Research Report

Strength Testing in the Air Force

Current Processes and Suggestions for Improvements

Carra S. Sims, Chaitra M. Hardison, Maria C. Lytell, Abby Robyn,
Eunice C. Wong, Erin N. Gerbec

RAND CORPORATION

For more information on this publication, visit www.rand.org/t/RR471

Library of Congress Cataloging-in-Publication Data

Sims, Carra S.
Strength testing in the Air Force : current processes and suggestions for improvements / Carra S. Sims, Chaitra M. Hardison, Maria C. Lytell, Abby Robyn, Eunice C. Wong, Erin N. Gerbec.
pages cm
Includes bibliographical references.
ISBN 978-0-8330-8435-4 (pbk. : alk. paper)
1. United States. Air Force—Physical training. 2. Physical fitness—Measurement.
3. Physical fitness—Evaluation. 4. United States. Air Force—Personnel management.
I. Rand Corporation. II. Title.

UG638.S56 2014
358.4'15—dc23

2014040609

Published by the RAND Corporation, Santa Monica, Calif.

Preface

Since 1987, the Strength Aptitude Test, a test of physical strength, has been used by the Air Force to screen and classify enlisted personnel to their career specialties. The decision to institute the test was the culmination of several years of research on physical skills testing. However, more than 20 years later, the Strength Aptitude Test as a screening and classification tool in the Air Force has yet to be reevaluated. RAND was therefore asked to evaluate the usefulness, validity, and fairness of the Strength Aptitude Test for classifying enlisted airmen to their career specialties. This report provides the results of our evaluation.

The research reported here was commissioned by the Air Force Directorate of Force Management Policy (AF/A1P) and conducted within the Manpower, Personnel, and Training Program of RAND Project AIR FORCE as part of a project from fiscal years 2010 to 2011. This report should be of interest to those involved or interested in Air Force policy, procedures, and practices for classifying enlisted personnel to job specialties.

RAND Project AIR FORCE

RAND Project AIR FORCE (PAF), a division of the RAND Corporation, is the U.S. Air Force's federally funded research and development center for studies and analyses. PAF provides the Air Force with independent analyses of policy alternatives affecting the development, employment, combat readiness, and support of current and future air, space, and cyber forces. Research is performed in four programs: Force Modernization and Employment; Manpower, Personnel, and Training; Resource Management; and Strategy and Doctrine. The research reported here was prepared under contract FA7014-06-C-0001.

Additional information about PAF is available on our website:
http://www.rand.org/paf/

Contents

Figures

Tables

Summary

Since 1987, the Strength Aptitude Test (SAT), a test of physical strength, has been used by the Air Force to screen and classify enlisted personnel to their career specialties. The decision to institute the test was the culmination of several years of research on physical skills testing. However, over the past 20 years, the Air Force has not reevaluated the test as a screening and classification tool. RAND was therefore asked to evaluate the current status of the SAT regarding its usefulness, validity, and fairness for classifying enlisted airmen. This report provides the results of our study.

Our evaluation began with an initial review of the SAT, existing evidence regarding its validity, and the current processes for developing cut scores on the SAT. Based on that initial review, we concluded that, while strength testing is needed in the Air Force, the SAT and the current processes for establishing the minimum requirements for entry into certain Air Force specialties (AFSs) may not be ideal. In particular, we identified a number of gaps in the evidence supporting current processes and determined that three research efforts would be worthwhile in helping to close the gaps. This study undertook the first two research efforts; however, as part of our conclusions we provide insight into how the third effort might be conducted. The three research efforts are as follows:

1. *More information is needed on how the SAT is actually used in practice*. To address this need, we conducted a series of in-person observations of the SAT being administered to applicants in a variety of Military Entrance Processing Stations (MEPS) across the United States. We also interviewed recruits being tested as well as the personnel at the MEPS who regularly administer the tests—the liaison non-commissioned officers or LNCOs.
2. *The process for setting cut scores should be updated*. We concluded that the manner in which information about physical job requirements is collected might be deficient because it involves only limited input. As an alternative, we explored collecting this information using an online survey.
3. *The SAT should be further validated and its validity should be compared to that of other strength and stamina measures.* The particular gap that should be filled is the link between test performance and on-the-job performance. Research on the SAT has not adequately explored this issue. Although this avenue of research was beyond the scope of this study, we describe how such a study might be conducted.

