranked available job and their simulated job AFQT (column 3b) is set to that job's AFQT. For individuals who share the same AFQT, their simulated job AFQT is set to the average job AFQT of the group of jobs they are matched to. This averaging is what leads them to be treated identically.[14] We then calculate simulated relative ability (column 4b) by taking each individual's own AFQT minus simulated job AFQT.

Figure 4 illustrates how the distributions of actual and simulated relative ability vary across AFQT bins. Comparing the median across bins makes clear that lower aptitude individuals tend to be underqualified, while higher aptitude individuals tend to be overqualified. Nonetheless, the 10-90 percentile ranges reveal substantial overlap across AFQT bins in the distribution of relative ability. While there is more variation in actual relative ability both across and within aptitude bins, there is also sizable variation for simulated relative ability.

[13]The correlation for the displayed rows is 0.70 versus 0.69 for the full sample.

[14]For example, three individuals have an AFQT of 0.95. Our simulation assigns two to Nuclear Forensics, which has the highest job AFQT (0.92), and the third to Weather Technician, which has the second highest job AFQT (0.82). They are all assigned the same simulated job AFQT of 0.89, which is the average across the three jobs.

To construct our instruments, we calculate simulated ability surplus and deficit using simulated relative ability and the same functional form as in equation (2). Since there are two potentially endogenous variables, there are two first stages. Each endogenous variable is regressed on the two simulated instrumental variables plus the other controls appearing in equation (3). As Table 3 reveals, the instruments are highly relevant and effective F-statistics indicate that there is not a weak instruments problem (Montiel Olea and Pflueger 2013).[15] As expected, ability surplus is best predicted by simulated ability surplus and ability deficit by simulated ability deficit. Contrary to what might be expected, the signs of the coefficients on the off-diagonals are positive. This is an indication that we do not have the correct functional form for the first stages, which are predicting variables that have kink points at zero with a parsimonious model. However, consistency does not require the functional form of the first stages to be correctly specified, but rather that the instruments are uncorrelated with the error term in the second stage.

4 Results

In this section, we report the OLS and IV estimates for how ability surplus and deficit affect outcomes across the three time horizons. To interpret magnitudes, we discuss the effects of 10 percentage point (pp) increases in the two measures (i.e., increases of 0.10). These correspond to roughly one standard deviation for both ability surplus and deficit.

4.1 Short Run

We start by reporting results for short-run outcomes. These outcomes occur during technical training for the assigned job, which takes place following graduation from basic training. The typical training takes 3-4 months to complete, with more involved training taking up to 6 months.

[15] Andrews et al. (2019) suggest computing the effective F-statistic and using the usual Stock-Yogo critical value of about 10.

Our first outcome is whether an individual graduates from technical training. Column 1 of Table 4 reports OLS estimates based on equation (3). A 10 pp increase in ability surplus is linked to a 0.8 pp increase in graduation, while a similar increase in ability deficit reduces graduation by 3.4 pp. For context, the mean graduation rate is 94% and, as a reminder, 10 pp is roughly one-standard deviation. Column 2 reports the IV estimate for this same outcome, but reaches a very different conclusion. A 10 pp increase in ability surplus causes a 1.5 pp *decrease* in graduation, and a similar increase in ability deficit causes a small and statistically insignificant *increase* in graduation.

The stark difference between OLS and IV suggests that even in this setting with centralized job assignments at the beginning of individual's careers, there is a large amount of selection. Selection is likely to be even greater in settings where individuals have more latitude to choose which job to pursue, or are observed later in their career after repeated sorting. One explanation for the contrast between columns 1 and 2 is that high-ability individuals who endogenously choose to be in less-challenging jobs do so because they prefer them, and are thus more likely to persist to graduation. Those that are sorted for exogenous reasons into less-challenging jobs are demotivated to graduate.

In the row labeled "p-endogeneity" we test whether the IV estimates differ from their OLS counterparts using a Hausman test. The test strongly rejects equality. This likely indicates bias in the OLS estimates, although we recognize it is also possible that differences arise due to heterogeneous treatment effects. We reject equality of the OLS and IV estimates not just for this outcome, but for all short-, medium-, and long-run outcomes. Hence, while we continue to report the OLS estimates for comparison purposes, we focus the discussion on the IV estimates for the remainder of the paper.

Columns 3-6 decompose the graduation outcome into two subcomponents: graduating on time and graduating with delay. Graduating with delay occurs when an individual needs to repeat technical training, but ends up eventually graduating. Fourteen percent of individuals graduate with delay. Focusing on the IV decomposition in columns 4 and 6, the drop in graduation due to ability surplus is partly due to a drop in on-time graduation and partly due to a drop in delayed graduation. For ability deficit, the decomposition reveals the null impact on graduation masks two opposing forces. Individuals with a one standard deviation larger ability deficit (i.e., a 10 pp increase) are 2.9 pp less likely to graduate on time, but this is more than offset by a 3.6 pp increase in graduating with delay. In other words, while these individuals initially struggle, they appear to be motivated to keep trying until they succeed.

Near the end of technical training, individuals take tests measuring their mastery of the skills needed for their assigned job. For individuals who do not graduate technical training, we do not observe test scores. To deal with this, we create a composite outcome that can be constructed for everyone: an indicator for whether an individual both graduates and scores above the job-specific average on the exams. The rationale for this composite outcome is that graduating and scoring above average are both good outcomes. The IV estimate for this outcome is reported in column 8. Relative to an overall mean of 43%, a 10 pp increase in ability surplus increases the likelihood of graduating and scoring above the mean by 2.8 pp, while a similar increase in ability deficit reduces it by 7.8 pp.[16] These results indicate that overqualification helps with mastery of skills by the end of technical training, while underqualification makes learning more difficult.

It would also be interesting to know the effects of ability surplus and deficit on the likelihood of scoring high on the test conditional on graduating. We can calculate these "intensive margin" responses if we assume that compliers have average likelihoods of graduating and of scoring high conditional on graduating (see Appendix B for details). While this is not a mild assumption, we view these estimates to be a useful complement to those for the composite

[16] In our data, 10% of graduates are missing technical training test scores for unknown reasons; in columns 7 and 8 we code the composite outcome as zero for these cases, which implicitly treats missingness as a negative outcome. If we alternately code the composite outcome as one, the estimates are larger in absolute value for ability surplus (0.517, s.e.=0.097) and deficit (-0.851, s.e.=0.146).

outcome. The rows labeled "im-surplus" and "im-deficit" in Table 4 report these intensive margin responses. The estimates are large in absolute value and statistically significant, implying that the effect on the joint outcome operates primarily through improved achievement. This reflects that surplus and deficit push graduation in the opposite direction of the composite outcome of graduating and scoring high.

Table 4 reports two additional tests related to functional form. As a reminder, we allow ability surplus and deficit to enter separately into the regression. One alternative is a linear model, where all that matters is relative ability as described in equation (1). We test this formally in the row labeled "p-linear", which reports the p-value for a test that the coefficients on ability surplus and deficit are equal but opposite in sign. For three out of four IV regressions, the null hypothesis of linearity is rejected. In the next row labeled "p-absolute" we instead test whether the effects of ability surplus and deficit have the same signs and magnitudes, i.e., that the absolute value of relative ability is what matters. This is the assumption several papers in the literature rely on for identification (e.g., Fredriksson et al. 2018; Guvenen et al. 2020). For the IV regressions, this alternative specification is always rejected.

4.2 Medium Run

We next turn to medium-run outcomes observed 2 and 3 years after entry, after an individual has had a chance to work in their job for some time. The sample size is somewhat smaller compared to the short run, since outcomes are only available for cohorts that entered early enough in our sample period to be tracked for the requisite number of years (see Online Appendix A).

Our first measure is whether an individual separates, as measured by 2- and 3-year attrition. These measures are cumulative, in that they include any effects on attrition during technical training. Job separation is most often due to poor performance or disciplinary problems. Table 5 reveals that ability surplus increases attrition both 2 and 3 years after entry, while ability deficit reduces it. Column 4 reports that a 10 pp increase (i.e., a one standard deviation increase) causes attrition to rise by 4.9 pp after 3 years. Compared to the effect on dropping out during technical training (Table 4, column 2), this impact is three times larger. In contrast, a 10 pp increase in ability deficit causes attrition to fall by 3.9 pp. These patterns are consistent with overqualified individuals valuing their jobs less and underqualified individuals being motivated to work hard to retain their jobs.

Behavioral violations at both the 2- and 3-year time horizons are also consistent with this story. Ability surplus leads to more behavioral problems, as measured by criminal incidents, while ability deficit leads to a reduction. Specifically, a 10 pp increase in ability surplus increases criminal involvement after three years by 19% relative to the mean ($0.10 \times 0.083 / 0.044$) while a similarly-sized increase in ability deficit causes a 46% reduction.

4.3 Long Run

Our long-run outcomes are measured 4-5 years out, near the end of a typical 4- or 6-year enlistment term. While earlier career advancements depend largely on tenure, promotion to become a non-commissioned officer is competitive. Promotion follows a tournament model, where only the most qualified candidates advance. Individuals who are unsuccessful can compete in future rounds; we focus on an individual's first opportunity. We measure long-term job performance using the key inputs to this promotion decision: whether an individual has moved far enough up the job ladder, a job-specific exam, a general knowledge exam, and performance evaluations by superiors.

Since advancing sufficiently to be eligible for promotion is an achievement in itself, this is our first long-run outcome. Column 2 of Table 6 shows that ability surplus reduces promotion eligibility, while ability deficit increases it. Comparing the magnitudes to Table 5 column 4, effects on 3-year attrition explain over two-thirds of the the effects on promotion-eligibility. The remainder is driven by either a failure to progress through the normal steps or additional exits after the 3-year mark.

Our second outcome is performance on the job-specific exam which measures mastery of job knowledge. The test is specific to an individual's job, so we normalize it to be mean zero with a standard deviation of one within jobs. To deal with selection into promotion eligibility, and hence the availability of this test score, we create a composite indicator of success which equals one if an individual both takes the test (i.e., is promotion eligible) and scores above average. We again find opposing effects of over- and underqualification, but in this case a surplus is beneficial and a deficit is harmful. As seen in Table 6 column 4, a 10 pp in ability surplus increases the composite outcome by 3.8 pp, which represents a 14% increase relative to the mean. In contrast, a similar increase in ability deficit results in a 20% decrease. Our estimates of the intensive margin responses in the last two rows of the table imply that the effects on the joint outcome operate primarily through improved achievement, which is not surprising since the effects on promotion eligibility run counter to those on the composite outcome.

Our third outcome is performance on the general test of Air Force knowledge, including history, organization, regulations, and practices. This is a common test taken by all individuals, and is hence normalized across all individuals regardless of their jobs. We again create a composite binary outcome for whether an individual both takes the test and scores above average. The estimates stand in stark contrast to the job knowledge tests, with opposite-signed results. For this common test, a 10 pp increase in ability surplus decreases the composite outcome by 5.6 pp, for a 20% drop relative to the mean. For ability deficit, there is a 13% increase relative to the mean. The statistically significant intensive margin estimates suggest that at least part of the composite effects are driven by effects on achievement.

