Institute for Research on Poverty
Discussion Paper no. 1017-93

The "Misnorming" of the U.S. Military's Entrance Examination and Its Effect on Minority Enlistments

Joshua D. Angrist
Department of Economics
Hebrew University of Jerusalem
and
National Bureau of Economic Research

September 1993

This research was supported by the Institute for Research on Poverty Small Grants Program for 1992–1993 and by the National Science Foundation. Special thanks go to Mike Dove and the staff at the Defense Manpower Data Center for data on applicants to the military and to Michael Sarel for outstanding research assistance. Thanks also go to participants in the IRP Small Grants Research Seminar held in May 1993 for helpful comments. The views expressed here are solely those of the author alone and in no way reflect the views of any U.S. government agency or employee.

Abstract

The score a prospective recruit must earn on the military's entrance examination was raised in 1980 in response to an error discovered in the score scale previously used. Raising this score led to a reduction in enlistments, especially among minorities. Recent plans to reduce the military have also had an adverse impact on service opportunities for minority applicants. This paper considers empirical and theoretical aspects of the relationship between entrance standards and minority representation in the military, focusing on racial differences in the proportion of qualified applicants who enlist. The results suggest that increases in test score standards have a large impact on minority enlistment not only because minorities have lower scores, but also because qualified minority applicants are far more likely than other qualified applicants to enlist.

The "Misnorming" of the U.S. Military's Entrance Examination and Its Effect on Minority Enlistments

1. INTRODUCTION

Many blacks in the United States regard the military as an attractive employment option. The Defense Department's current efforts to reduce the Armed Forces may therefore frustrate the plans of a lot of young blacks. One approach being used to cut the number in uniform is to increase entrance standards (Defense Department 1992). If military service is a desirable career path for blacks, tougher admissions requirements could have an especially large impact on them. Indeed, Boesel (1992) and Laurence (1992) predict that higher entrance standards will disqualify a greater percentage of minority applicants than of white applicants.

Admissions standards have been raised from time to time. Probably the most famous episode occurred in 1980, when the score a prospective recruit must earn on the military's entrance examination--the Armed Services Vocational Aptitude Battery (ASVAB)--was raised in response to an error discovered in the scoring scale previously used. What is now known as the "misnorming" of the ASVAB had a profound impact on enlistments, particularly among blacks (U.S. Congress 1989). In this paper I take a new look at the effect the ASVAB misnorming had on white and black enlistments during the two years following the incident.

The next section presents data documenting recent changes in social representation in the military. Section 3 outlines a simple behavioral model of the effect of entrance criteria on enlistment in the military. The model is not specific to the military selection process and highlights phenomena common to other application and screening processes such as those used to determine attendance at selective colleges. Section 4 describes military applicant data for the period 1976–1982 and provides a brief history of military screening in the late 1970s. The episode now known as the "misnorming" of the Armed Services Vocational Aptitude Battery (ASVAB) is recounted, and a strategy is outlined for

using this episode to estimate the effect of screening criteria on social representation in the Armed Forces. Section 5 presents empirical results from this strategy, and Section 6 concludes.

2. SOCIAL REPRESENTATION IN THE MILITARY: 1979–1992

In 1979, only 64 percent of Army recruits were high school graduates, and 46 percent had Armed Forces Qualification Test (AFQT) test scores between the 10th and 30th percentiles of the national youth AFQT score distribution.[1] Beginning in 1980, however, test scores and schooling levels of newly enlisted soldiers improved steadily. This improvement was partly attributable to the correction of incorrect ASVAB scoring procedures in the late 1970s (described more fully below) and to legislative limits on the number of low-scoring enlistments;[2] it was also due to poor civilian labor market conditions and new packages of veterans benefits that made military service relatively attractive for many young people.[3]

By 1987, 93 percent of all new recruits had a high school diploma, and 95 percent had scores in the top 70 percent of the AFQT score distribution (Categories I–III). At the same time, the fraction of minority enlistments fell from an all time high of 35.5 percent in 1980 to 24.2 percent in 1983, rising again to 28.1 percent in 1987 (Defense Department 1988, p. II-33). The fraction of nonwhites in the U.S. population rose over the same period from 15 to 18 percent. It is important to note, however, that average minority representation during the period 1976–1980 was a more modest 25 percent, actually somewhat less than the 1985–1987 average of 27 percent. This stability in average representation was partly attributable to improved representation in services other than the Army, at a time when nonwhite representation in the Army declined dramatically. Nevertheless, correction of the ASVAB test scoring error and increases in education and score-related entrance requirements appear to have made enlistment more difficult for minority applicants (Cooper 1977; Eitelberg 1988; U.S. Congress 1989; Defense Department 1992).

The Bush administration's plan to reduce the size of the Armed Forces by 25 percent over fiscal years 1991–1995 aggravated concerns about opportunities for minorities in the military (Laurence 1992). Recent Defense Manpower Data Center (DMDC) statistics on enlistments, tabulated by race and test scores and presented in Tables 1 and 2, appear to substantiate these fears. The number of enlistments by men and women without prior service fell by 27 percent between 1989 and 1992, from 278,000 to 202,000. There was a nearly 30 percent decline in male enlistments. Enlistments by white males declined by only 25 percent, however, while enlistments by black men--the group that appears to have been hardest hit by military downsizing--fell by 47 percent.

An important and related change in the characteristics of newly enlisted soldiers is documented in Table 2, which shows recent enlistment data tabulated by AFQT group. In 1989, there were 17,700 new enlistments by men with AFQT scores in Category IV. In 1992, there were 340. Because black applicants are much more likely to have scores in Category IV, the elimination of Category IV enlistments clearly contributed to the decline in black representation in new enlistments from 20 percent in 1989 to 15 percent in 1992.

But the elimination of Category IV enlistments alone cannot explain so large a decline. Excluding all Category IV enlistments from the 1989 data only reduces black representation to 18 percent. Moreover, enlistments by men without a high school degree are disproportionately likely to be by white applicants. Tabulations of data not shown in the table indicate that excluding both Category IV enlistments and enlistments by men without a high school degree would leave black representation at 18.5 percent. Thus, the precipitous decline in black representation also suggests that increases in entrance standards can have a larger impact on minority applicants than forecasts based solely on the distribution of entrance qualifications would predict. In the next section I outline a simple behavioral model of enlistment and minority representation.

TABLE 1

Number of Enlistments (in Thousands), by Race

Year	Men				Women				Total
	White	Black	Hispanic	Other	White	Black	Hispanic	Other	
1989									
Enlistments	168.5	48.6	15.1	7.6	23.4	11.4	2.2	1.2	278
Percentage	70.3	20.3	6.3	3.2	61.3	29.8	5.8	3.1	
1990									
Enlistments	136.7	37.7	13.8	6.0	18.4	8.5	1.8	.9	224
Percentage	70.4	19.4	7.1	3.1	62.2	28.7	6.1	3.0	
1991									
Enlistments	133.2	27.6	13.2	5.8	16.8	6.3	1.7	.9	206
Percentage	74.1	15.4	7.3	3.2	65.4	24.5	6.6	3.5	
1992									
Enlistments	126.9	26.3	13.1	3.5	19.3	7.2	2.2	1.1	202
Percentage	74.7	15.4	7.7	2.1	64.8	24.2	7.4	3.7	

Source: Defense Manpower Data Center.