Use of the SAT at Military Entrance Processing Stations

To better understand the operational use of the SAT, we observed the test administration process at four medium- to large-sized MEPS locations, interviewed recruits taking the test, and interviewed the Air Force staff at the observation sites that screen recruits and administer the SAT. We also interviewed test administrators at four other medium- to small-sized MEPS sites. Our aim through these site visits was to investigate the condition of the incremental lift machines used in test administration; to determine if the test protocol is being consistently administered

across locations and in the way it was designed to be used; and to gain insight into recruits' reactions to the SAT. We offer several recommendations that could improve test administration and, in turn, assignment of career fields, over the long run.

Incremental Lift Machines

In general, the machines we observed were in good working order, though we did identify a few machines in need of repair or replacement. We also learned that some MEPS may have more than one machine and any extras could be used as replacements for those that are damaged. Some differences exist in terms of visible information regarding use of the machines and where machines are located at the MEPS. But in general, these differences did not appear to impact test administration in any significant way.

Recommendation: Conduct a full inventory of SAT machines on a regular basis (every few years). Identify damaged machines and replace or repair as needed.

SAT Administration

We observed a total of 34 recruits taking the SAT. Many aspects of the administrations that we observed were fairly consistent across recruits, LNCOs, and locations, and consistent with the way in which the test administration was originally conceptualized. We did, however, discover some variations in the administration that could meaningfully impact test results. The test begins with a 40-pound lift (the minimum requirement) and increases in ten-pound increments to a maximum lift of 110 pounds. As an example of one variation in administration, most of the LNCOs stop the test at 100 pounds, even though the intended administration is to continue to 110 pounds and record 110 pounds for a final score if the recruit successfully completes the lift. LNCOs frequently stop at 100 pounds because no job currently requires a higher score. As a result, a score of 100 could mean that a recruit cannot lift 110, or it could mean that the recruit tested at a location that stops at 100 pounds and never had a chance to lift to 110. This variation adds error to the information collected that could make it less useful because it restricts variance that would be useful to inform validation efforts, as well as imposing limitations on identifying whether an airman is or is not qualified for a particular job should requirements change.

Other variations concern whether recruits take the test individually or in groups and whether the encouragement they might receive in groups measurably affects their results—for some recruits it might be a positive motivator; for others a source of embarrassment. Whether such differences affect performance on the SAT needs to be investigated. LNCOs also differ as to whether they allow a "second chance" to complete a lift, especially if a recruit wanted a specific job but had not qualified for it.

Recommendation: To eliminate potential sources of error in test administration, send new instructions to all MEPS locations and develop a standardized training procedure for all LNCOs. Additionally, every few years, audit the implemented procedures, retrain LNCOs and redistribute official administration protocols to help ensure that the proper protocol is maintained over time.

Recruit Knowledge of the SAT

Another major difference in test administration was how much information the LNCOs divulged before and during the test. For example, some LNCOs tell recruits what score they need for a particular job and how the test will be administered, including the starting and incremental weights; others do the opposite and tell the recruits nothing during the test, including their final score. Many LNCOs believe that recruits learn about the test from recruiters, so there is no need to explain the test once the recruits arrive at the MEPS. This view is in stark contrast to the information gained from the recruits interviewed: Only half said they had heard about the strength test before arriving at the MEPS; 38 percent knew it was used to qualify for certain jobs; but only 11 percent knew how much weight they had to lift to qualify for a preferred Air Force job. Having prior knowledge and understanding of the SAT and having the opportunity to prepare could significantly affect test scores. It could mean the difference between qualifying for a desired job or not. Advice to practice could be particularly important for recruits who have no experience lifting weights or using weight machines.