Why is there an increase in job knowledge associated with an ability surplus, but a decrease in general knowledge? One possible explanation is that when overqualified individuals are placed into a job which is easy for them, they are not motivated to put in much effort, consistent with the effects we observed for the short- and medium-run. This would explain why they do worse on the general knowledge test (after conditioning on their own ability, as we do in the regressions), which is benchmarked relative to individuals across all jobs. These overqualified individuals may also not work hard to learn the specific knowledge associated with their job, but nonetheless find it relatively easy to learn the information compared to the relevant comparison group of others who are less able but in the same job.

Column 8 considers another input into the promotion decision. Individuals receive a performance report score which is based on supervisors' evaluations of their job performance and decorum on- and off-duty. Since most promotion-eligible individuals get perfect evaluations (73%) and attriting individuals are missing evaluations, we create a composite measure which equals one if an individual is promotion eligible and has a perfect performance score. A 10 pp increase in ability surplus reduces this composite outcome by 7.5 pp, with the intensive margin estimate indicating that not all of this reduction is due to being ineligible. Ability deficit does not have a statistically significant effect.

Our final long-run outcome, presented in column 10, is the promotion outcome itself. Only 18% of individuals in our sample achieve promotion, and the success rate for eligibles is just 31%. A 10 pp rise in ability surplus increases the chances of being both eligible and promoted by 4.5 pp, which translates to a 25% increase relative to the mean. A similar increase in ability deficit reduces promotion by 34%. Since ability surplus reduces eligibility but increases promotion, it follows that it also increases promotion among eligibles, as seen in the intensive margin estimate. Following the same logic but running in the opposite direction, ability deficit increases eligibility but reduces promotion, thus also decreasing promotion among eligibles.

Since promotion occurs within jobs, these results line up with the findings that ability surplus increases, while ability deficit decreases, job-specific knowledge. In other words, even though overqualification causes individuals to be more likely to attrit and do worse on a general subject test, they still stand out relative to others in the same job as being the most qualified for promotion. And even though underqualification causes individuals to persist and aquire more general military knowledge, they are at a disadvantage when it comes to competing in a tournament with more able individuals.

5 Robustness and Heterogeneity

We now explore robustness and heterogeneity. For these analyses, we use the IV model and focus on a subset of outcomes. We chose two outcomes from each of the three time horizons: graduation from technical training, graduating and scoring above average on technical training exams, 3-year retention, 3-year behavior, being eligible for promotion and scoring above average on the job-specific exam, and promotion.

5.1 Robustness

We start by testing sensitivity to finer controls for own ability. Our empirical design is meant to compare otherwise similar individuals assigned to jobs of different average ability levels. To control for an individual's ability level, our main specification includes a third-order polynomial in own AFQT. In Table 7 we instead include indicators for each AFQT percentile so the comparison is between individuals with identical ability levels but assigned to different jobs. The estimates are similar.

We next test whether the results change when measures related to cohort quality are added as controls. Recall that the identifying variation in our instruments for ability surplus and deficit comes from two sources: variation in the set of available jobs at entry and in the set of individuals competing for these jobs. This raises the possibility that some of the effects we observe could be attributed to peer effects during training. Hence, Table 8 adds two control variables: the mean AFQT of an individual's peers in the same entry cohort, and an individual's own AFQT percentile rank within their cohort. The former variable is designed to capture linear-in-means peer effects which have been commonly estimated and critiqued in the literature (e.g., Carrell et al. 2009; Imberman et al. 2012; Lavy and Schlosser 2011; Sacerdote 2014). The latter variable is designed to capture the effect of ordinal rank, which has been shown to matter in the military and education settings (e.g., Chesney and Carrell 2024; Denning et al. 2023; Elsner et al. 2021; Murphy and Weinhardt 2020). The estimates for ability surplus and deficit are largely unchanged, indicating that our results are unlikely to be driven by peer effects.

The results from several additional robustness checks appear in the Online Appendix. First, we use an alternative functional form for own ability relative to the job. Instead of including ability surplus and deficit, we include indicators for quintiles of relative ability. While the estimates are noisy, the signs line up with those implied by our main specification (22 out of 24 times) and the estimated effects are greatest in the tail quintiles (Appendix Table C1). Second, we change the definition of the job market used to construct the instruments to allow for potential competition from adjacent cohorts. This could arise due to delays between when individuals first commit and when they are called to basic training, at which point jobs and career fields are assigned. Rather than sorting individuals and jobs based on the week they enter basic training, if we broaden the window to ± 4 weeks of entry, the results are qualitatively unchanged (Appendix Table C2). Finally, the basic training records have less comprehensive coverage for early cohorts, while technical training and promotion records have less complete coverage for later cohorts (see Appendix A). Excluding early or later cohorts yields broadly similar conclusions (Appendix Tables C3 and C4).

5.2 Heterogeneity

We now explore whether the effects of being over- and underqualified are heterogeneous, starting with males versus females in Table 9. To do this, we interact our ability surplus and deficit measures with a female indicator variable. In our sample, females are a minority, comprising only one-fifth of the sample.[17] For men, being underqualified for a job has the

[17] Summary statistics broken down by gender (and by race/ethnicity) can be found in Table 1.

same effects seen for the entire sample: it improves retention and behavior, but reduces job knowledge and promotion. Relative to men, ability deficit negatively impacts outcomes for women. Men's graduation increases by 13.5 pp for a 10 pp increase in ability deficit, whereas women's graduation falls sharply by 34.1 pp. This same differential response shows up for 3-year attrition. While both men and women are negatively affected on the job-specific test score outcomes (columns 2 and 5), the negative effect on these achievement-related outcomes is more than double for women. Interestingly, the positive effects of ability surplus on skill acquisition are also magnified for women.

These patterns indicate that women are more sensitive to ability mismatch. A potential explanation from the literature is that women's decisions on whether to persist in a field of study are more responsive to relative performance signals, such as grades (e.g., Kugler et al. 2021; Ost 2010; Rask and Tiefenthaler 2008). There is also evidence that women tend to shy away from competition (Astorne-Figari and Speer 2019; Niederle and Vesterlund 2007). Women who are in jobs for which they are underqualified likely receive negative relative performance signals and face stiffer competition, while the reverse is likely true for those with an ability surplus.

In Table 10, we analyze heterogeneity by race and ethnicity. In our sample, 15% are Black (non-Hispanic) and 10% are Hispanic. Similar to the pattern observed for women, Blacks tend to be more negatively affected by ability deficit relative to the reference group (non-Hispanic, non-Black individuals). For example, an ability deficit harms Black skill acquisition during technical training by roughly twice as much. The effects for ability surplus on skill acquisition are more puzzling, as the effects in the short run and long run are opposite one another. For Hispanics, neither ability surplus nor deficit has statistically significant differential effects. These findings suggest that ability deficits exacerbate existing racial disparities for Blacks in skill acquisition and career advancement, but not for Hispanics.

6 Outside Labor Market Incentives

Summing up our findings, overqualification causes individuals to attrit at higher rates, both during technical training and after they are working in their assigned jobs. It also results in more behavioral problems, lower performance on the general knowledge tests taken by all workers, and worse performance evaluations by their superiors. On the other hand, overqualification results in better performance relative to others in the same job: job-specific test scores rise both during technical training and while on the job, and these individuals are more likely to be promoted. These patterns suggest that overqualified individuals are less motivated, but when judged relative to their others in their job still outperform them.

Underqualification results in the polar opposite: it decreases attrition, reduces behavioral problems, increases general knowledge, and improves performance evaluations, but also lowers performance on job-specific tests and harms promotion prospects. These patterns are consistent with underqualified individuals being motivated to put forth more effort, but struggling to compete with others in the same job.

While we cannot directly observe motivation, we do have a measure for the incentive to invest in one's assigned job related to future outside market options. Indeed, one of the recruiting pitches made by the military is that it provides valuable career training. During military service, earnings depend only on rank and step, so there is no variation across jobs. Nonetheless, different military jobs have substantially different civilian prospects. Most individuals do not spend their entire career in the military; during our time period, the average length of service is 9 years for enlisted personnel (Air Force Personnel Center News Service 2005). To capture the value of career incentives, we use mean earnings (from 2006, inflated to 2025 dollars) for the Standard Occupational Classification codes which correspond to individuals' assigned jobs (see footnote 11).

In Table 11 we estimate our baseline IV model with this measure of civilian earnings potential as the outcome. Column 1 reveals that when ability surplus increases by 10 pp, outside earnings fall by approximately $8,200. In contrast, a 10 pp increase in ability deficit causes outside market earnings to rise by $9,800. For context, average outside earnings are roughly $73,000. Column 2 presents results using log earnings as the outcome. A 10 pp increase in ability surplus is predicted to reduce outside earnings by 10%, while a similar increase in ability deficit is associated with a 13% increase.

In other words, after conditioning on ability level (as we do in our regressions), individuals who are overqualified are in jobs which have lower outside earnings, while the reverse is true for those who are underqualified. Given the pattern of findings summarized above, these differential future earnings incentives appear to demotivate overqualified individuals while motivating those who are underqualified. This is consistent with individuals being willing to make larger investments in careers which have higher returns (Becker 1964).

7 Conclusion

Despite the importance of ability mismatch for the labor market, the challenge of endogenous selection into jobs has made estimating these effects empirically difficult. By leveraging plausibly exogeneous variation in job assignments in the US Air Force, we provide the first quasi-experimental estimates of how skill mismatch affects early career success. Our findings suggest overqualified individuals are less motivated to invest and persist in their jobs, though they still outperform others in the same roles. On the other hand, underqualified individuals appear to put forth more effort, but struggle to acquire the job-specific skills necessary to compete successfully for promotion. Consistent with differential incentives, individuals who are overqualified are in jobs which are less valuable in terms of future outside earnings potential, while the reverse is true for those who are underqualified.

These findings provide insights into the tradeoffs firms face when making hiring decisions. An overqualified hire may be more productive relative to others in the same job, but comes with the risk of less commitment and higher turnover. For an underqualified hire, the tradeoffs are in the opposite direction. Ultimately, firms need to weigh benefits versus costs depending on which margins they value most. The same set of tradeoffs exist when allocating a fixed set of workers to jobs or tasks with a firm, but with the added wrinkle that reallocations which increase one worker's relative ability must lower someone else's.

While our study provides novel evidence on the consequences of ability mismatch for labor market outcomes, we are only able to scratch the surface on how firms should weigh the tradeoffs we document. Although we do not have monetary values for other outcomes, there are figures for the US Air Force indicating that re-training an individual costs \$15,671 and attrition costs \$41,135 (Manacapilli et al. 2012, inflated to 2025 dollars). Combined with our estimates, placing an individual in a job where they have a one-standard deviation higher ability deficit is predicted to impose \$569 worth of retraining costs (via graduation delays) while saving \$1,612 from reduced attrition (within 3 years). In contrast, an individual who is overqualified by the same amount is not predicted to affect training costs, but imposes costs of \$1,995 through increased attrition. Effort and productivity effects would also need to be taken into account for a complete accounting. Future work could dig deeper into whether firms optimally trade off the costs and benefits of hiring workers who are over- or underqualified.

References

Air Force Personnel Center News Service (2005). Service demographics offer snapshot of force. https://www.af.mil/News/Article-Display/Article/134672/service-demographics-offer-snapshot-of-force/. Published April 7, 2005. Accessed 4 June 2025.

Andrews, I., Stock, J. H., and Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, 11(1):727–753.