Notes: Percentages are of the total by sex and may not add to 100 because of rounding.

TABLE 2

Number of Enlistments (in Thousands), by AFQT Group

Year	Men				Women			
	I–II	IIIa	IIIb	IV	I–II	IIIa	IIIb	IV
1989								
Enlistments	90.0	61.3	67.9	17.7	14.4	12.2	11.1	.09
Percentage	37.9	25.6	28.3	7.4	37.8	32.0	29.1	.00
1990								
Enlistments	76.8	52.3	56.9	6.8	12.2	10.7	6.5	.02
Percentage	39.6	26.9	29.3	3.5	41.2	36.1	22.0	.00
1991								
Enlistments	78.1	49.2	49.8	1.06	11.1	9.7	4.7	.01
Percentage	43.4	27.4	27.7	.01	43.3	37.7	18.3	.00
1992								
Enlistments	77.4	50.1	43.3	.34	12.7	10.7	6.2	.03
Percentage	45.1	29.2	25.2	.00	42.6	35.9	20.8	.00

Source: Defense Manpower Data Center.

Notes: Percentages are of the total by sex and may not add to 100 because of rounding and because of unknown AFQT scores. AFQT Group I includes percentile scores 93–100, AFQT Group II includes percentile scores 65–92, AFQT Group IIIa includes percentile scores 50–64, AFQT Group IIIb includes percentile scores 31–49, and AFQT Group IV includes percentile scores 10–30. AFQT Group V includes percentile scores 1–9. Applicants with scores in Group V are barred from enlistment by law.

3. THE EFFECT OF ADMISSIONS CRITERIA ON MINORITY REPRESENTATION: THEORY

The original rationale for military entrance testing was to screen out applicants who could not be trained to master the skills required of soldiers. Thus, standards should primarily reflect technological advances or changes in the medical knowledge needed by soldiers (Karpinos 1967.) In practice, however, standards have often been changed to accommodate shifting demands for military manpower and changes in manpower availability (Eitelberg, Lawrence, Waters, and Perelman 1984; Karpinos 1967.)

Entrance standards affect the social mix in the military in at least three ways. First, there is the direct effect: higher standards mean fewer individuals from different groups are qualified. If the distribution of qualifications differs across groups, then the direct effect works to change the social composition of the armed forces. Second, entrance standards affect the incentive to apply to the military. Applicants who believe they have little chance of satisfying entrance criteria are less likely to go through a costly and time-consuming enlistment process. Third, changes in standards may change the distribution of enlistment probabilities in the pool of qualified applicants. For example, if applicants in certain groups who are disqualified by increased standards would not have enlisted anyway, then increased standards would have little impact on enlistments by members of this group. On the other hand, groups with a high propensity to follow through with enlistment are more likely to be hurt by increases in standards.

These three features of the applicant screening process are highlighted by a simple behavioral model based on comparisons of expected utilities and costs. First, define the events: E = successful enlistment, A = application for enlistment, Q = qualified for enlistment, and M = membership in a demographic group of interest such as a minority group. Because E occurs only if A and Q occur, the probability of successful enlistment by a member of group M can be decomposed as follows:

$$\Pr[E \mid M] = \Pr[E \mid M, A, Q] \bullet \Pr[Q \mid M, A] \bullet \Pr[A \mid M]. \quad (1)$$

In words, successful enlistment begins with an application, applicants must be found to be qualified, and finally, qualified applicants must follow through with actual enlistment.

The direct effect of qualification standards is $\Pr[Q \mid M, A]$. This probability is given by the distribution of applicant qualifications and military entrance standards; it has no behavioral component. For example, if applicant qualifications are denoted by q and the entrance standard is κ, then the probability of meeting this standard is $\Pr[q > \kappa \mid M, A]$.

The most important modification of the direct effect of entrance standards on minority applicants stems from the fact (documented below) that the probability of enlistment conditional on application and qualification is significantly higher for minorities than for members of other groups. This probability, $\Pr[E \mid M, A, Q]$, multiplies the direct effect of changes in qualifications. Therefore, a given number of disqualified minority applicants would have a larger impact on minority enlistments than would the same number of white disqualifications on white enlistments.

Behavioral components of the enlistment process also include the impact of standards on the decision to apply, an event which occurs with probability $\Pr[A \mid M]$. To model this decision, suppose that the utility of voluntary service and civilian alternatives can be characterized by random variables u_v and u_c. Suppose also that potential applicants to the military face a cost of application of the amount θ, and that the applicants' subjective probability of meeting enlistment standards is p_q. Finally, suppose that the applicants' subjective probability of following through with enlistment, if qualified, is p_e. If the applicants' expectations are realistic, then $p_q = \Pr[Q \mid M, A]$ and $p_e = \Pr[E \mid M, A, Q]$, but this need not be the case. An applicant who chooses to apply by comparing expected utilities enlists if

$$u_v[p_q p_e] + u_c[1 - p_q p_e] - \theta > u_c,$$

or, equivalently,

(2) $$[u_v - u_c]p_q p_e > \theta.$$

Changes in qualifications can be seen to affect the probability of application through their impact on p_q and p_e. These probabilities are subjective assessments by individuals in the pool of applicants. But potential applicants' evaluation of the chances of acceptance are tied to actual acceptance standards by virtue of the recruiter practice of giving prospective applicants a "mini-AFQT" or other preliminary screening test before beginning the official application process.[4] Thus, it seems likely that an increase in entrance standards will tend to reduce an applicant's subjective probability of meeting enlistment standards, p_q. This reduces the incentive to enlist because it reduces the expected payoff from application. Groups for whom the increase in standards reduces p_q the most are more likely to be discouraged from applying.

The application decision rule, (2), also has the simple implication that anyone who chooses to apply to the military (and faces positive application costs) expects military service to be more attractive than a civilian career at the time of application. Yet applications data show that large numbers of qualified applicants do not follow through with enlistment. Orvis and Gahart (1990, p. vii) suggest that one reason many qualified applicants do not enlist is that, "even after application, civilian job opportunities, social support for enlisting, college plans, and finances have substantial effects on the enlistment decisions of young men."[5]

To model this process of reassessment, I assume that information about civilian opportunities can be characterized by random variables added to applicants' original evaluations of civilian and military opportunities. In particular, revised assessments are characterized by:

$$\tilde{u}_v = u_v + \eta_v \quad \text{and} \quad \tilde{u}_c = u_c + \eta_c,$$

where η_v and η_c are random variables representing new information. Then we can write

(3) $\Pr[E \mid M, A, Q] = \Pr [\, \tilde{u}_v - \tilde{u}_c > 0 \mid (u_v - u_c) > \theta/p_q p_e,\ q > \kappa,\ M]$.

Another way to write this is

(4) $\Pr[E \mid M, A, Q] = \Pr [u_v - u_c > \eta \mid (u_v - u_c) > \theta/p_q p_e,\ q > \kappa,\ M]$,

where $\eta \equiv \eta_c - \eta_v \equiv [(u_v - u_c) - (\tilde{u}_v - \tilde{u}_c)]$. Note that even if applicants' revised assessments of civilian and military opportunities are independent of their original assessments, the *change* in relative assessments, η, is likely to be positively correlated with $(u_v - u_c)$.