Recommendation: Issue new guidance to recruiters requiring them to fully inform recruits about the SAT and encourage preparation. Specifically, recruiters should make sure that recruits understand the nature of the test and how it relates to career field assignments.

Strength Requirements Survey

RAND developed a web-based survey for defining strength requirements in career fields. The survey asked respondents in eight AFSs to describe aspects of the job's physical requirements that are vital for defining strength requirements. They are as follows:

- **The *types* of physical actions**—such as lifting, pushing, throwing. Different actions require different types of strength.
- **The *level* of the action**—that is, how much ***weight*** is involved and the ***duration*** of the action. The same action can have very different strength requirements depending on the weight of the object.
- **The *frequency* and *importance* of the actions**. Actions that occur rarely or that are of little importance are less essential in defining minimum strength requirements. In contrast, those activities that occur frequently or are vital to successful performance are central to defining minimum requirements.

The first step in establishing cut scores on any test involves clearly defining the requirements of the job. In the case of establishing requirements for strength testing, it is critical to have a solid understanding of the type of physically demanding tasks that are required for the job, as well as their importance on the job, the frequency with which they occur, and for how long that physical activity is sustained. The survey we developed was designed specifically to address these key aspects of AFS-specific job demands.

Our assessment of the survey results was conducted, in part, to determine whether a survey such as this would be a viable alternative to the current method used by the Air Force for collecting information about physical job demands. And we determined that it was. For example, overall, the average ratings of frequency, importance, duration, and performance without

mechanical assistance calculated from the survey responses revealed some differences by AFS. Respondents in the specialties with higher minimum strength scores reported more frequent requirements for particular physical activities than respondents in other specialties—such as more requirements to push objects without assistance, or more-frequent requirements to rotate, push, and carry. Analysis of movement type not only identified most frequent movement types—waist-level, chest-level, and on-the side movements appeared consistently across all AFSs—but also revealed interesting patterns that differentiated among specialties. The representation of specific results, described in this report, further illustrate the validity of the survey tool.

Our assessment of the survey results led to recommendations in two areas: (1) the methodology for converting job demands into career field strength requirements, and (2) identifying career field physical demands.

Methodology for Setting Strength Scores

Our review of the current methodology for converting job demands information into SAT strength scores revealed that many elements of the program are unsupported, and other key elements that should be considered in the method are absent (including duration and importance of various tasks). As a result, we believe the process should be changed to consider a broader range of factors that more accurately reflect physical demands and, in doing so, accurately document the elements of the methodology to provide a basis for continued evaluation.

Recommendation: Establish a new method for converting job demands information into SAT cut scores. In developing a new process, explore the following factors:

- Use well-established approaches for setting standards.
- Compensate for gains expected from basic training.
- Consider task importance and duration, in addition to frequency and percentage of people performing the task.
- Consider a wider variety of physical demands, such as those that may emphasize stamina in addition to those that require strength.
- Use score crosswalks instead of regression equations for converting information about the force associated with one action to another.

Career Field Physical Demands

The Air Force does not collect data on the physical demands of the job in the processes currently used to collect data on occupational tasks within specialties. Our findings agree with those of the Government Accountability Office in 1996. The results of our survey suggest that the Air Force could add survey items similar to those in the Strength Requirements Survey to address this deficiency.

Recommendation: Add items addressing physical demands to the Air Force's occupational analysis survey. We recommend adopting the Strength Requirements Survey for this purpose. Prior to use, implement the following improvements to the survey tool:

- *Increase the screening tool threshold* so it is higher for branching respondents to more detailed questions about physical activity, thereby providing more differentiation between specialties with low versus high physical demands.