Andrews, R. J., Imberman, S. A., and Lovenheim, M. F. (2017). Risky business? The effect of majoring in business on earnings and educational attainment. NBER Working Paper 23575.

Anelli, M. (2020). The returns to elite university education: A quasi-experimental analysis. *Journal of the European Economic Association*, 18(6):2824–2868.

Arcidiacono, P., Aucejo, E. M., and Hotz, V. J. (2016). University differences in the graduation of minorities in STEM fields: Evidence from California. *American Economic Review*, 106(3):525–62.

Arcidiacono, P. and Lovenheim, M. (2016). Affirmative action and the quality-fit trade-off. *Journal of Economic Literature*, 54(1):3–51.

Astorne-Figari, C. and Speer, J. D. (2019). Are changes of major major changes? The roles of grades, gender, and preferences in college major switching. *Economics of Education Review*, 70:75–93.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Columbia University Press (for NBER), New York.

Black, S. E., Denning, J. T., and Rothstein, J. (2023). Winners and losers? The effect of gaining and losing access to selective colleges on education and labor market outcomes. *American Economic Journal: Applied Economics*, 15(1):26–67.

Bleemer, Z. (2022). Affirmative action, mismatch, and economic mobility after California's Proposition 209. *The Quarterly Journal of Economics*, 137(1):115–160.

Bleemer, Z. and Mehta, A. (2022). Will studying economics make you rich? A regression discontinuity analysis of the returns to college major. *American Economic Journal: Applied Economics*, 14(2):1–22.

BLS (2006). Occupational employment and wage statistics (OEWS) tables, 2006. https://www.bls.gov/oes/tables.htm. Accessed 10 May 2025.

BLS (2010). Crosswalks between the 2010 (SOC) and systems used by other federal and international statistical agencies. https://www.bls.gov/soc/soccrosswalks.htm. Accessed 10 May 2025.

BLS (2018). Crosswalks between the 2018 (SOC) and systems used by other federal and international statistical agencies. https://www.bls.gov/soc/2018/crosswalks_used_by_agencies.htm. Accessed 10 May 2025.

Campbell, S., Macmillan, L., Murphy, R., and Wyness, G. (2022). Matching in the dark? Inequalities in student to degree match. *Journal of Labor Economics*, 40(4):807–850.

Carrell, S. E., Fullerton, R. L., and West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3):439–464.

Chesney, A. J. and Carrell, S. E. (2024). Brains versus brawn: Ordinal rank effects in job training. *Journal of Public Economics*, 238:105197.

Deming, D. and Kahn, L. B. (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, 36(S1):S337–S369.

Denning, J. T., Murphy, R., and Weinhardt, F. (2023). Class rank and long-run outcomes. *Review of Economics and Statistics*, 105(6):1426–1441.

Department of Defense (2019). Population representation in the military services: Fiscal year 2019 summary report.

Dillon, E. W. and Smith, J. A. (2020). The consequences of academic match between students and colleges. *Journal of Human Resources*, 55(3):767–808.

DMDC (2024). Crosswalk files. https://www.onetcenter.org/crosswalks.html. O*NET Resource Center. Accessed 25 April 2025.

Duncan, G. J. and Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review*, 1(1):75–86.

Eide, E. R., Hilmer, M. J., and Showalter, M. H. (2016). Is it where you go or what you study? The relative influence of college selectivity and college major on earnings. *Contemporary Economic Policy*, 34(1):37–46.

Elsner, B., Isphording, I. E., and Zölitz, U. (2021). Achievement rank affects performance and major choices in college. *The Economic Journal*, 131(640):3182–3206.

Fredriksson, P., Hensvik, L., and Skans, O. N. (2018). Mismatch of talent: Evidence on match quality, entry wages, and job mobility. *American Economic Review*, 108(11):3303–38.

Guvenen, F., Kuruscu, B., Tanaka, S., and Wiczer, D. (2020). Multidimensional skill mismatch. *American Economic Journal: Macroeconomics*, 12(1):210–44.

Hastings, J. S., Neilson, C. A., and Zimmerman, S. D. (2013). Are some degrees worth more than others? Evidence from college admission cutoffs in Chile. NBER Working Paper 19241.

Hersch, J. (1991). Education match and job match. *The Review of Economics and Statistics*, 73(1):140–144.

Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics*, 91(4):717–724.

Imberman, S. A., Kugler, A. D., and Sacerdote, B. I. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, 102(5):2048–2082.

Kahn, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, 17(2):303–316.

Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111.

Kugler, A. D., Tinsley, C. H., and Ukhaneva, O. (2021). Choice of majors: Are women really different from men? *Economics of Education Review*, 81:102079.

Lavy, V. and Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3(2):1–33.

Levy, F. and Murnane, R. J. (1992). US earnings levels and earnings inequality: A review of recent trends and proposed explanations. *Journal of Economic Literature*, 30(3):1333–1381.

Lindenlaub, I. (2017). Sorting multidimensional types: Theory and application. *The Review of Economic Studies*, 84(2):718–789.

Lise, J. and Postel-Vinay, F. (2020). Multidimensional skills, sorting, and human capital accumulation. *American Economic Review*, 110(8):2328–76.

Liu, K., Salvanes, K. G., and Sørensen, E. (2016). Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review*, 84:3–17.

Lytell, M. C., Kuhn, K., Haddad, A., Marquis, J. P., Lim, N., Hall, K. C., Stewart, R., and Wenger, J. W. (2015). *Navy and Air Force Active-Duty Drawdowns in the Mid-2000s*, pages 39–52. RAND Corporation.

Manacapilli, T., Matthies, C. F., Miller, L. W., Howe, P., Perez, P. J., Hardison, C. M., Massey, H. G., Greenberg, J., Beighley, C., and Sims, C. S. (2012). Reducing attrition in selected Air Force training pipelines. *RAND Project Air Force Technical Report*.

McGowan, M. A. and Andrews, D. (2017). Labor market mismatch and labor productivity: Evidence from PIAAC data. In *Skill Mismatch in Labor Markets*. Emerald Publishing Limited.

McGuinness, S., Pouliakas, K., and Redmond, P. (2018). Skills mismatch: Concepts, measurement and policy approaches. *Journal of Economic Surveys*, 32(4):985–1015.

Montiel Olea, J. L. and Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3):358–369.

Mountjoy, J. (2024). Marginal returns to public universities. NBER Working Paper 32296.

Mountjoy, J. and Hickman, B. R. (2021). The returns to college(s): Relative value-added and match effects in higher education. NBER Working Paper 29276.

Murphy, R. and Weinhardt, F. (2020). Top of the class: The importance of ordinal rank. *The Review of Economic Studies*, 87(6):2777–2826.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Oreopoulos, P., von Wachter, T., and Heisz, A. (2012). The short-and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, 4(1):1–29.

Ost, B. (2010). The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review*, 29(6):923–934.

Perry, A., Wiederhold, S., and Ackermann-Piek, D. (2014). How can skill mismatch be measured? New approaches with PIAAC. *Methods, Data, Analyses*, 8(2):38.

Rask, K. and Tiefenthaler, J. (2008). The role of grade sensitivity in explaining the gender imbalance in undergraduate economics. *Economics of Education Review*, 27(6):676–687.

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics*, 6(1):253–272.

Sattinger, M. (1993). Assignment models of the distribution of earnings. *Journal of Economic Literature*, 31(2):831–880.

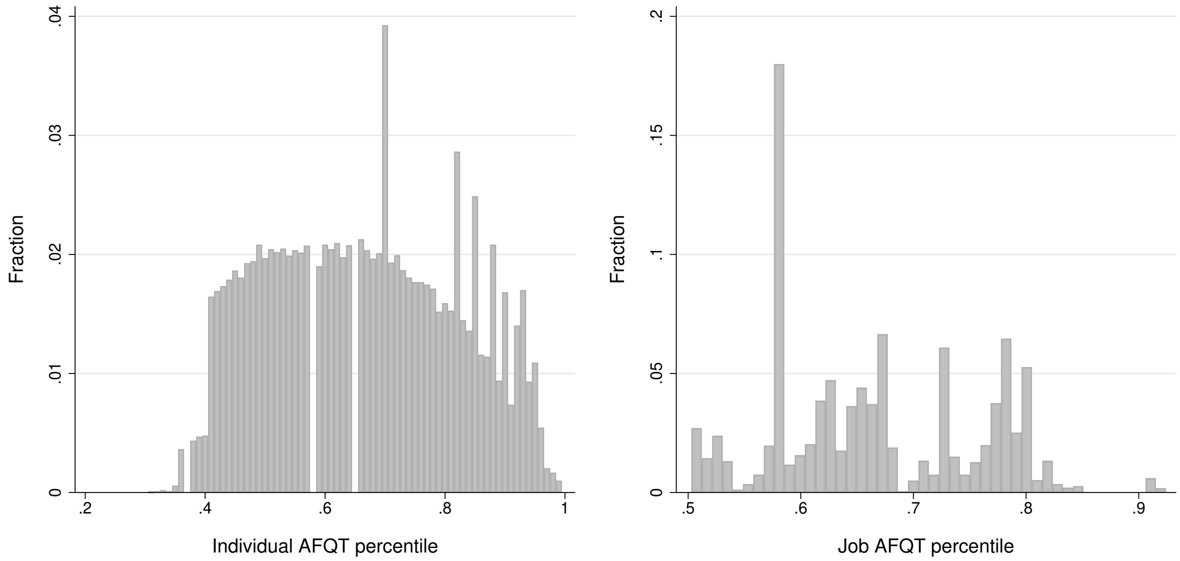
Schwandt, H. and von Wachter, T. (2019). Unlucky cohorts: Estimating the long-term effects of entering the labor market in a recession in large cross-sectional data sets. *Journal of Labor Economics*, 37(S1):S161–S198.

Segal, D. (2004). Development and evaluation of the 1997 ASVAB score scale. Defense Manpower Data Center, Technical Report 2004-002.

Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32(4):711–754.

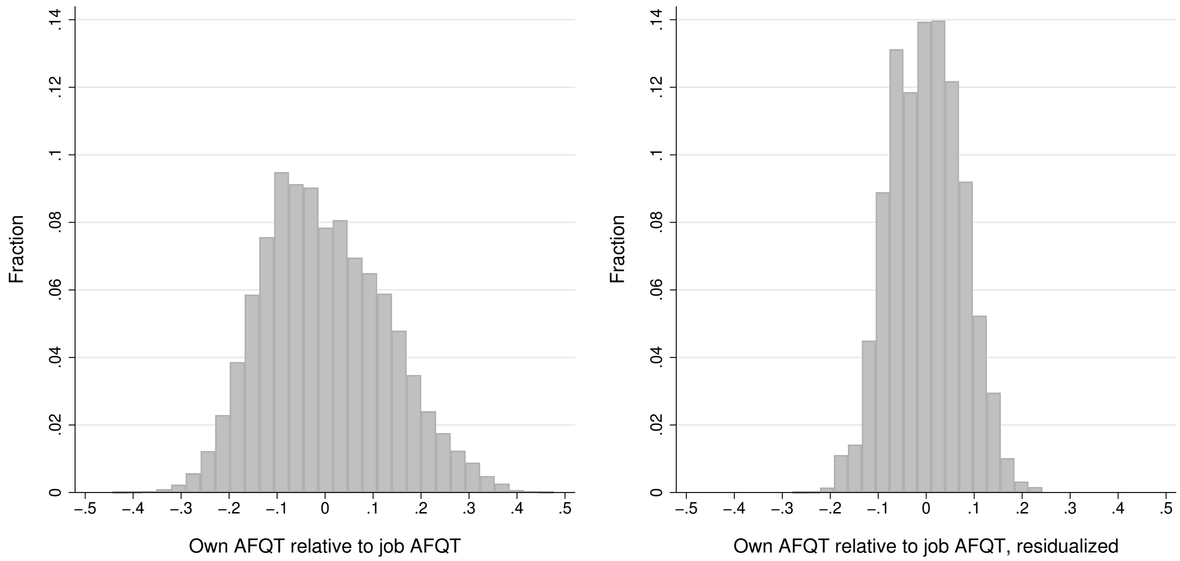
Zimmerman, S. D. (2019). Elite colleges and upward mobility to top jobs and top incomes. *American Economic Review*, 109(1):1–47.