Changes in enlistment standards can be seen to affect the probability of enlistment by applicants through the conditioning arguments in (3). Consider first conditioning on application. An applicant is someone for whom it is true that $(u_v - u_c) > \theta/p_q p_e$. If an increase in standards raises the application threshold ($\theta/p_q p_e$), then conditioning on application raises the probability that $(u_v - u_c)$ exceeds any constant number. On the other hand, this effect is moderated by the fact that η, the stochastic enlistment threshold, is positively correlated with $(u_v - u_c)$. In the extreme case where the revised assessments, $\tilde{u}_v$ and $\tilde{u}_c$, are independent of the original assessments in the pool of qualified applicants, conditioning on application has no impact on the probability of enlistment.

The last effect captured in equation (4) is a function of the correlation between qualifications, *q*, and individual assessments of civilian and military opportunities. Conditioning on qualification ($q > \kappa$) can either reduce or increase the probability of enlistment when enlistment standards are raised.

This is because qualifications can be either negatively or positively correlated with $\tilde{u}_v - \tilde{u}_c$. Since the military ranks applicants largely on the basis of measures of general human capital (schooling and test scores), an increase in standards could work to reduce $\tilde{u}_v - \tilde{u}_c$ in the pool of qualified applicants. On the other hand, *in the population choosing to apply*, an increase in certain types of qualification standards could work to select those applicants with the best military prospects.

4. APPLICANT SCREENING AND ENLISTMENT: 1976–1982

In January 1976, the Armed Forces introduced a new entrance examination for the purpose of screening applicants to all services. The new ASVAB test, like previous military exams, consisted of a number of subtests that were used individually and in the form of composites. The most important composite was the AFQT, which was used to screen out unqualified applicants and to keep track of the quality of entrants to the military. Recruiters typically attempt to get as many high-AFQT applicants as possible.

A history of the AFQT and other applicant screening tests in use since World War II is presented in Table 3. The AFQT originated as part of the Army General Qualifications Test (AGCT), used to screen and classify servicemen and women during World War II. Eight AFQT test forms common to all services were used as part of service-specific test batteries from 1951 to 1973. In any ongoing testing program like the AFQT, an effort is made to ensure that the test score scale has the same meaning when new versions of the test are introduced. The process of establishing equivalence is called test-equating or norming. AFQT test forms were normed using a technique known as equipercentile equating, described in Maier and Truss (1983, appendix B) and briefly summarized in the appendix to this paper.

Between 1973 and 1975, the military did not require each service to use a common AFQT measure when screening and classifying applicants. During this period, the individual services

TABLE 3

History of Armed Services Applicant Screening Tests

Test Forms	Event	Date
AGCT and NGCT	Used by the Army and Navy for classification and assignment	WWII
AGCT	Used by the Army for enlistment screening	late 1940s
AFQT 1–6 (composite from subtests common to service-specific batteries, e.g., ACB, NBTB)	Implemented by Army	1950
	Implemented by all services	1951
	Each new AFQT form normed to WWII Mobilization Population in 1944 through readministration of AGCT (which included AFQT subtests)	
	Congress outlaws inductions from Mental Group V	June 1951
AFQT 7–8 (composite from subtests common to service-specific batteries)	Used in Vietnam era	1960–1972
	Normed to WWII Mobilization Population as before	
	ASVAB 2 used in high school testing program	1968
	Draft ends	July 1973
Service-specific composites and batteries	ASVAB 3 in Navy, Air Force	1973–1975
	Army uses ACB-73	1973
	ASVAB 4 developed but never used	
	Service-specific norms applied	

(table continues)

TABLE 3 (continued)

Test Forms	Event	Date
ASVAB 5/6/7	Implemented by Defense Department using "original score scale"	January 1976
	Initially normed against Navy and Air Force recruits using ASVAB 2/AFQT as reference and low-ability applicants using ACB-73/AFQT as reference	
	Early evidence of misnorming	April 1976
ASVAB 5/6/7	New "operational score scale" introduced; this lowered scores in Mental Groups I, II, and III	September 1976
	Navy revises mental group definitions to generate ability distribution closer to AFQT from NBTB-75	April 1977
	Center for Naval Analyses (CNA) publishes report showing original scale was much too easy, and 1976 operational scale only partially corrected this	January 1979
	Second CNA study suggests major norming errors at low aptitudes and produces a new scale	May 1979
	Results from Army Research Institute (ARI) study confirm second CNA study and produce another new scale	October 1979
	Three independent studies confirm misnorming and produce ultimate "corrected score scale" normed by the Educational Testing Service to AFQT 7A	Spring 1980

(table continues)

TABLE 3 (continued)

Test Forms	Event	Date
	Corrected score scale adopted, normed to WWII Mobilization Population, used to retrospectively evaluate recruit quality for 1976–1980	October 1980
ASVAB 8/9/10	Replaces ASVAB 5/6/7	October 1980
	ASVAB 8/9/10 normed and score scale calibrated using orthodox test equating procedures and the Vietnam-era AFQT 7A as reference test under WWII Mobilization Population norms	
	Corrected score scale for ASVAB 5/6/7 validated against ASVAB 8/9/10 scale	March 1983
ASVAB 8A	Test form administered to all participants in the National Longitudinal Survey of Youth (NLSY) by NORC to produce a new 1980 Profiles of American Youth Study reference population	Fall 1980
	New 1980 Profiles of American Youth Study score scale introduced	October 1984
	1980 Profiles of American Youth Study norms become standard; historical AFQT data reported against these norms using WWII Mobilization Population/1980 Youth Study conversion tables[a]	
	Corrected WWII Mobilization Population and 1980 Profiles of American Youth Study scales are shown to be similar	

(table continues)

TABLE 3 (continued)

Test Forms	Event	Date
ASVAB 11/12/13	New test forms introduced	October 1984
	1980 Profiles of American Youth Study norms become standard	

Source: Information in this table is compiled from ASVAB Working Group (1980), Maier and Truss (1983), and Maier and Sims (1986).

Notes: AGCT and NGCT are the Army and Navy General Classification Tests. ACB is the Army Classification Battery, NBTB is the Navy Basic Test Battery. ASVAB is the Armed Services Vocational Aptitude Battery. The AFQT is a composite computed from scores on subtests in batteries such as ASVAB, ACB, and NBTB after 1950. From 1950–1953, the AFQT composite was the same as that on the original AGCT, including Verbal, Arithmetic Reasoning, and Spacial Relationships subtests. From 1953–1973 a Tool Knowledge subtest was added. For ASVAB 5/6/7, the AFQT composite was the sum of subtest scores for Word Knowledge, Arithmetic Reasoning, and Spacial Relationships. For ASVAB 8/9/10, the AFQT composite was the sum of Word Knowledge, Arithmetic Reasoning, Paragraph Comprehension, and half the score on the Numerical Operations subtest.

[a]Conversion of historical Military Entrance Processing Command data to 1980 score scales performed in 1985 by Les Willis of the DMDC on the basis of unpublished tables underlying Figures 1–3, pp. 1–18 in Maier and Sims (1986).

employed a variety of different tests, with service-specific scoring norms. The introduction of the ASVAB 5/6/7 forms in 1976 marked a return to a service-wide AFQT. Initially, however, no attempt was made to accurately equate the score scale to norms from the previous reference population of men at risk of being drafted for service in World War II. A variety of ad hoc procedures were used instead.[6] Consequently, AFQT percentile scores on ASVAB 5/6/7 did not reflect the same level of ability as did earlier scores.