- *Add questions about other types of physical job demands,* such as muscular or cardiovascular endurance, to gain a more complete picture of physical requirements.
- *Consider tailoring the survey to specific tasks*, mapping the physical demand questions to the comprehensive list of career field tasks already used in the Air Force occupational analysis survey.
- *Compare survey responses to other evaluations of job demands*, such as interviews with career field managers, focus groups, interviews with job incumbents, or in-person site visits. This process is consistent with work already done in preparing for and interpreting results of the occupational analysis surveys already administered, and it could be important in accurately setting the minimum strength scores for some Air Force specialties.

In analyzing results, gender and skill-level differences in survey responses should be compared when measuring job demands. If gender differences are identified on the survey, further examination of why perceptions of the job requirements might differ by gender should be explored before setting the minimum cut point for each specialty. Skill-level differences should similarly be explored to determine if the most physically demanding work is undertaken by a subset of skill levels, and then evaluate how those differences should be considered in setting minimum strength requirements for the career field. In addition to analyzing gender and skill-level differences, analysis of all data must be conducted within the context of the entire career field to obtain an accurate picture of how commonplace particular physical requirements are and, in turn, how relevant they are in defining strength scores.

Most important, perhaps, is the fact that survey tools, such as those described here, must be continually refined to ensure that they adequately capture specialty-specific physical demands—as these requirements can evolve with changes in technology or work processes. Thus the Air Force, if adopting this or a similar data-collection tool, will need to conduct periodic checks, such as meetings with career field managers and AFS subject matter experts, to ensure that results are accurate.

Examining the SAT in Comparison to Other Tools

The link between test performance and on-the-job performance is critical for determining the overall effectiveness of a test. However, research on the SAT has not adequately explored this issue.

Recommendation: Begin collecting data on the SAT and other alternative tools before and after basic training for use in future validation studies. First, collect data on the SAT and other measures both prior to and after basic training; then wait for several months and collect data on subsequent performance outcomes. Examine whether there are differences across AFSs and by gender in which test is the best predictor of performance.

The results of such a study would lay the groundwork for determining which tests are most predictive, which tests show the least amount of predictive bias against key subgroups (i.e., race and gender), and whether one test should be used for certain AFSs and another test used for a different group of AFSs. Ultimately, such a study is necessary for determining whether the Air Force should continue to use the SAT.

Conclusion

It is clear that some Air Force career fields require high levels of strength. For those specialties, failure to screen for strength capability could have negative consequences. Personnel who are not strong enough to handle the objects involved in the job could be injured while attempting the work or could cause others to be injured around them. Moreover, if these individuals are trained for a job that they ultimately cannot perform, the Air Force risks losing that training time and effort if individuals must ultimately be reclassified. Thus, the Air Force should not abandon the idea of strength testing or eliminate the use of the SAT without finding a suitable replacement. While the SAT is in use, administration of the test should adhere to specific guidelines to ensure the fairness and effectiveness of the scores. At the same time, alternative tests should be pursued, and the existing cut scores should be reexamined to make sure that they are not set too high or too low for a given specialty. The survey developed in this study offers one such alternative and, with some modification as we describe, could be incorporated into occupational surveys already in use by the Air Force.

Acknowledgments

We have many people to thank for contributing to this multiyear project. We would like to begin by thanking our project sponsors (with their office symbols at the time of the work), Maj Gen Dunbar (A1P), Mr. John Park (A1PF), and Dr. Lisa Hughes (A1PF) for their support of the project. We also would like to thank A1P and AFPC personnel, particularly Dr. Paul DiTullio (A1PF), Mr. Ken Schwartz (AFPC/DSYX), and Mr. Johnny Weissmuller (AFPC/DSYX), who tirelessly offered guidance and assistance throughout the project.