Figure 1: Distribution of ability across individuals and jobs



Notes: The left panel shows the distribution of individual AFQT percentile scores. Some percentile scores are unpopulated, while others have spikes, due to the integer rounding done by the Department of Defense in the mapping of raw AFQT scores to percentiles (see Segal (2004), pp. 13-15). The right panel shows the distribution of individual's job AFQT scores, defined as the average AFQT score of individuals in the same job leaving out one's own cohort. The spike at 0.58 is driven by the job "Security Forces", which is the most common job. There are 90,543 individuals and 134 different jobs.

Figure 2: Distribution of relative ability



Notes: The left hand panel plots the distribution of relative ability, where relative ability is defined as an individual's own AFQT percentile minus their job AFQT percentile. The right hand panel plots the distribution of residualized relative ability after flexibly controlling for an individual's own AFQT using a third-order polynomial.

Figure 3: Variation in average ability of individuals and available jobs by cohort

Notes: Open circles plot the average AFQT percentile score of individuals within a weekly basic training graduation cohort, while grey circles plot the average job AFQT for the set of jobs available to the cohort. For ease of viewing, the graph presents means by week, whereas a job market is defined by career field by week.

Figure 4: Distributions of actual and simulated relative ability, by AFQT decile

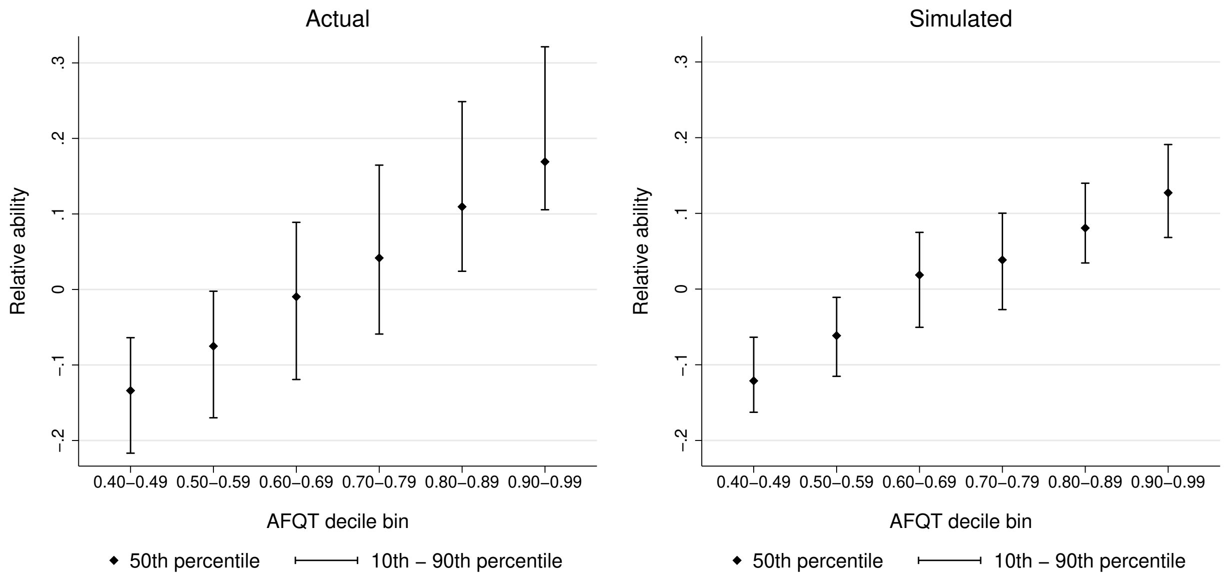


Table 1: Summary statistics

	All (1)	Male (2)	Female (3)	Black (4)	Hispanic (5)	White or other (6)
Own AFQT percentile	0.67	0.68	0.62	0.59	0.62	0.69
Job AFQT percentile	0.67	0.67	0.64	0.63	0.65	0.67
Relative ability	-0.00	0.00	-0.01	-0.04	-0.03	0.01
Ability surplus (conditional on surplus$\geq$0)	0.11	0.12	0.10	0.09	0.10	0.12
Ability deficit (conditional on deficit>0)	0.10	0.10	0.08	0.10	0.10	0.10
Fraction with ability surplus	0.47	0.49	0.40	0.31	0.37	0.51
Gender: female	0.21	0.00	1.00	0.29	0.25	0.19
Race/ethnicity: Black (non-Hispanic)	0.15	0.14	0.22	1.00	0.00	0.00
Race/ethnicity: Hispanic	0.10	0.09	0.11	0.00	1.00	0.00
Race/ethnicity: other non-White	0.06	0.06	0.07	0.00	0.00	0.09
Age: 19-20 years	0.39	0.39	0.36	0.39	0.39	0.39
Age: 21-22 years	0.15	0.16	0.14	0.14	0.15	0.16
Age: 23+ years	0.10	0.11	0.09	0.08	0.11	0.11
Education: has some college	0.08	0.07	0.08	0.06	0.08	0.08
Family status: married	0.08	0.08	0.09	0.05	0.11	0.09
Family status: has 1+ dependents	0.07	0.07	0.07	0.05	0.09	0.07
Health: no physical limitations	0.78	0.79	0.77	0.82	0.79	0.78
Health: medical failure on initial exam	0.09	0.09	0.10	0.08	0.08	0.09
Health: BMI overweight or obese	0.34	0.38	0.18	0.30	0.39	0.34
Enlistment option: has conduct waiver	0.07	0.08	0.04	0.05	0.07	0.08
Enlistment option: has medical waiver	0.05	0.05	0.05	0.04	0.04	0.05
Enlistment option: term > 4 years	0.51	0.56	0.34	0.44	0.45	0.53
Enlistment option: advanced pay grade	0.23	0.23	0.25	0.24	0.25	0.23
Census division: New England	0.03	0.03	0.03	0.01	0.02	0.04
Census division: Middle Atlantic	0.10	0.10	0.10	0.10	0.08	0.11
Census division: East North Central	0.12	0.12	0.11	0.09	0.04	0.14
Census division: West North Central	0.08	0.08	0.08	0.03	0.02	0.10
Census division: South Atlantic	0.20	0.20	0.21	0.40	0.19	0.16
Census division: East South Central	0.08	0.08	0.07	0.11	0.01	0.08
Census division: West South Central	0.16	0.16	0.16	0.16	0.30	0.14
Census division: Mountain	0.07	0.07	0.07	0.03	0.08	0.08
Census division: Pacific	0.15	0.15	0.17	0.08	0.25	0.16
Career field: mechanical	0.30	0.35	0.13	0.18	0.26	0.33
Career field: administrative	0.07	0.04	0.19	0.16	0.11	0.05
Career field: general	0.43	0.39	0.59	0.49	0.46	0.42
Career field: electronics	0.20	0.22	0.10	0.17	0.17	0.21
N	90,543	71,753	18,790	14,015	8,704	67,824

Notes: Each column shows means for different samples. See Section 3.1 for details on the construction of the variables in the first 6 rows. The definitions for the other variables are detailed in Appendix Table A4.

Table 2: Example of sorting algorithm used to construct the simulated instruments

Individual		Job AFQT		Relative ability		Job title	
Rank	AFQT	Actual	Sim	Actual	Sim	Actual	Simulated
(1)	(2)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
1	0.95	0.92	0.89	0.03	0.06	Nuclear Forensics	Nuclear Forensics
2	0.95	0.78	0.89	0.17	0.06	Network Integration	Nuclear Forensics
3	0.95	0.59	0.89	0.36	0.06	Electrical	Weather Technician
4	0.94	0.78	0.80	0.16	0.14	Integrated Avionics (Instr./Controls)	Precision Equipment
5	0.94	0.78	0.80	0.16	0.14	Network Integration	Precision Equipment
6	0.94	0.78	0.80	0.16	0.14	Avionics (F16)	Cryptography
7	0.94	0.78	0.80	0.16	0.14	Integrated Avionics (Elec. Warfare)	Cryptography
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
36	0.80	0.78	0.78	0.02	0.02	Integrated Avionics (Elec. Warfare)	Integrated Avionics (Comm./Nav.)
37	0.80	0.73	0.78	0.07	0.02	Aircraft Elec. / Environ.	Integrated Avionics (Comm./Nav.)
38	0.79	0.80	0.78	-0.01	0.01	Cryptography	Integrated Avionics (Comm./Nav.)
39	0.78	0.80	0.78	-0.02	0.00	Cryptography	Avionics (F16)
40	0.78	0.78	0.78	0.00	0.00	Integrated Avionics (Elec. Warfare)	Avionics (F16)
41	0.78	0.78	0.78	0.00	0.00	Integrated Avionics (Elec. Warfare)	Avionics (F16)
42	0.76	0.78	0.78	-0.02	-0.02	Network Integration	Integrated Avionics (Elec. Warfare)
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
72	0.49	0.63	0.62	-0.14	-0.13	Aircraft Armament (F16)	Aircraft Armament (F15)
73	0.49	0.63	0.62	-0.14	-0.13	Aircraft Armament (F16)	Aircraft Armament (F15)
74	0.48	0.73	0.62	-0.25	-0.14	Aircraft Elec. / Environ.	Aircraft Armament (B1)
75	0.48	0.59	0.62	-0.11	-0.14	Electrical	Aircraft Armament (B1)
76	0.47	0.62	0.59	-0.15	-0.12	Aircraft Armament (B1)	Electrical
77	0.47	0.51	0.59	-0.04	-0.12	HVAC/Refrigeration	Electrical
78	0.41	0.63	0.51	-0.22	-0.10	Aircraft Armament (F16)	HVAC/Refrigeration

Notes: The table illustrates how the instrument is constructed for an example cohort; see Section 3.3 for details of the algorithm. Column 1 lists an individual's rank based on their AFQT percentile, which is reported in column 2. The remaining columns show the actual and simulated job AFQT, relative ability, and job title.