Errors in the ASVAB 5/6/7 score scale were first discovered in late 1976 and were partially corrected when a new "operational score scale" was introduced shortly thereafter. But in early 1979, studies by the Center for Naval Analyses found that the operational score scale was still too easy and that large numbers of new recruits would not have been accepted for service under traditional entrance standards. By late 1979 new sets of corrected score scales had been produced by the ASVAB Working Group (in the Office of the Assistant Secretary of Defense for Manpower, Reserves, and Logistics.)

In early 1980 the ASVAB Working Group, in cooperation with the Educational Testing Service, constructed a corrected score scale for ASVAB 5/6/7 that had been carefully normed to a Vietnam-era AFQT test (Form 7A) in samples of eleventh and twelfth graders. At the same time, a new set of forms, ASVAB 8/9/10, was also normed to AFQT 7A (Maier and Truss 1983, Appendix H). In October 1980, ASVAB 8/9/10 replaced ASVAB 5/6/7, and the corrected score scale for ASVAB 5/6/7 became the official scale for the purposes of historical comparisons. Validation studies suggested that AFQT scores on the corrected score scale reflected the same level of ability as pre-1973 AFQT tests. Note, however, that even though scores on ASVAB 5/6/7 and ASVAB 8/9/10 were both normed to AFQT 7A, AFQT scores from the two sets of forms may not be comparable because the formula for the AFQT composite changed (see the notes to Table 3.)[7]

For the purposes of this project, the importance of changes in score scales between 1976 and 1982 is that these changes led to large differences in the probability that low-ability applicants would be accepted into the Armed Forces. An applicant whose AFQT score from a correctly normed test put him in Mental Group IV would have had a much higher chance of being accepted into the military if he had applied between 1976 and 1979 than in late 1980 or later. An applicant in Mental Group V would not have been accepted at all except for misnorming. The incorrect norming of the original ASVAB therefore generates a natural experiment that can be used to identify the effects of entrance standards on social representation in the military.

The data analyzed here come from annual DMDC files containing information on all applicants and entrants to the military in a fiscal year. One computerized record is generated each time an applicant makes contact with military recruiters or examiners. These records indicate the date and type of contact (examination results, enlistment into the Delayed Entry program [DEP], enlistment to active duty, and discharge from DEP without active-duty enlistment). Each record also reports basic demographic information, physical examination results and test scores, and terms and conditions of enlistment for entrants.[8] Files for fiscal years 1976 to 1982 contain the population of applicants used in this study.

Applicant records do not indicate whether applicants eventually enlist in the military. But both applicants' and entrants' records contain Social Security Numbers (SSNs) that I have used to link the two types of records. Details of this link are outlined in Table 4. The match was accomplished by splitting each fiscal-year file, recording one to two million contacts, into separate data sets containing records for applicants and all other contacts. At this stage only applicants and entrants without prior military service who were identified as being sixteen to twenty-five years old when applying or entering the military were retained (row 1 of Table 4 shows initial record counts for applicants). Additional selection criteria retained only seventeen- to twenty-one-year-old applicants

TABLE 4

Military Applicant Population (in Thousands), 1976–1982

Characteristics	Total	Calendar Year 1976	1977	1978	1979	1980	1981	1982
1. Aged 16–25 at application or entry; without prior military service	5,372	908	699	602	666	866	936	695
2. Keep if aged 17–21 at application, with valid sex, race codes, & AFQT scores; if tested on ASVAB 5/6/7 or a later form; if no education beyond college at application	4,186	589	539	491	542	699	766	560
3. Drop observations with negative waiting times or greater than two years waiting time between application and entry	4,160							
4. Drop duplicate applicant records (retain most recent)	3,727	(actual number of duplicate records: 432,509)						
5. Percentage identified as entrants in match		50	50	53	52	49	45	48
6. Applicants nonwhite by race and sex		white men: 2,179 white women: 453			nonwhite men: 865 nonwhite women: 231			

Notes: Data are from applicant record tapes for FY76–FY82; calendar year data for 1982 are for FY82 only (through September 30, 1982). Observations with a calendar year inconsistent with the fiscal year were dropped at row 1. Match procedures are described in text.

who were tested on ASVAB 5/6/7 or later forms, and dropped those without valid sex, race, and AFQT variables. Applicants with more than a college degree at entry were also discarded. The resulting record counts are in row 2 of Table 4.

For each applicant's file for a fiscal year, I searched for a record with the same SSN in the entrants file for that same year, as well as the entrants files for two fiscal years ahead. Applicants for whom an entrants record was found, and whose record identification code showed active-duty service, were identified as veterans in the match. Multiple entrants in a three-year period were given the most recent entrants record, and duplicate applicant records were discarded. The final applicants/entrants data set had approximately 3.7 million observations.

5. THE EFFECT OF ADMISSIONS CRITERIA ON MINORITY REPRESENTATION: ESTIMATION

5.1 Graphical Evidence

In the remainder of this paper, I focus on male applicants without prior military service, who applied to enter the military in fiscal years 1976–1982.[9] Figures 1 and 2 plot data from the 1976–1982 applicants/entrants match showing the probability of enlistment by year and correctly normed AFQT score. These figures provide dramatic evidence of the effects of the ASVAB misnorming. For whites and nonwhites separately, the figures plot eight lines for the time series of enlistment probabilities in the eight AFQT mental groups, minus the probability of enlistment for that mental group in 1982. For example, the first point plotted on the line labeled 4b in Figure 1 indicates that applicants whose correctly normed AFQT score put them in Mental Group IVb were roughly 25 percentage points more likely to enlist in 1976 than in 1982. In each of the lowest mental groups (IVa, IVb, IVc, and V) there was a substantially higher probability of enlisting in 1976–1979

Figure 1 here

Figure 2 here

than in 1980–1982. By 1981 there was almost no difference in the probability of enlisting by mental group (relative to the base year, 1982.)

Figures 3 and 4 further emphasize the effects of misnorming by subtracting the line for AFQT Group I applicants from the lines for all other groups in order to control for common-year effects. For example, Figure 3 indicates that white applicants whose correctly normed AFQT score put them in Mental Group IVb were almost 30 percentage points more likely to enlist in 1977 and 1978 than in 1982, after controlling for year effects common to all AFQT groups. Figures 3 and 4 show that even Mental Group 3b was affected by the ASVAB misnorming.

5.2 Measurement Framework

The return to traditional test score standards after correction of the ASVAB misnorming in 1980 resulted in the disqualification of a disproportionately large number of black applicants. Eitelberg (1988, Table 46) estimates that 25.6 percent of all the men who enlisted between 1976 and 1980 and 41 percent of the black men who enlisted in this period would not have qualified under traditional standards. In this subsection, I focus on the adverse impact of higher standards from the point of view of the applicant. Using the applicants/entrants match, I am able to decompose the total effect of higher standards into the direct effect and a behavioral component operating through the conditional probability $\Pr[E \mid M, A, Q]$.