We also are very grateful to the many individuals who gave their time for interviews. We thank those from the 711th Human Performance Wing, including Dr. Gregory Zehner, HyegJoo Choi, Dr. Joe McDaniel, Beth Rogers, Col Timothy Wells, and Dr. Joseph Pellettiere for their insights into how the cut scores are established in the Air Force and how the SAT itself was developed. We thank the various Liaison Noncommissioned Officers (LNCOs) who answered our questions about how the SAT is administered at the MEPS (their names are excluded to protect confidentiality). We also thank the career field managers who participated in our interviews and assisted in recruiting participants for our online survey including CMSgt Akers, CMSgt Bingman, CMSgt Cummins, CMSgt Healy, CMSgt Hodges, SMSgt Zidack, and SMSgt Knoblauch.

In addition, special recognition is offered to Mr. Jeff Scott (AFMA) and Mr. Paul Collier (Booz-Allen-Hamilton) for helping us overcome some of the technical obstacles involved in the online survey.

We also want to thank our RAND colleagues who made this project possible, or at least, easier to complete. Al Robbert provided guidance throughout the project. Paul Emslie and Perry Firoz offered expertise on Air Force personnel data. Judy Mele and Nancy Campbell completed many data pulls and data cleaning activities for this data-intensive project. Herb Shukiar pretested the survey to find bugs in the programming. Julie Ann Tajiri provided ongoing administrative support.

We would also like to thank our reviewers who provided their time and effort to help strengthen this paper.

Last, and by no means least, we thank the many airmen who agreed to participate in interviews, test-administration observations, and the online survey. Without their time and effort, our study would not have been possible.

Abbreviations

A1PF	Force Management Division
AERA	American Educational Research Association
AETC	Air Education and Training Command
AF/A1P	Air Force Directorate of Force Management Policy
AFE	Aircrew Flight Equipment
AFECD	Air Force Enlisted Classification Directory
AFMA	Air Force Management Agency
AFPC	Air Force Personnel Center
AFS	Air Force Specialty
AFU-AS	A-10, F-15 & U-2 Avionics Systems
AGE	aircraft ground equipment
AP-TTP	Aerospace Propulsion – Turboprop/Turboshaft
APA	American Psychological Association
ASVAB	Armed Services Vocational Aptitude Battery
CS	Cyber Surety
DMDC	Defense Manpower Data Center
EOD	Explosive Ordnance Disposal
EEOC	Equal Employment Opportunity Commission
GAO	Government Accountability Office
ILD	incremental lift device
ILM	incremental lift machine
LNCO	liaison non-commissioned officer
LRU	line replaceable units
MAFS	Manned Aircraft Maintenance – Aircraft Fuel Systems
MEPS	Military Entrance Processing Stations
NCME	National Council on Measurement in Education

NCO	non-commissioned officer
OAD	Occupational Analysis Division [of AETC]
PAF	Project AIR FORCE
SAT	Strength Aptitude Test
SF	Security Forces
SME	subject matter expert
SS	Surgical Service

1. Introduction

The Strength Aptitude Test (SAT) is used to screen and classify personnel for the strength requirements in enlisted Air Force career fields. Although the Army was initially investigating the SAT's usefulness as a formal classification tool and also used it for career counseling, it was first introduced in its current form in the late 1980s after studies suggested it would be a useful screening tool. However, relatively little research has been conducted on the Air Force's strength test in the three intervening decades. The test has stood "as is" with little reevaluation, while jobs have changed over time.

As a result of the lack of continued research, the SAT has met with some controversy. Some have called for reviews of the effectiveness of the measure and its consequences for the representation of women in certain career fields. For example, in 1996 the Government Accountability Office (GAO) questioned the use of the SAT for personnel classification and argued that SAT validation studies do not adequately account for gender differences for the following reasons:

- Male and female scores on the SAT have been grouped together for analyses, despite clear gender differences in performance on the test and different distributions of strength abilities across genders.
- The testing protocol for the SAT has not allowed individuals to perform the test to their maximum potential because individuals cannot get into comfortable lifting positions.
- Shorter individuals have been found to need more upper-body strength to perform as well on the test as taller individuals.
- Men, and especially women, improved in their performance on the SAT after just two weeks of basic training. However, airmen are not afforded the opportunity to retake the SAT after basic training.