Table 3: First stage

	Ability surplus	Ability deficit
	(1)	(2)
Simulated ability surplus	0.692**	0.121**
	(0.010)	(0.005)
Simulated ability deficit	0.180**	0.689**
	(0.009)	(0.010)
DV mean	0.053	0.053
N	90,543	90,543
Effective F-statistic	3,719	3,158

Notes: Control variables include indicators for the gender, race/ethnicity, age, education, family status, health, enlistment, and geographic categories listed in Table 1. All specifications also include cubic terms for the individual's AFQT percentile score and interactions between basic training graduation year and career field. Standard errors are reported in parentheses and are clustered by basic training graduation cohort (year $\times$ week). The effective F-statistic is calculated according to Montiel Olea and Pflueger (2013).
** 5%, * 10% significance level

Table 4: Short-run technical training outcomes (<6 months)

	Graduate		Graduate on time		Graduate with delay		> avg tech score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ability surplus	0.081**	-0.152**	0.329**	-0.051	-0.248**	-0.101	1.262**	0.279**
	(0.016)	(0.042)	(0.029)	(0.071)	(0.025)	(0.066)	(0.036)	(0.101)
Ability deficit	-0.337**	0.071	-0.688**	-0.292**	0.351**	0.363**	-1.256**	-0.784**
	(0.031)	(0.066)	(0.047)	(0.123)	(0.035)	(0.101)	(0.044)	(0.141)
IV	No	Yes	No	Yes	No	Yes	No	Yes
DV mean	0.943	0.943	0.798	0.798	0.144	0.144	0.426	0.426
N	90,543	90,543	90,543	90,543	90,543	90,543	90,543	90,543
p-linear	0.000	0.160	0.000	0.003	0.013	0.006	0.901	0.000
p-absolute	0.000	0.017	0.000	0.142	0.000	0.001	0.000	0.000
p-endogeneity		0.000		0.000		0.016		0.000
im-surplus							1.271**	0.562**
im-deficit							-0.858**	-1.008**

Notes: Each pair of columns shows the results from OLS and IV specifications, for the technical training outcome indicated in the column heading. See the notes to Table 3 for the list of included control variables. The fourth outcome (>avg tech score) is a composite variable for whether an individual both graduates technical training and scores above average on the job-specific exams. We report p-values for a series of tests: "p-linear" is for whether the coefficients on ability surplus and ability deficit are equal and opposite in sign, "p-absolute" is for whether they are equal, and "p-endogeneity" is for whether the IV estimates for ability surplus and deficit equal their OLS counterparts. The rows labeled "im-surplus" and "im-deficit" present the intensive margin effects for scoring above the mean conditional on graduating technical training, calculated as described in Appendix B. Standard errors are reported in parentheses and are clustered by basic training graduation cohort (year × week).
** 5%, * 10% significance level

Table 5: Medium-run outcomes (2-3 years)

	2-year attrition		3-year attrition		2-year behavior		3-year behavior	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ability surplus	0.148**	0.371**	0.165**	0.485**	0.023*	0.081**	0.015	0.083**
	(0.021)	(0.052)	(0.028)	(0.066)	(0.012)	(0.030)	(0.016)	(0.036)
Ability deficit	-0.160**	-0.346**	-0.222**	-0.392**	-0.091**	-0.193**	-0.094**	-0.201**
	(0.032)	(0.082)	(0.042)	(0.125)	(0.017)	(0.051)	(0.026)	(0.079)
IV	No	Yes	No	Yes	No	Yes	No	Yes
DV mean	0.093	0.093	0.137	0.137	0.030	0.030	0.044	0.044
N	75,498	75,498	62,449	62,449	66,690	66,690	44,934	44,934
p-linear	0.766	0.761	0.250	0.471	0.003	0.046	0.020	0.137
p-absolute	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
p-endogeneity		0.000		0.000		0.029		0.086

Notes: Each pair of columns shows the results from OLS and IV specifications, for the medium-run outcome indicated in the column heading. The attrition variables are indicators for separating due to poor performance within 24 or 36 months of entry. The misbehavior variables are indicators for any criminal incidents within 24 or 36 months of entry. The sample is limited to those who entered early enough for the outcomes to be observed (see Online Appendix A). For other details, see the notes to Table 4.
** 5%, * 10% significance level

Table 6: Long-run promotion outcomes (4-5 years)

	Promotion eligible		> avg job knowledge score		> avg USAF knowledge score		Perfect performance evals		Promoted	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ability surplus	-0.352**	-0.684**	0.484**	0.375**	-0.307**	-0.559**	-0.362**	-0.745**	0.519**	0.448**
	(0.062)	(0.116)	(0.058)	(0.113)	(0.056)	(0.133)	(0.062)	(0.127)	(0.051)	(0.108)
Ability deficit	0.424**	0.503**	-0.196**	-0.553**	0.395**	0.379*	0.256**	0.329	-0.257**	-0.612**
	(0.058)	(0.242)	(0.057)	(0.225)	(0.054)	(0.225)	(0.062)	(0.222)	(0.045)	(0.183)
IV	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
DV mean	0.580	0.580	0.277	0.277	0.281	0.281	0.424	0.424	0.180	0.180
N	34,148	34,148	34,148	34,148	34,148	34,148	34,148	34,148	34,148	34,148
p-linear	0.392	0.442	0.001	0.438	0.273	0.430	0.219	0.089	0.001	0.429
p-absolute	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.000
p-endogeneity		0.006		0.086		0.067		0.005		0.067
im-surplus			1.127**	1.212**	-0.236**	-0.392**	-0.181**	-0.422**	1.085**	1.140**
im-deficit			-0.687**	-1.371**	0.328**	0.234	-0.093	-0.069	-0.670**	-1.327**

Notes: Each pair of columns shows the results from OLS and IV specifications, for the long-run outcome indicated in the column heading. The first outcome (promotion eligible) is an indicator for being eligible for promotion to non-commissioned officer. The remaining outcomes are composite variables that are set to zero for those who are not promotion eligible and thus do not have records. The sample is limited to those who entered early enough for the outcomes to be observed (See Online Appendix A). For other details, see the notes to Table 3.
** 5%, * 10% significance level

Table 7: Robustness to more flexible controls for individual ability

	Short run		Medium run		Long run	
	Graduate	> avg tech score	3-year attrition	3-year behavior	> avg job score	Promoted
	(1)	(2)	(3)	(4)	(5)	(6)
Ability surplus	-0.145**	0.277**	0.490**	0.069*	0.381**	0.462**
	(0.040)	(0.101)	(0.069)	(0.036)	(0.114)	(0.106)
Ability deficit	0.016	-0.712**	-0.297**	-0.179**	-0.512**	-0.571**
	(0.060)	(0.136)	(0.112)	(0.070)	(0.198)	(0.157)
IV	Yes	Yes	Yes	Yes	Yes	Yes
DV mean	0.943	0.426	0.137	0.044	0.277	0.180
N	90,543	90,543	62,449	44,934	34,148	34,148

Notes: The samples and IV specifications are the same as those in Tables 4, 5, and 6, except that the the control set includes indicators for each AFQT percentile instead of a cubic in the individual's AFQT percentile score. Due to sparsity, we include a single indicator for AFQT scores below the 40th percentile.
** 5%, * 10% significance level

Table 8: Robustness to controls for measures of peer quality

	Short run		Medium run		Long run	
	Graduate	> avg tech score	3-year attrition	3-year behavior	> avg job score	Promoted
	(1)	(2)	(3)	(4)	(5)	(6)
Ability surplus	-0.163**	0.238**	0.494**	0.085**	0.413**	0.444**
	(0.044)	(0.106)	(0.069)	(0.038)	(0.116)	(0.113)
Ability deficit	0.099	-0.668**	-0.406**	-0.214**	-0.673**	-0.606**
	(0.072)	(0.153)	(0.139)	(0.093)	(0.255)	(0.208)
Cohort peer mean AFQT	0.029	0.221*	0.077	-0.062	-0.145	-0.046
	(0.053)	(0.129)	(0.097)	(0.060)	(0.163)	(0.144)
AFQT percentile rank in cohort	-0.025*	-0.105**	0.011	0.011	0.089**	-0.004
	(0.013)	(0.031)	(0.027)	(0.022)	(0.045)	(0.036)
IV	Yes	Yes	Yes	Yes	Yes	Yes
DV mean	0.943	0.426	0.137	0.044	0.277	0.180
N	90,543	90,543	62,449	44,934	34,148	34,148

Notes: The samples and IV specifications are the same as those in Tables 4, 5, and 6, except that two peer-related variables are added to the control set: i) cohort peer mean AFQT, defined as the average AFQT score for individuals graduating basic training in the same week and career field, and ii) own AFQT percentile rank within that cohort.
** 5%, * 10% significance level

Table 9: Heterogeneity by gender

	Short run		Medium run		Long run	
	Graduate	> avg tech score	3-year attrition	3-year behavior	> avg job score	Promoted
	(1)	(2)	(3)	(4)	(5)	(6)
Ability surplus	-0.134**	0.176	0.393**	0.117**	0.370**	0.489**
	(0.046)	(0.109)	(0.066)	(0.042)	(0.132)	(0.125)
Ability deficit	0.135**	-0.672**	-0.504**	-0.209**	-0.401*	-0.564**
	(0.068)	(0.146)	(0.129)	(0.083)	(0.235)	(0.192)
Surplus × female	-0.002	0.703**	0.262*	-0.124	0.470*	-0.131
	(0.072)	(0.175)	(0.150)	(0.080)	(0.285)	(0.252)
Deficit × female	-0.476**	-0.923**	0.670**	-0.110	-1.003**	-0.204
	(0.112)	(0.219)	(0.181)	(0.123)	(0.318)	(0.268)
IV	Yes	Yes	Yes	Yes	Yes	Yes
DV mean if male	0.946	0.431	0.128	0.047	0.290	0.187
DV mean if female	0.930	0.411	0.174	0.031	0.231	0.157
N	90,543	90,543	62,449	44,934	34,148	34,148

Notes: The samples and IV specifications are the same as those in Tables 4, 5, and 6, except that the ability surplus and deficit variables are interacted with an indicator for whether the individual is female. Interactions between this indicator and the cubic in own AFQT percentile are also added to the control set. ** 5%, * 10% significance level

Table 10: Heterogeneity by race/ethnicity

	Short run		Medium run		Long run	
	Graduate	> avg tech score	3-year attrition	3-year behavior	> avg job score	Promoted
	(1)	(2)	(3)	(4)	(5)	(6)
Ability surplus	-0.174**	0.235**	0.475**	0.088**	0.488**	0.496**
	(0.042)	(0.107)	(0.073)	(0.038)	(0.117)	(0.113)
Ability deficit	0.099	-0.710**	-0.521**	-0.165**	-0.319	-0.465**
	(0.069)	(0.143)	(0.133)	(0.079)	(0.243)	(0.188)
Surplus × Black	0.122	0.612**	0.003	0.058	-0.655*	-0.421
	(0.088)	(0.206)	(0.211)	(0.162)	(0.357)	(0.269)
Deficit × Black	-0.137	-0.606**	0.511**	-0.109	-1.228**	-0.645**
	(0.093)	(0.177)	(0.208)	(0.142)	(0.306)	(0.173)
Surplus × Hispanic	0.158	0.065	-0.120	-0.176	-0.208	0.218
	(0.101)	(0.251)	(0.197)	(0.137)	(0.406)	(0.390)
Deficit × Hispanic	0.043	0.195	0.246	-0.060	-0.178	-0.096
	(0.093)	(0.210)	(0.223)	(0.186)	(0.367)	(0.321)
IV	Yes	Yes	Yes	Yes	Yes	Yes
DV mean if White/other	0.943	0.448	0.132	0.037	0.298	0.198
DV mean if Black	0.934	0.346	0.177	0.076	0.195	0.113
DV mean if Hispanic	0.949	0.391	0.115	0.045	0.242	0.151
N	90,543	90,543	62,449	44,934	34,148	34,148

Notes: The samples and IV specifications are the same as those in Tables 4, 5, and 6, except that the ability surplus and deficit variables are interacted with indicators for whether the individual is Black or Hispanic. Innteractions between these indicators and the cubic in own AFQT percentile are also added to the control set.