The statistical analysis is based on a 30 percent random sample of the population described in row 2 of Table 4.[10] There are 699,947 white applicants and 286,918 nonwhite applicants in the sample. Table 5 provides descriptive statistics for the sample. Columns 1 and 4 show the proportion of white and nonwhite applicants meeting basic enlistment standards. For the purposes of the table, an applicant was determined to have satisfied the basic standard if he had a high school diploma and an AFQT score of 16 or more, or a GED or no high school diploma but an AFQT score of 31 or more. This is roughly equivalent to the minimum standard for entrance into the Army prevailing

Figure 3 here

Figure 4 here

TABLE 5

Qualification Probabilities: Proportion Meeting Basic Test Score and Schooling Standards

	Whites			Nonwhites		
	Mean Proportion	Interaction Terms		Mean Proportion	Interaction Terms	
Year	Qualifying (1)	AFQT Group IV (2)	AFQT Group V (3)	Qualifying (4)	AFQT Group IV (5)	AFQT Group V (6)
1976	.84	.276	.018	.57	.154	.011
		(.003)	(.006)		(.008)	(.010)
1977	.87	.374	.016	.60	.207	.010
		(.003)	(.007)		(.008)	(.010)
1978	.87	.368	.009	.61	.187	.008
		(.003)	(.007)		(.008)	(.011)
1979	.87	.374	.002	.61	.188	.003
		(.003)	(.007)		(.008)	(.011)
1980	.86	.295	.0002	.59	.145	.0002
		(.003)	(.006)		(.007)	(.010)
1981	.76	-.090	-.000	.48	-.121	.000
		(.003)	(.006)		(.007)	(.010)
1982	.84	reference year		.64	reference year	
Sample		699,947			286,918	

Notes: Columns 1 and 4 show the proportion of applicants meeting the following criteria: high school graduates with AFQT scores of 16 or greater, or GED and non-high-school graduates with AFQT scores of 31 or greater. For the purposes of this calculation, AFQT scores are determined on the basis of score scales in force at the time of application (incorrect 1944 norms until October 1980, correct 1944 norms thereafter). The remaining columns report the interaction terms in a regression of the qualification dummy on year effects, main effects for AFQT Groups IIIB, IV, and V, and year-by-AFQT-group interaction effects for these three groups. Standard errors are in parentheses. Sample is 30 percent of row 2 in Table 4, minus duplicate observations and observations on veterans with negative waiting times or waiting times greater than two years between application and service.

throughout the sample period. I used the AFQT scaled scores in force at the time of application (based on incorrect 1944 norms until October 1980, and correct 1944 norms thereafter)[11] to assign the qualification dummy. White applicants were substantially more likely to meet minimum enlistment standards than nonwhites, and for both groups there was a substantial drop in qualification proportions in 1981 after the ASVAB misnorming had been corrected.

Columns 2, 3, 5, and 6 report the interaction terms in a regression of the qualification dummy on year effects, main effects for AFQT Groups IIIB, IV, and V, and year-by-AFQT interaction terms for these three groups. The AFQT dummies on the right-hand side of the estimating equation indicate "true" AFQT scores calculated on the basis of a correctly normed (1980) score scale. The estimating equation is:

$$(5) \qquad Q_{it} = \alpha + \sum_t d_{it}\beta_t + \sum_k a_{ik}\gamma_k + \sum_t \sum_k (d_{it} \cdot a_{ik})\, \delta_{tk} + \varepsilon_{it}$$

$$t = 76, 77, 78, 79, 80, 81$$

$$k = \text{AFQT Group IIIb, AFQT Group IV, AFQT Group V,}$$

where Q_{it} is a dummy variable that indicates whether applicant i met basic schooling and test score standards at the time of application; d_{it} is a dummy variable that equals 1 if applicant i applied in year t; and a_{ik} is a dummy variable that equals 1 if i has a correctly scaled AFQT score in group k. The regression coefficients β_t and γ_k are period and AFQT-group main effects. The interaction terms, δ_{tk}, show how much more likely low-ability recruits with AFQT scores in group k were to be classified as meeting basic standards during the misnorming period than they were subsequently. The interaction terms in this equation were computed in the same manner as the interaction terms plotted in Figures 3 and 4.

The estimates in Column 2 of Table 5 indicate that a white applicant whose correctly scaled score would have placed him in AFQT Group IV was 37 percentage points more likely to have been classified as meeting basic standards in 1977 than in 1982. The corresponding figure for nonwhite applicants is 21 percentage points.

The reason the impact of misnorming on qualifications is lower for nonwhites is that nonwhite applicants were more likely than white applicants to be high school graduates.

Columns 1 and 4 of Table 6 show the percentage of applicants successfully enlisting. Here the percentages for whites and nonwhites are much closer than those for qualification, although white applicants were consistently more likely to enlist than nonwhite applicants. The probability of enlistment dropped sharply between 1980 and 1981, again reflecting the correction of test-score misnorming. The interaction terms in columns 2, 3, 5, and 6 correspond to those reported for qualifications in Table 5. These are interaction terms from a version of equation (5) where the dependent variable is enlistment. A white applicant whose true AFQT score would put him in AFQT Group IV was 22 percentage points more likely to enter the military in 1977 than in 1982. The corresponding figure for nonwhites is 20 percent.

Overall, the estimates in Tables 5 and 6 suggest that, conditional on correctly scaled AFQT test scores, test-score misnorming led to similar increases in white and nonwhite enlistments. Yet the impact of misnorming on qualification was much stronger for whites than for nonwhites. To explain this finding, note that the relationship between qualifications and enlistment can be decomposed using the following identity:

$$(6)\ \Pr[E_{it}=1 \mid b_i, A_{it}=1, Q_{it}] = \Pr[E_{it}=1 \mid b_i, A_{it}=1, Q_{it}=0] + Q_{it}\{\Pr[E_{it}=1 \mid b_i, A_{it}=1, Q_{it}=1] - \Pr[E_{it}=1 \mid b_i, A_{it}=1, Q_{it}=0]\},$$

TABLE 6

Enlistment Probabilities: Proportion Successfully Enlisting

	Whites			Nonwhites		
	Mean Proportion	Interaction Terms		Mean Proportion	Interaction Terms	
Year	Enlisting (1)	AFQT Group IV (2)	AFQT Group V (3)	Enlisting (4)	AFQT Group IV (5)	AFQT Group V (6)
1976	.53	.194	.112	.43	.161	.123
		(.005)	(.013)		(.010)	(.013)
1977	.51	.219	.128	.45	.196	.141
		(.005)	(.013)		(.010)	(.013)
1978	.54	.222	.138	.49	.193	.110
		(.006)	(.014)		(.010)	(.013)
1979	.54	.213	.066	.45	.174	.062
		(.005)	(.014)		(.010)	(.013)
1980	.53	.131	.007	.40	.101	.013
		(.005)	(.013)		(.009)	(.013)
1981	.48	-.012	-.004	.35	-.025	-.001
		(.005)	(.013)		(.009)	(.013)
1982	.52	reference year		.43	reference year	
Sample		699,947			286,918	

Notes: Columns 1 and 4 show the proportion of applicants enlisting to active duty. The remaining columns report the interaction terms in a regression of an enlistment dummy on year effects, main effects for AFQT Groups IIIB, IV, and V, and year-by-AFQT-group interaction effects for these three groups. Standard errors are in parentheses. Sample is 30 percent of row 2 in Table 4, minus duplicate observations and observations on veterans with negative waiting times or waiting times greater than two years between application and service.

where b_i is a variable representing applicant i's correctly scaled AFQT score. For the purposes of the analysis that follows, $\Pr[E_{it}=1 \mid b_i, A_{it}=1, Q_{it}=0]$ is not always zero because the qualification variable, $Q_{it}=0$, is only an approximate indicator of applicant qualifications.[12]