The GAO report also criticized the speed with which the Air Force resurveys career fields to ensure their SAT cutoff scores for entry are up to date. Taking the critiques of the validation evidence and resurveying of career fields together, the GAO report concluded that the Air Force "will run the risk of denying servicemembers' entry into occupations based on invalid or outdated strength requirements," especially in those "merged occupations that have not been resurveyed" (p. 9).

The questions raised by the GAO highlight some of the potential issues with the operational use of the SAT. However, investigation into the process used to identify minimum qualifications for specific career fields has not been conducted. In addition, there is little published documentation on the SAT, making even basic inquiries into the nature of the test difficult.

The Air Force Directorate of Force Management Policy (AF/A1P) turned to RAND to provide a report documenting what is known about the SAT as well as a more detailed investigation of the SAT minimum qualification scores. What follows is a discussion of what we learned in the process of that investigation.

Approach

The research reported here is part of a multiyear effort. It started with an initial review of the SAT and the current processes used for developing cut scores on the SAT. After reviewing those processes, we identified a number of gaps in the evidence supporting the current processes. However, it is also clear from the career field documentation and interviews with subject matter experts in the Air Force that many jobs are quite physically demanding. As a result, we concluded that while strength testing is needed in the Air Force, the SAT and the current processes used for establishing the minimum requirements for entry into certain Air Force Specialties (AFSs) may not be ideal. More specifically, we determined that the following three research efforts would be worthwhile:

- More information on how the test is actually used in practice is needed.
- The process for setting cut scores should be updated.
- The SAT should be further validated, and its validity should be compared to that of other strength and stamina measures.

In the next phase of the project, we set out to collect data addressing the first two research efforts (i.e., collecting more information about SAT administration and updating the process for setting cut scores). Information regarding these research efforts serves to answer some of the GAO's stated concerns as well as to address the larger question regarding the usefulness, fairness, and validity of the SAT.

To address the need for more information about how the test is used in practice, we conducted a series of in-person observations of the SAT being administered to applicants at a variety of Military Entrance Processing Stations (MEPS) across the United States. In addition to the observations, we also conducted interviews with the people being tested and with the personnel at the MEPS who regularly administer the test.

With respect to the second suggestion, updating the method for setting cut scores, we explored a change to one aspect of the method—namely, the manner in which information about physical job requirements is collected and how that information might be applied to set cut scores in a better manner. Because we were concerned that the current method for collecting this information may be deficient (in that it only involves input from a few people at usually only three base locations), we set out to explore collecting this information using an online survey.

Online surveys are commonly used for conducting job analyses, and the Air Force itself administers an occupational analysis survey to every enlisted AFS every three years that collects job task information but does not currently collect information about the *physical* requirements for successful execution of those tasks. Therefore, we set out to explore whether we could develop questions to address those physical requirements and whether the resulting survey items would be a useful addition to the current occupational analysis survey. We administered this survey to eight AFSs that had a variety of strength requirements (as ascribed by the current cut score system) to examine its functioning and identify needed changes to the content if the Air Force decides to incorporate it into the existing occupational analysis surveys.

Lastly, although resources were not available within the current project budget and timeline to pursue work addressing the third recommendation, we do offer several suggestions regarding the work that is needed.

Organization of the Report

The next five chapters address the use of the SAT and its validation as a classification tool in the Air Force. We begin in Chapter 2 with background on strength testing, including how the SAT is used by the Air Force today, as well as a review of research that has been conducted on strength testing in civilian employment settings. Chapter 2 ends with a discussion of the three areas needing further investigation. In Chapter 3, we describe the results of our interviews and in-person observations at the MEPS. The next two chapters describe our initial work on developing an alternative method for defining job requirements. Those chapters describe the methodology (Chapter 4) and report the results (Chapter 5) of the web survey that RAND developed to assess physical strength requirements. Chapter 6 concludes with our recommendations for the Air Force's use of the SAT in the future, along with a discussion of the research that is still needed to support its continued use.