** 5%, * 10% significance level

Table 11: Outside market incentives

	Potential civilian earnings	Log(potential earnings)
	(1)	(2)
Ability surplus	-81,548**	-0.976**
	(5,206)	(0.059)
Ability deficit	97,650**	1.313**
	(5,407)	(0.065)
IV	Yes	Yes
DV mean	72,943	11.163
N	90,543	90,543

Notes: The sample and IV specification is the same as in Table 4. Potential civilian earnings are imputed based on occupation codes as described in Section 6.
** 5%, * 10% significance level

ONLINE APPENDICES

Job Mismatch and Early Career Success

Julie Berry Cullen, Gordon B. Dahl, Richard De Thorpe

A Data Appendix

A.1 Data Sources and Analysis Sample

We obtained the dataset through a data agreement with the USAF. The dataset was originally compiled under contract with the Human Resources Research Organization in order to conduct a series of studies of force management issues. The sample is all active duty, non-prior service individuals who enlisted in the Air Force between October 1, 2001 (start of FY 2002) and September 30, 2006 (end of FY 2006). The Defense Manpower Data Center (DMDC) provided enlistment information for these individuals. Using their Social Security numbers, these records were merged to data from other military systems in order to track individuals from entry through basic training, subsequent technical training, and service in units for up to five years. The data sources for each stage are listed in Table A1.

For our purposes, we need to observe individuals both in basic training and in technical training. Basic training data identify our proxy for the job market that an individual participates in, while technical training data provide information on the job to which an individual is assigned. Date of entry is a key determinant of coverage, with early entrants unlikely to be observed in the Basic Training Management System (BTMS) extract and late entrants unlikely to be observed in the Technical Training Management System (TTMS) extract. Thus, we restrict attention to the subset (128,729 of 155,761) that entered between March 15, 2002 and May 31, 2006 with non-missing aptitude scores, as detailed in the top section of Table A2. Summary statistics for background characteristics for this sample are shown in column 1 of Table A3.

Starting from this baseline sample, we make the additional exclusions detailed in the bottom section of Table A2 to build our analysis sample. First, an individual has to be observed in and graduate from basic training. Figure A1 panel a shows that availability of BTMS records hovers around 80% for entry cohorts prior to the start of the second quarter of 2003, when it jumps to 90% and then steadily increases to near complete coverage for the most recent cohorts. The characteristics of those whose training records were preserved in the system are typical other than for timing of entry (column 2 of Table A3), as are the 92% of these entrants that graduate from basic training (column 3 of Table A3). Figure A1 panel b shows that the graduation rate is relatively constant across entering cohorts.

Second, after graduating, an individual has to be observed in the TTMS database which indicates the specific job for which the individual receives technical training. Only those that attend courses managed by the Air Education and Training Command appear in the database. Those assigned training managed by another service (e.g., explosive ordinance disposal) or who attend Air Force schools under different management (e.g., pararescue) do not appear. Figure A1 panel c shows that around 90% of graduates from early basic training cohorts have TTMS records, and this drops to around 75% for later cohorts that are less likely to have completed technical training at the time the extract was pulled. Among those with TTMS records, our final sample selection is to drop linguists since these individuals are required to demonstrate language proficiency on a test we do not observe. Those (non-linguists) with TTMS records tend to be from earlier entry cohorts than the typical basic training graduate, though they are not otherwise too dissimilar (column 4 of Table A3). This final column is our analysis sample.

As is clear from the progression across Figures A1 panels a-c, our analysis sample captures only a subset of entries that occur over our analysis period. The incomplete coverage due to basic training records that were not retained and airmen who failed to graduate basic training are not likely to be endogenous to job assignment, since basic training is generic. However, the incomplete coverage due to missing technical training records may be related to job assignments through two channels. We do not observe jobs for i) individuals assigned to technical training outside of TTMS coverage, or ii) individuals from later cohorts assigned to jobs with extended training programs, since the training records are only for completed training spells. To address the first issue, we define job markets excluding these individuals, effectively treating these as segmented markets. To address the second, we show robustness to excluding later cohorts.

A.2 Individual Characteristics

The individual characteristics used in our analysis are detailed in Table A4. Most of these are gathered at regional military entrance processing stations (MEPS) to determine whether a recruit is eligible for service. These stations administer medical and aptitude examinations, conduct extensive interviews, and provide job counseling. A physical profile is generated by a medical examiner to indicate any physical limitations. Scores on subsections of the Armed Services Vocational Aptitude Battery (ASVAB) are used to construct the percentile on the Armed Forces Qualifying Test (AFQT), as well as percentiles on four composite aptitude areas that map to career fields: mechanical, administrative, general, and electronics. A minimum AFQT score is required for an individual to be eligible to serve, while the aptitude area scores determine the set of jobs a recruit is eligible to pursue. Recruits who would otherwise be disqualified, such as for medical conditions or prior law violations, can be granted waivers. When enlisting, individuals can choose active duty terms of 4 or 6 years, with the longer option bringing an accelerated promotion schedule across pay grades. The entering pay grade is usually the lowest (E-1), but is raised to E-2 or E-3 for those with qualifying college credits or who participated in high school ROTC-type programs.

A.3 Jobs and Job Markets

We characterize job markets by all those graduating basic training at the same time in the same career field. In the USAF, jobs map to Air Force Specialty Codes (AFSCs). In our analysis sample, there are 134 distinct jobs.

We infer the career field relevant to an individual from the assigned job. Since individuals list one of the career fields among their job preference set when first enlisting, we presume their potential and assigned jobs cluster in this field. In most cases, the career field categorization is direct, as the job corresponds to a single field. For the 16% of cases where the job can be accessed via two career fields so that the relevant field is ambiguous, we assign individuals to the one for which they have the greatest predictive similarity (based on area aptitude scores and gender) to others already allocated to the field.

After this process, 30%, 7%, 43%, and 20% are allocated to the mechanical, administrative, general, and electronic career field, respectively. Figure A2 shows patterns in the size of each group across calendar quarters, including the shared contraction in 2005.

Rather than characterizing jobs based on career field aptitude scores, we characterize jobs based on the average AFQT scores of individuals assigned to them. We choose this approach so that the measure of relative ability, which is the difference between own AFQT and job AFQT, is directly comparable across career fields. Further, the correlation between individuals' AFQT scores and their career field aptitude scores is high (0.85).

A.4 Outcomes

The definitions of the short-, medium-, and long-run outcomes we use in our analysis are detailed in Table A5.

The short-run outcomes are observed during the phase of training that follows basic training, which is technical training for assigned jobs. Technical training typically lasts 3-4 months, and rarely lasts more than 6 months. Given the way we construct our analysis sample, these outcomes are never missing.

Our medium-run outcomes include indicators for two- and three-year retention and criminal incidents. These are always observed for those in our analysis sample who enter early enough for sufficient time to have passed when the administrative data extracts were pulled.

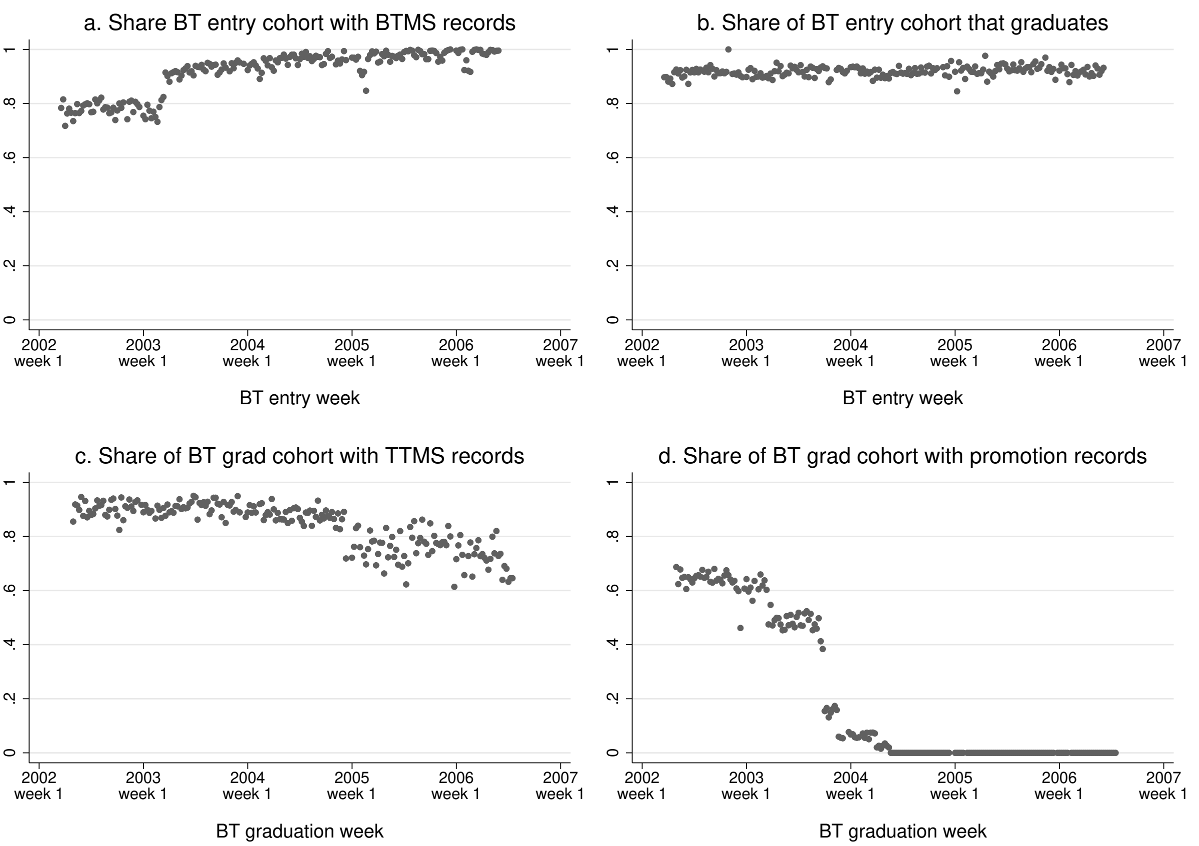
Our long-run outcomes are based on promotion under the Weighted Airman Promotion System (WAPS). Most airmen graduate basic training with a rank of E-1 and automatically advance in steps to E-4 after three years of service. Promotion to the first non-commissioned officer level, E-5, is then competitive. To be eligible to compete, individuals need to meet minimum requirements for time in service, time in pay grade E-4, and job skill level (achieved through on-the-job training or coursework) by an annual cut-off date.

For promotion-eligible individuals, total promotion scores are calculated based on points received for time in service, time spent in pay grade level E-4, enlisted performance report scores, awards and decorations, and scores on two exams. Most receive perfect performance scores, and few have awards or decorations. One of the exams, the Specialty Knowledge Test, assesses job-specific knowledge. The other, the Promotion Fitness Exam, assesses general knowledge about the Air Force. Competing with individuals not yet successfully promoted from prior cohorts, those with the highest total promotion scores within each specific job (AFSC) are selected to fill available vacancies.