To identify the effect of changes in qualifications on the probability of enlistment, I combine (6) with a linear model:

$$\Pr[E_{it}=1 \mid b_i, A_{it}=1, Q_{it}=0] = \alpha_b + \beta_t + \varepsilon_{it}$$

$$\Pr[E_{it}=1 \mid b_i, A_{it}=1, Q_{it}=1] - \Pr[E_{it}=1 \mid b_i, A_{it}=1, Q_{it}=0] = \phi_t,$$

where α_b is an AFQT-group effect, β_t is a period effect, ε_{it} is an idiosyncratic component for individual i at time t, and ϕ_t is the time-varying effect of qualifications. This leads to the estimating equation

$$\Pr[E_{it}=1 \mid b_i, A_{it}=1, Q_{it}] = \alpha_b + \beta_t + Q_{it}\phi_t + \varepsilon_{it}. \quad (7)$$

It is necessary to control for period effects because enlistment rates and the quality of enlistments generally follow the unemployment rate (Dale and Gilroy 1985). In principle, the parameters of equation (7) can be estimated using a linear probability model or other functional form to regress an enlistment dummy on a qualifications dummy, along with period effects and AFQT-group effects. Coefficients on interactions of Q_{it} with dummies for demographic groups may show how the effect of qualification (or disqualification) differs for different groups.

It is important to note, however, that Ordinary Least Squares (OLS) estimates of equation (7) do not necessarily lead to estimates of ϕ_t with a causal interpretation. In other words, the resulting estimates of ϕ_t may not tell us how likely an arbitrarily or exogenously disqualified applicant would have been to enlist. This is because the behavior of qualified applicants may differ systematically from the behavior of those who are not qualified, and because the military uses information on enlistments when setting qualification standards.[13] One way to overcome this problem is to assume that although ε_{it} may be correlated with Q_{it}, it is not correlated

with b_i in the population where A_{it}=1. Then, taking expectations of both sides of (7) conditional on b_i and A_{it}=1 alone, we have

$$(8) \qquad \Pr[E_{it}=1 \mid b_i, A_{it}=1] = \alpha_b + \beta_t + \Pr[Q_{it}=1 \mid b_i, A_{it}=1]\phi_t.$$

The first step in estimating equation (8) is to estimate $\Pr[Q_{it}=1 \mid b_i, A_{it}=1]$. Then causal estimates of ϕ_t can be computed from a regression of the estimated probability of enlistment by year and true AFQT group on AFQT-group effects, year effects, and the estimated probability of meeting basic qualification standards by year and true AFQT group.

This estimation strategy can be interpreted as a decomposition of the probability of enlistment conditional on application into a component which represents the direct effect of qualifications, $\Pr[Q_{it}=1 \mid b_i, A_{it}=1]$, and a behavioral component, ϕ_t, which indicates how likely qualified applicants are to follow through with enlistment. One simple way to estimate ϕ_t is to divide the interaction terms in Table 5 (showing the relationship between true AFQT scores and qualifications each year) by the interaction terms in Table 6 (showing the relationship between true AFQT scores and basic enlistments in each year). More generally, estimates of ϕ_t can be computed using Two-Stage Least Squares (2SLS). 2SLS estimates are computed by regressing Q_{it} on a full set of AFQT-group and year effects *and interaction terms*, and then regressing E_{it} on the fitted values from this "first-stage" equation using equation (8) in the second stage. In this case, ϕ_t is identified by the exclusion of interaction terms from the second stage.[14]

5.3 Estimation Results

Table 7 reports OLS and 2SLS estimates of equation (7) for two different measures of applicant qualifications (2SLS estimates of equation (7) are the same as weighted least squares estimates of equation (8) when $\Pr[Q_{it}=1 \mid b_i, A_{it}=1]$ is replaced by regression estimates [Theil 1971].) The first measure is the basic qualifications dummy used in Table 5. The second is an attempt to more precisely code the exact age, test score,

and schooling standards used by the Army in each year from 1976–1982. Information used to compute the Army qualifications dummy is from Eitelberg et al. (1984, Appendix A).

Columns 1 and 2 of Table 7 report OLS estimates of the coefficient ϕ_t in equation (8), where Q_{it} is the basic qualifications dummy. The first row of the table shows a pooled estimate for all years. The interpretation of these estimates is that disqualifying a white applicant reduces his probability of enlistment by .315, while disqualifying a nonwhite applicant reduces his probability of enlistment by .320.

The OLS estimates summarize the correlation between qualifications and enlistment conditional on year effects and AFQT-group effects. But since qualification is related to personal characteristics (like age and schooling) that are associated with the propensity to enlist for other reasons, this correlation probably does not have a causal interpretation. The 2SLS estimates could differ from the OLS estimates because the 2SLS estimates are constructed by discarding all variation in qualifications other than that associated with changes in the way AFQT scores were scaled over time. Thus, the 2SLS estimates can be thought of as the results from a natural experiment designed to estimate the effect of higher test-score cutoffs.

2SLS estimates, reported in columns 3 and 4, show a much larger impact of disqualification on enlistment, as well as a substantial difference in the impact of disqualification by race. The estimates in columns 3 and 4 indicate that, on average across all years in the sample, disqualifying a

TABLE 7

OLS and 2SLS Estimates of the Effect of Basic Standards on Enlistment

	Regressor							
	Basic Qualifications				Army Qualifications			
	OLS		2SLS		OLS		2SLS	
Year	Whites (1)	Nonwhites (2)	Whites (3)	Nonwhites (4)	Whites (5)	Nonwhites (6)	Whites (7)	Nonwhites (8)
1976–1982	.315	.320	.477	.575	.263	.293	.448	.559
	(.002)	(.002)	(.007)	(.015)	(.002)	(.002)	(.007)	(.015)
1976	.279	.282	.325	.439	.257	.266	.251	.407
	(.004)	(.005)	(.013)	(.018)	(.004)	(.005)	(.013)	(.018)
1977	.261	.274	.311	.432	.213	.245	.228	.407
	(.005)	(.005)	(.014)	(.018)	(.005)	(.005)	(.014)	(.018)
1978	.232	.268	.298	.455	.106	.196	.170	.452
	(.005)	(.005)	(.014)	(.018)	(.004)	(.005)	(.012)	(.020)
1979	.299	.335	.367	.500	.226	.301	.275	.480
	(.005)	(.005)	(.014)	(.018)	(.005)	(.005)	(.014)	(.019)
1980	.365	.384	.442	.542	.348	.371	.368	.512
	(.004)	(.005)	(.013)	(.018)	(.004)	(.005)	(.013)	(.018)
1981	.351	.345	.423	.549	.337	.333	.380	.526
	(.003)	(.004)	(.008)	(.015)	(.003)	(.004)	(.008)	(.016)
1982	.334	.320	.441	.585	.321	.321	.397	.562
	(.004)	(.005)	(.009)	(.017)	(.004)	(.005)	(.010)	(.017)

Notes: The table reports coefficients from a regression of an enlistment dummy on a qualification dummy. Other covariates include a full set of year effects and main effects for AFQT Groups IIIB, IV, and V. The 2SLS estimates are identified by the exclusion of year-by-AFQT-group interaction terms, which were also included in the first-stage equation used to compute fitted values of the qualification dummy. Standard errors are in parentheses. Sample is 30 percent of row 2 in Table 4, minus duplicate observations and observations on veterans with negative waiting times or waiting times greater than two years between application and service.

white applicant reduces his probability of enlistment by .477, whereas disqualifying one nonwhite applicant reduces his probability of enlistment by .575. Put differently, the reduction in nonwhite enlistments caused by disqualifying one thousand nonwhite applicants would be 20 percent larger (575 - 477 = 98) than the reduction in white enlistments caused by disqualifying one thousand white applicants. 2SLS estimates using the Army qualifications dummy show an average 25 percent differential impact by race. The fact that minority applicants are typically less likely to be qualified serves to accentuate this differential impact.