Given the time-in-service requirements, individuals in our sample would have had to graduate basic training in one of our earlier cohorts to possibly be observed in the promotion extract. Figure A1 panel d shows that the share that is promotion eligible hovers near 65% for basic training graduation cohorts through mid-March 2003. This share then drops to around

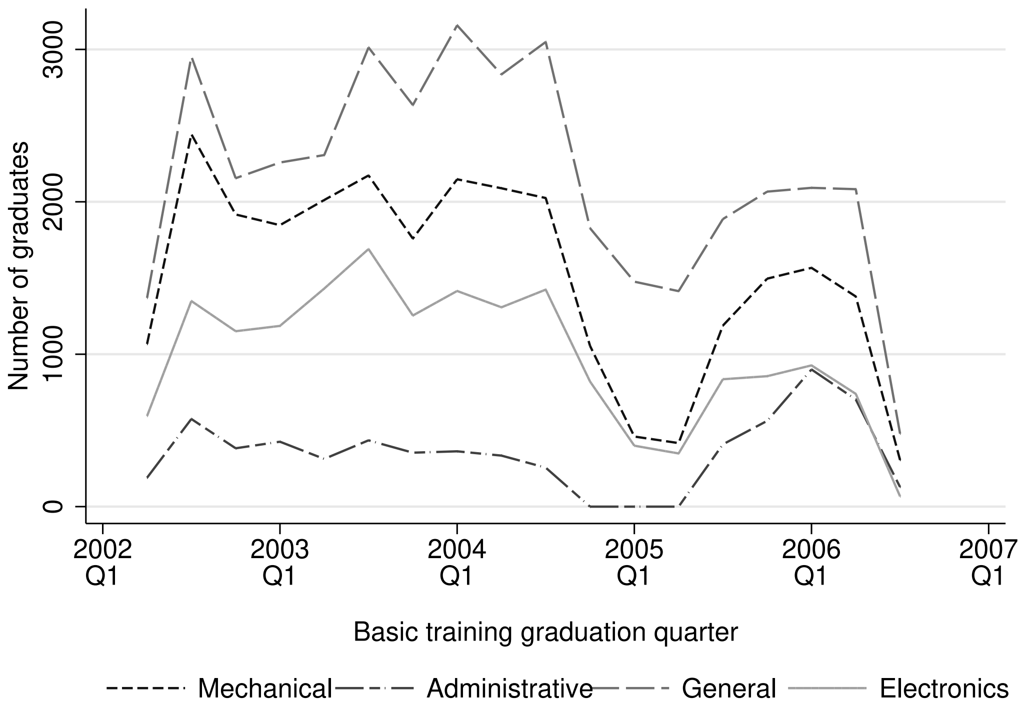
50% before falling off dramatically after mid-September 2003. Thus, when studying these long-term outcomes, we only consider basic training graduation cohorts that graduated by mid-September 2003 and also show robustness to restricting the sample further to those occurring prior to mid-March 2003.

Figure A1: Data coverage over time



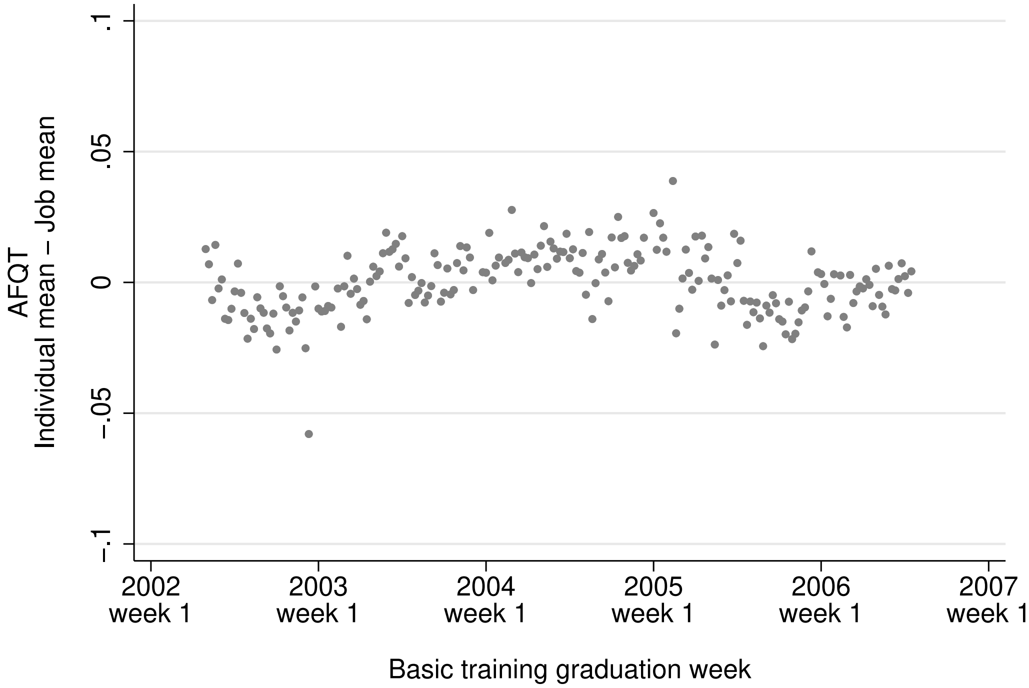
Notes: Panel a shows the share of the baseline sample of entrants that have basic training (BT) records available. Panel b shows the share of those with basic training records that graduate basic training. Panel c shows the share of basic training graduates that have records from technical training. Panel d shows the share of basic training graduates with technical training records that also have promotion records available.

Figure A2: Number of basic training graduates over time by career field



Notes: The figure shows the number of basic training (BT) graduates in our analysis sample by quarter of graduation and career field.

Figure A3: Average individual ability minus average job ability by cohort



Notes: This graph shows the difference between the two variables plotted in Figure 3. The figure shows the average AFQT percentile score for individuals minus the average AFQT percentile score for available jobs, by basic training graduation cohort.

Table A1: Data sources and sample frames

Phase	Data source	Sample frame
Entry into service	Defense Manpower Data Center (DMDC)	FY02-06 active-duty non-prior service entries
Basic training	Basic Training Management System (BTMS)	Subset with record in database as of March 2007 extract
Technical training	Technical Training Management System (TTMS)	Subset with record in database as of January 2007 extract
In-service attrition	DMDC Loss Files	Recorded up through August 31, 2007 24 months: Entered by Aug 31, 2005 36 months: Entered by Aug 31, 2004
In-service criminal activity	DMDC Defense Incident-based Reporting System	Recorded up through December 31, 2006 24 months: Entered by Dec 31, 2004 36 months: Entered by Dec 31, 2003
Promotions	Military Personnel Data System	Subset with record in database as of February 2007 extract (Graduated basic training by mid-September 2003)

Table A2: Sample construction

Sample frame	Number of individuals
All FY02-06 active-duty non-prior service enlisted Air Force entrants	155,761
Drop if entry week cohort has fewer than 100 individuals	70
Drop if entered prior to March 15, 2002 (low BTMS coverage)	15,144
Drop if entered after May 31, 2006 (low TTMS coverage)	10,376
Drop if missing AFQT score	1,442
Entrants March 15, 2002 to May 31, 2006 (with non-missing test scores)	128,729
Drop if not observed in BTMS extract	11,310
Drop if does not graduate from basic training	9,596
Drop if basic training graduation cohort has fewer than 100 individuals	132
Drop if not observed in TTMS extract	16,320
Drop if linguist or fewer than 10 basic training peers in same career field	828
Analysis sample	90,543

Table A3: Summary statistics

	Baseline	In BTMS	Graduated basic training	In TTMS
	(1)	(2)	(3)	(4)
Own AFQT percentile	0.68	0.68	0.68	0.67
Gender: female	0.23	0.23	0.22	0.21
Race/ethnicity: Black (non-Hispanic)	0.14	0.15	0.15	0.15
Race/ethnicity: Hispanic	0.09	0.09	0.10	0.10
Race/ethnicity: other non-White	0.06	0.06	0.06	0.06
Age: 19-20 years	0.39	0.39	0.39	0.39
Age: 21-22 years	0.16	0.16	0.16	0.15
Age: 23+ years	0.11	0.11	0.11	0.10
Education: has some college	0.09	0.08	0.08	0.08
Family status: married	0.08	0.09	0.09	0.08
Family status: has 1+ dependents	0.07	0.07	0.07	0.07
Health: no physical limitations	0.78	0.78	0.78	0.78
Health: medical failure on initial exam	0.09	0.09	0.09	0.09
Health: BMI overweight or obese	0.33	0.34	0.34	0.34
Enlistment option: has conduct waiver	0.07	0.07	0.07	0.07
Enlistment option: has medical waiver	0.05	0.05	0.05	0.05
Enlistment option: term >4 years	0.48	0.50	0.50	0.51
Enlistment option: advanced pay grade	0.25	0.24	0.25	0.23
Census division: New England	0.03	0.03	0.03	0.03
Census division: Middle Atlantic	0.10	0.10	0.10	0.10
Census division: East North Central	0.12	0.12	0.12	0.12
Census division: West North Central	0.08	0.08	0.08	0.08
Census division: South Atlantic	0.20	0.20	0.20	0.20
Census division: East South Central	0.08	0.08	0.08	0.08
Census division: West South Central	0.16	0.16	0.16	0.16
Census division: Mountain	0.07	0.07	0.07	0.07
Census division: Pacific	0.15	0.15	0.15	0.15
Fiscal year of entry: 2002	0.16	0.14	0.14	0.15
Fiscal year of entry: 2003	0.28	0.26	0.26	0.28
Fiscal year of entry: 2004	0.26	0.27	0.27	0.28
Fiscal year of entry: 2005	0.15	0.16	0.16	0.15
Fiscal year of entry: 2006	0.16	0.17	0.17	0.14
Observations	128,729	117,419	107,691	90,543

Notes: Each column shows means for a different sample, starting with the broadest in column 1 and then sequentially narrowing to the analysis sample across columns 2-4.

Table A4: Description of individual characteristics

Variable name	Variable description
AFQT percentile	Percentile score (nationally normed) on the Armed Forces Qualifying Test (AFQT), constructed from the math and verbal subsections of the Armed Services Vocational Aptitude Battery (ASVAB)
Gender: female	Indicator for female
Race/ethnicity	Indicators for race/ethnicity groups ("other non-White" includes Asian/Pacific Islander, American Indian/Alaskan, other, and missing)
Age	Age (in years) at entry date
Education: has some college	Attended college for at least 1 semester
Family status: married	Indicator for married
Family status: has 1+ dependents	Indicator for has 1 or more dependents
Health: no physical limitations	Indicator for determined to "have a high level of medical fitness, fit for any military assignment" along the 6 physical factors evaluated
Health: medical failure on initial exam	Indicator for having a medical failure (such as for weight) at the initial MEPS physical exam, which might be resolved by the time of entry and thus not require a waiver for enlistment
Health: BMI overweight or obese	Indicator for BMI in overweight or obese range
Enlistment option: has conduct waiver	Required waiver to enlist due to a prior law violation
Enlistment option: has medical waiver	Required waiver to enlist due to a disqualifying medical condition
Enlistment option: term > 4 years	Initial contract term is more than 4 years
Enlistment option: advanced pay grade	Pay grade at time of entry is above pay grade E-1
Census division: New England	MEPS located in the New England Census division: Boston MA, Springfield MA, Portland ME
Census division: Middle Atlantic	Albany NY, Buffalo NY, New York NY, Syracuse, NY, Harrisburg PA, Philadelphia PA, Pittsburgh PA
Census division: East North Central	Chicago IL, Indianapolis IN, Detroit MI, Lansing MI, Cleveland OH, Columbus OH, Milwaukee WI
Census division: West North Central	Des Moines IA, Omaha NE, Kansas City MO, St. Louis MO, Fargo, ND, Minneapolis MN, Sioux Falls SD
Census division: South Atlantic	Jacksonville FL, Tampa FL, Atlanta GA, Baltimore MD, Charlotte NC, Raleigh NC, San Juan PR, Fort Jackson SC, Beckley WV, Richmond VA
Census division: East South Central	Montgomery AL, Louisville KY, Jackson MS, Knoxville TN, Nashville TN, Memphis TN
Census division: West South Central	Little Rock AR, New Orleans LA, Shreveport LA, Oklahoma City OK, Amarillo TX, Dallas TX, El Paso TX, Houston TX, San Antonio TX
Census Division: Mountain	Phoenix AZ, Denver CO, Boise ID, Albuquerque

	NM, Butte MT, Salt Lake City UT
Census Division: Pacific	Anchorage AK, Los Angeles CA, Oakland CA, Sacramento CA, San Diego CA, Honolulu HI, Portland OR, Seattle WA, Spokane WA

Table A5: Outcome variable construction and description

Variable name	Variable description
Short-run technical training outcomes (<6 months)	
Graduate	Indicator for graduates technical training
Graduate on time	Indicator for graduates on time
Graduate with delay	Indicator for graduates after repeating training
Greater than average technical training score	Indicator for graduates and has an above average final course grade score, compared to others in the same job-specific technical training course (i.e., AFSC).
Medium-run retention and behavior outcomes (2-3 years)	
2- and 3-year retention	Indicators for retention (i.e., no attrition) within 2 and within 3 years of entry. Separations that are not considered attrition are those that are due to expiration of terms of service, early release, entry into an officer commissioning program or military service academy, or disability/death. Separations that are considered attrition are those due to disciplinary infractions or other offenses, behavior disorders, failures to meet physical standards, poor performance, and pregnancy/parenthood. The small share (<1.5%) that separates for reasons that could not be determined are treated as having been retained.
2- and 3-year behavior	Indicators for any violent or property criminal incidents within 2 and within 3 years of entry
Long-run promotion outcomes (4-5 years)	
Promotion eligible	To be eligible for promotion to E-5, the individual must have 6 months in pay grade E-4, 36 months in service, and a job skill level of 5 ("journeyman"), as well as have taken the general and job specialty knowledge exams.
Promotion eligible and above average job score	Indicator for promotion eligible and scoring above the mean on the specialty knowledge exam among those competing for promotion within the same job in our sample. The 100-question multiple-choice exam is specific to the AFSC. The questions are drawn from career development courses that provide ongoing technical training for skill upgrading.
Promotion eligible and above average USAF knowledge score	Indicator for promotion eligible and scoring above the mean on the general knowledge exam among all those competing for promotion in our sample. The general knowledge (i.e., promotion fitness) exam tests a wide range of Air Force knowledge, including history, organization, regulations, and practices. The raw score is the number of correct responses out of the 100 items on the exam.
Perfect performance evaluations	Perfect scores on the most recent enlisted performance reports, which are annual evaluations of performance, character, and career progression.
Promoted	Indicator for successful promotion to E-5 on first attempt

B Calculation of Intensive Margin Effects

For several measures of performance, we only observe them if the individual has not attrited from the sample. For example, we only observe a test score if the individual persists, and therefore takes the test. To deal with this, we create composite binary outcomes which equal one if the performance measure is both observed (e.g., taking the test) and favorable (e.g., scoring above the mean on the test), and zero otherwise. These composite outcomes are interesting in themselves, as both components can be thought of as positive outcomes.

It would also be interesting to know the effects of ability surplus and deficit on conditional outcomes – e.g., how ability surplus and deficit affect performance on a test conditional on an individual taking the test. We call these "intensive margin" effects, and report our estimates of these in the rows labeled im-surplus and im-deficit in Tables 4 and 6. In this appendix, we explain how we calculate these effects and the underlying assumption.

To understand how we calculate the intensive margin effects, consider a test which measures performance. We are interested in how ability surplus (s) and ability deficit (d) affect the probability an individual scores high on the test (H=1) conditional on taking the test (T=1). Below, we lay out the procedure for ability surplus; a similar logic holds for ability deficit.

Using the definition of conditional expectation, the probability an individual both takes the test and scores high on it can be written as:

$$\mathbb{P}_x(H \cap T) = \mathbb{P}_x(H|T)\mathbb{P}_x(T) \tag{1}$$

where the subscript x indicates the probabilities are evaluated based on the observable and unobservable characteristics of compliers. Taking the total derivative with respect to ability surplus, s, yields:

$$\frac{d\mathbb{P}_x(H \cap T)}{ds} = \mathbb{P}_x(H|T)\frac{d\mathbb{P}_x(T)}{ds} + \mathbb{P}_x(T)\frac{d\mathbb{P}_x(H|T)}{ds} \tag{2}$$

Rearranging terms yields:

$$\frac{d\mathbb{P}_x(H|T)}{ds} = \left[\frac{d\mathbb{P}_x(H \cap T)}{ds} - \mathbb{P}_x(H|T)\frac{d\mathbb{P}_x(T)}{ds}\right] / \ \mathbb{P}_x(T) \tag{3}$$

Intuitively, if ability surplus has no impact on test-taking (i.e., $\frac{d\mathbb{P}_x(T)}{ds}$=0), the effect on the intensive margin would simply be our estimate of the impact on the composite outcome, $\frac{d\mathbb{P}_x(H \cap T)}{ds}$, rescaled by the likelihood of taking the test, $\mathbb{P}_x(T)$. If ability surplus affects test-taking, some of the estimated impact on the composite outcome would instead be attributable to the extensive margin.

We can estimate two of the terms on the right-hand side, $\frac{d\mathbb{P}_x(H \cap T)}{ds}$ and $\frac{d\mathbb{P}_x(T)}{ds}$, using our instrumental variables strategy (using the composite outcome and an indicator for taking the test as the dependent variables). The remaining two terms, $\mathbb{P}_x(H|T)$ and $\mathbb{P}_x(T)$, are expected means for compliers, and hence unobservable. We assume these probabilities are the same for compliers as they are for all individuals, and estimate them using the entire sample. Despite this strong assumption, we view the intensive margin calculations to be a useful exercise to complement our estimates for the composite outcomes.

To evaluate the statistical significance of the intensive margin effects, we use 500 bootstrap iterations.

C Additional Results

Table C1: Alternative functional form for relative ability

	Short run		Medium run		Long run	
	Graduate	> avg tech score	3-year attrition	3-year behavior	> avg job score	Promoted
	(1)	(2)	(3)	(4)	(5)	(6)
Relative ability quintile 1	-0.015	0.096**	0.081**	0.015	0.096**	0.092**
	(0.010)	(0.024)	(0.018)	(0.010)	(0.029)	(0.027)
Relative ability quintile 2	0.010	0.073**	0.017	0.004	0.052**	0.050**
	(0.008)	(0.017)	(0.013)	(0.008)	(0.026)	(0.022)
Relative ability quintile 4	-0.004	-0.066**	-0.015	-0.005	-0.007	-0.027
	(0.009)	(0.017)	(0.016)	(0.010)	(0.029)	(0.024)
Relative ability quintile 5	0.012	-0.071**	-0.063**	-0.022	-0.053	-0.056*
	(0.012)	(0.024)	(0.024)	(0.016)	(0.035)	(0.030)
IV	Yes	Yes	Yes	Yes	Yes	Yes
DV mean	0.943	0.426	0.137	0.044	0.277	0.180
N	90,543	90,543	62,449	44,934	34,148	34,148

Notes: The samples and IV specifications are the same as those in Tables 4, 5, and 6, but replacing ability surplus and deficit with quintiles of relative ability, where quintile 1 is the top quintile and corresponds to the largest ability surpluses.
** 5%, * 10% significance level

Table C2: Robustness to broader job market definition: ±4 weeks

	Short run		Medium run		Long run	
	Graduate	> avg tech score	3-year attrition	3-year behavior	> avg job score	Promoted
	(1)	(2)	(3)	(4)	(5)	(6)
Ability surplus	-0.261**	0.129	0.560**	0.065	0.240**	0.384**
	(0.045)	(0.108)	(0.071)	(0.041)	(0.122)	(0.123)
Ability deficit	0.275**	-0.539**	-0.472**	-0.262**	-0.616**	-0.659**
	(0.078)	(0.153)	(0.153)	(0.087)	(0.262)	(0.203)
IV	Yes	Yes	Yes	Yes	Yes	Yes
DV mean	0.943	0.426	0.137	0.044	0.277	0.180
N	90,543	90,543	62,449	44,934	34,148	34,148

Notes: The samples and IV specifications are the same as those in Tables 4, 5, and 6, but we change the definition of the job market to include those cohorts graduating basic training up to 4 weeks earlier and later.

** 5%, * 10% significance level

Table C3: Robustness to excluding early entering cohorts

	Short run		Medium run		Long run	
	Graduate	> avg tech score	3-year attrition	3-year behavior	> avg job score	Promoted
	(1)	(2)	(3)	(4)	(5)	(6)
Ability surplus	-0.099**	0.208*	0.446**	0.015	0.758**	0.775**
	(0.050)	(0.120)	(0.088)	(0.057)	(0.215)	(0.197)
Ability deficit	0.042	-0.725**	-0.259	-0.083	-0.878**	-0.765**
	(0.073)	(0.163)	(0.164)	(0.133)	(0.421)	(0.318)
IV	Yes	Yes	Yes	Yes	Yes	Yes
DV mean	0.946	0.430	0.128	0.042	0.245	0.133
N	66,345	66,345	38,253	20,739	9,955	9,955

Notes: The IV specifications are the same as those in Tables 4, 5, and 6, but the sample is restricted to entry cohorts with near complete availability of basic training records. These individuals enter after the first quarter of 2003 (see Figure A1, panel a).
** 5%, * 10% significance level

Table C4: Robustness to excluding later entering cohorts

	Short run		Medium run		Long run	
	Graduate	> avg tech score	3-year attrition	3-year behavior	> avg job score	Promoted
	(1)	(2)	(3)	(4)	(5)	(6)
Ability surplus	-0.319**	0.457**	0.698**	0.164**	0.231*	0.353**
	(0.080)	(0.189)	(0.097)	(0.048)	(0.140)	(0.139)
Ability deficit	0.239*	-0.916**	-0.795**	-0.327**	-0.296	-0.546**
	(0.145)	(0.297)	(0.202)	(0.104)	(0.260)	(0.244)
IV	Yes	Yes	Yes	Yes	Yes	Yes
DV mean	0.931	0.421	0.154	0.045	0.297	0.207
N	21,042	21,042	21,042	21,042	21,042	21,042

Notes: The IV specifications are the same as those in Tables 4, 5, and 6, but the sample is restricted to entry cohorts with near complete availability of technical training and promotion records. These individuals graduate basic training before mid-March 2003 (see Figure A1, panel d).
** 5%, * 10% significance level