In each column, the time-varying 2SLS estimates of the effect of disqualification for whites and nonwhites indicate that the proportion of qualified applicants continuing on to enlist increased over time. This finding is consistent with the theoretical implication, discussed in Section 3, that as standards are raised the remaining pool of qualified applicants is more likely to enlist. The increase in enlistment probabilities is slightly larger for nonwhites.

5.4 The Behavioral Impact

The 1982 data for applicants illustrate the importance of the behavioral component of enlistment for the Armed Forces racial mix. In the 1982 applicant population analyzed here, roughly 268.4 thousand white male applicants satisfied basic test score and schooling standards, and roughly 73.9 thousand nonwhite male applicants satisfied these standards. A simple calculation suggests that an increase in basic standards achieved by disqualifying all applicants with AFQT percentile scores in Category IV (below the 30th percentile) would have left roughly 248.1 thousand white applicants qualified and 50.3 thousand nonwhite applicants qualified. The increase in standards therefore would have disqualified 20.3 thousand white applicants and 23.6 thousand nonwhite applicants.

If the proportion of qualified applicants enlisting was equal to the average proportion for both groups combined (roughly 50 percent), then 16 percent more nonwhite than white enlistments would be lost by the elimination of AFQT-IV's from the applicant pool. In numbers, 10.2 thousand white enlistments would be lost and 11.8 thousand nonwhite enlistments would be lost, a difference of 1.6 thousand. But the estimates in Table 7 suggest that the proportion of qualified applicants enlisting differs substantially by race. The impact of

disqualification is largest in 1982; estimates for 1982 imply that disqualification of one thousand white applicants eliminates 441 white enlistments, whereas disqualification of one thousand nonwhite applicants eliminates 585 nonwhite enlistments. Applying these proportions in the example above leads to an estimated loss of 8.9 thousand white enlistments and 13.8 thousand nonwhite enlistments. In this case, roughly 55 percent more nonwhite enlistments than white enlistments are lost by the elimination of AFQT-IV's from the applicant pool. This calculation suggests that the behavioral component of enlistment could more than triple the difference between the number of white and nonwhite enlistments lost when entrance standards are raised.

6. CONCLUSIONS

The most important change in entrance standards used to screen applicants for the All-Volunteer Forces was the unintentional ASVAB misnorming. In this paper, I used linked data on military applicants and entrants to show that the increase in entrance standards resulting from correction of the ASVAB misnorming had a larger impact on minority enlistments than on white enlistments for at least two reasons. First, minority applicants were less likely to satisfy military entrance criteria than white applicants. Second, a disqualified minority applicant was otherwise more likely to enlist than was a disqualified white applicant.

For both racial groups, increases in standards led to higher rates of enlistment among qualified applicants. This increase was slightly larger for nonwhite applicants. Overall, however, estimates suggest that since the ASVAB misnorming episode, which resulted in higher test-score cutoffs, minority enlistments have decreased an average of 20 to 25 percent more than have white enlistments. This difference has occurred independently of any discouraging impact that higher standards may have had on potential applicants.

Much of the research on job performance in the military suggests that high school graduation status is a better predictor of success on the job than are test scores. For example, a number of studies suggest that the potentially unqualified men admitted during the ASVAB misnorming episode, as well as a group of low-ability men admitted in the late 1960s and early 1970s, did not perform significantly worse than other groups in the military (e.g., Ramsberger and Means 1987). Recent studies of groups of Navy (Cooke and Quester 1992) and

Army (Fernandez 1992) recruits suggest that for some population subgroups, there is an association between test scores, job performance, and recruit retention rates. But the Navy study suggests that schooling levels are a more important predictor of military job performance.

Although minority applicants to the military have poorer test scores than white applicants, minority applicants are slightly more likely than white applicants to have graduated from high school (Defense Department 1992). Overall, in fiscal year 1991, 97 percent of new recruits had a high school diploma, and most of the remainder had a GED or other alternative credential (Defense Department 1992). This high proportion, however, partly reflects recent increases in test-score cutoffs that could be relaxed in the future. For example, in 1987, when test-score cutoffs were lower, 93 percent of applicants were high school graduates. Increases in standards based on school graduation and school performance might therefore be a practical alternative that could lead to better-performing recruits without disqualifying a disproportionately large number of minority applicants.

Appendix: Equipercentile Equating and the AFQT

Equipercentile equating begins by first establishing a reference population and a reference test, against which all test scores are to be measured. The reference population for AFQT test forms until 1973 (through AFQT Form 8) is known as the World War II Mobilization Population. This is the population of all men on active duty in December 1944, most of whom took the AGCT. In equating AFQT score scales, the AGCT was used as a reference test. The score scale for each new AFQT test was determined by administering both the AGCT reference test and the new test to a sample assumed to have the same underlying ability distribution as the reference population. When equating tests, an effort is also usually made to administer the new test under the same conditions as those under which the reference population was originally tested.

The score scale for each new AFQT was determined by equating percentiles for both the new test form and the reference test in a sample of test takers. Formally, let X denote the raw score on the reference test, and let Y denote the raw score on the new test; let F and G denote the cumulative distribution functions. Then the scaled score on the new test is $Y^* = F^{-1}[G(Y)]$.

Test equating ensures that the distribution of the scaled score on the new test, Y^*, will be the same as the distribution of the score on the reference test. Scaled scores are then reported as percentiles of the reference test score distribution, and are said to be normed to the reference population. The hope is that percentile scores on the new test, having been normed to the reference population, indicate the same level of ability as percentile scores attained by the reference population. More sophisticated variants of the basic technique described here, such as Common Item Equating (Holland and Rubin 1982), can be used when the reference population is thought to differ importantly from the new population. Techniques are available to validate equating procedures.

Notes

[1]Statistics from U.S. Congress (1989, p. 5). AFQT mental groups are assigned to percentile scores as follows: Group V = percentile score 1–9, Group IV = percentile score 10–30, Group III = percentile score 31–64, Group II = percentile score 65–92, Group I = percentile score 93 and above.

[2]Effective October 1, 1981, the fraction of AFQT Category IV enlistments was limited to 80 percent for any fiscal year. This constraint was never binding. But no person with an AFQT score below 31 (Category IV or V) who was not a high school graduate could be accepted for enlistment (Defense Department 1988, p. II-23) after October 1981. Category V enlistments were outlawed in 1951.

[3]Experimental evidence on the importance of veterans benefits in stimulating enlistment is described by Fernandez (1982). Angrist (1993) shows that veterans benefits probably help to raise the level of postservice schooling and earnings for benefit users.

[4]Estimates of the proportion of prospective applicants who do not progress beyond exploratory discussions range from 10 to 20 percent. Those who fail to progress include those who fail the preliminary screening test or who lack the education credentials or who are involved with the legal authorities; those who seek information but decide not to follow up; and unemployed persons who seek immediate employment with no understanding of military requirements (Defense Department 1992).

[5]Enlistment is a multistage process typically involving several visits to different sites (Berryman, Bell, and Lisowski 1983). Applicants talk with recruiters, take a written test (perhaps after a preliminary shorter test given by recruiters) and undergo a medical examination, and can decide to enter the Delayed Entry Program and postpone enlistment for a year or more. The data analyzed below show that roughly two-thirds of the applicants who applied between 1976 and 1982 and eventually went on active duty did so in the same year as the year of application. Thirty percent went

on active duty in the year following the year of application, and 4 to 5 percent went on active duty in the second year after the year of application.

[6]For example, the upper part of the score scale was normed against Air Force and Navy recruits using ASVAB 2 as the reference test. The low end of the score scale was normed against Army applicants who had taken ACB-73. Some of the reference tests used by the individual services had also been incorrectly scored, and some reference score scales had been inflated by coaching.

[7]Also in 1980, a new set of norms was established when a nationally representative sample of high school students was given the ASVAB 8A test as part of the Profiles of American Youth Study undertaken in cooperation with the Labor Department. Score scales derived from the 1980 norms became standard in October 1984. A new set of test forms, ASVAB 9/10/11, was also introduced at this time. Since 1984, AFQT percentiles for the purpose of historical comparisons have been reported using 1980 norms. In practice, however, AFQT scores derived from 1980 norms differ little from scores derived from (corrected) WWII norms.

[8]Aspects of the applicant record-keeping system are described in the appendix to Berryman, Bell, and Lisowski (1983).

[9]Roughly 11 to 15 percent of new recruits were female in the period 1979–1991 (Defense Department 1992).

[10]After drawing this sample, I eliminated duplicate records using the procedure described above for the population, and discarded records on entrants with more than two years between entry and application or an entry date preceding the year of application.

[11]For men applying after October 1980, I converted the 1980 score scale provided on the USRDD tapes to a corrected 1944 score scale with the aid of conversion tables provided by Les Willis at the DMDC in Monterey, California.

[12]Some applicants with very low test scores are nevertheless accepted to the military. Entry standards also vary across the services. For example, the Marines have typically employed a lower test-score cutoff than the Army.

[13]The notion that qualifications are endogenous is supported by Eitelberg et al.'s (1984, p. 48) observation that ". . . aptitude and education Service screens are highly reactive to changes that occur in the recruiting environment."

[14]Angrist (1993) provides a detailed discussion of the use of interaction terms as instruments in 2SLS estimation.

References

Angrist, J. D. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." American Economic Review, 80 (June): 313–335.

_____. 1991. "The Draft Lottery and Voluntary Enlistment in the Vietnam Era." Journal of the American Statistical Association, 86(415): 584–595.

_____. 1993. "The Effect of Veterans Benefits on Education and Earnings." Industrial and Labor Relations Review, 46(4): 637–652.

ASVAB Working Group. 1980. History of the Armed Services Vocational Aptitude Battery, 1974-1980. Washington, D.C.: Office of the Assistant Secretary of Defense (Manpower, Reserves, and Logistics), March.

Berger, M. C., and B. T. Hirsch. 1983. "The Civilian Earnings Experience of Vietnam-Era Veterans." Journal of Human Resources, 18 (Fall): 455–479.

Berryman, S. E, R. M. Bell, and W. Lisowski. 1983. The Military Enlistment Process: What Happens and Can It Be Improved? Report R-2986-MRAL, RAND Corporation, Santa Monica, Calif., May.

Boesel, D. 1992. "Cutting Recruits: A Profile of the Newly Unqualified." Chapter 1 in Military Cutbacks and the Expanding Role of Education, Office of Educational Research and Improvement, U.S. Department of Education, Washington, D.C., October.

Cooke, T. C., and A. O. Quester. 1992. "What Characterizes Successful Enlistees in the All-Volunteer Force: A Study of Male Recruits in the U.S. Navy." Social Science Quarterly, 73(2): 238–252.

Cooper, R. V. L. 1977. Military Manpower and the All-Volunteer Force. Report R-1450-ARPA, RAND Corporation, Santa Monica, Calif., September.

Cutright, Phillips. 1974. "The Civilian Earnings of White and Black Draftees and Nonveterans." American Sociological Review, 39 (June): 317–327.

Dale, C., and C. Gilroy. 1985. "Enlistments in the All-Volunteer Force." American Economic Review, 75(3): 547–551.

Defense Department. 1988. Population Representation in the Military Services: Fiscal Year 1987. Office of the Assistant Secretary of Defense (Force Management and Personnel), Washington, D.C., August.

_____. 1992. Population Representation in the Military Services: Fiscal Year 1991. Office of the Assistant Secretary of Defense (Force Management and Personnel), Washington, D.C., October.

Eitelberg, Mark J. 1988. Manpower for Military Occupations. Human Resources Research Organization, Alexandria, Va., April.

Eitelberg, Mark J., J. Lawrence, B. Waters, and L. Perelman. 1984. Screening for Service: Aptitude and Education Criteria for Military Entry. Human Resources Research Organization, Alexandria, Va., September.

Fernandez, J. C. 1992. "Soldier Quality and Job Performance in Team Tasks." Social Science Quarterly, 73(2): 253–265.

Fernandez, R. L. 1982. Enlistment Effects and Policy Implications of the Educational Assistance Test Program. Report R-2935-MRAL, RAND Corporation, Santa Monica, Calif., September.

Grissmer, D. W. 1992. "Impact of the Military Drawdown on Youth Employment, Training, and Educational Opportunity." Chapter 3 in Military Cutbacks and the Expanding Role of Education, Office of Educational Research and Improvement, U.S. Department of Education, Washington, D.C., October.

Holland, P., and D. Rubin. 1982. Test Equating. New York: Academic Press.

Karpinos, Bernard D. 1967. "Mental Test Failures." Chapter 4 in The Draft: A Handbook of Facts and Alternatives, edited by Sol Tax, University of Chicago Press, Chicago.

Laurence, J. H. 1992. "Crew Cuts: Effects of the Defense Drawdown on Minorities." Chapter 2 in Military Cutbacks and the Expanding Role of Education, Office of Educational Research and Improvement, U.S. Department of Education, Washington, D.C., October.

Maier, M. H., and W. H. Sims. 1986. The ASVAB Score Scales: 1980 and World War II. Report No. 116, Center for Naval Analyses, Alexandria, Va.

Maier, M. H., and A. R. Truss. 1983. Original Scaling of ASVAB Forms 5/6/7: What Went Wrong? Research Contribution No. 457, Center for Naval Analyses, Alexandria, Va.

Orvis, B. R, and M. T. Gahart. 1990. Enlistment among Applicants for Military Service. Report R-3359-FMP, RAND Corporation, Santa Monica, Calif., January.

Ramsberger, P., and B. Means. 1987. Military Performance of Low-Aptitude Recruits: A Reexamination of Data from Project 100,000 and the ASVAB Misnorming Period. FR-PRD-87-31, Human Resources Research Organization, Alexandria, Va., September.

Theil, H. 1971. Principles of Econometrics. New York: John Wiley and Sons.

U.S. Congress. 1989. "Social Representation in the U.S. Military." Congressional Budget Office, Washington, D.C., U.S. Government Printing Office, October.