2. Background and Research on the Strength Aptitude Test

Strength Testing in the Air Force

In 1976, the Air Force instituted its first strength test, to measure what it called *Factor X*. This first Factor X test was considered experimental, and from 1977 to 1982 various types of strength tests were explored and studied empirically. Based on the results of those studies, the Air Force revised the Factor X test to involve a nine-step incremental lift process, renamed it the "Strength Aptitude Test," and began screening people using the new test in 1987. The SAT is a specific protocol using a specific type of incremental lift machine (ILM). We use the term "SAT" to refer to the entirety of the protocol as it is designed to be applied in the Air Force; alternative uses and protocols using the ILM refer to the machine itself.

How the SAT Is Used in the Air Force

Today, the SAT is still used at MEPS stations across the country for screening enlisted personnel for entry into the Air Force and into specific career fields. The very same machines that were introduced in 1987 (and pictured in Figure 2.1) are still being used today. The Air Force's ILMs are similar in many ways to weight-lifting machines that one might find in any local gym. For example, they include a weight stack that can be adjusted to accommodate varying weights and a lifting bar connected to the weight stack by a series of cables. The ILMs, however, were designed specifically for the military to meet a predetermined set of test specifications. These machines are often referred to as incremental lift machines or incremental lift devices (ILDs) because they allow users to start out lifting the bar with lowest weight setting (i.e., the weight of the bar alone), and to gradually increase the lift weight in increments of 10 pounds.

The Air Force's ILMs stand at more than 7 feet tall. This permits test takers to lift the bar smoothly past the 6-foot mark (i.e., the Air Force's required lift height) without abruptly hitting the top of the machine. The handle bar (shown in Figure 2.1) includes hand grips that rotate to accommodate the change in hand position that occurs as the lift progresses upwards (see Figures 2.1 and 2.2 for a comparison of the initial versus final hand positions; the overall motion is designed to be one smooth motion (although McDaniel, Skandis, and Madole, 1983, observed that individuals who could lift to shoulder height but not above were sometimes instructed in the "jerk" technique to complete the lift). The handle bar is designed to weigh exactly 40 pounds before adding any weight from the weight stack. Each weight in the weight stack weighs 10 pounds and each ILM accommodates a total lift weight of at least 110 pounds.[1]

[1] Aume (1984) details the prototype machine development.

Figure 2.1
The Incremental Lift Machine—Proper Starting Position for Initiating the Lift

SOURCE: Unpublished Air Force briefings.

The Air Force assigns scores on the ILM using a letter corresponding to each lift weight. The letter scores and the corresponding weight scores are shown in Table 2.1. All applicants must receive, at minimum, a score of “G” or 40 pounds to qualify for entry into the Air Force.

Figure 2.2
The Incremental Lift Machine—Bar Lifted to the Required Height

SOURCE: Unpublished Air Force briefings.

Table 2.1
SAT Scores and Corresponding Weight Values

Value Recorded in Personnel File	Corresponding Lift Weight
F	Less than 40 pounds (failing)
G	40 pounds
H	50 pounds
J	60 pounds
K	70 pounds
L	80 pounds
M	90 pounds
N	100 pounds
P	110 pounds

McDaniel, Skandis, and Madole (1983)[2] provide a number of specific suggestions regarding the use of the ILM as part of the SAT protocol. These include emphasizing the voluntary nature of the test, withholding information regarding test scores from recruits until after testing is completed to minimize motivation to overexertion, performing testing in private to minimize motivation to overexertion, disallowing multiple attempts at any single weight level, and blocking the information regarding the weight being lifted from view during testing. They also suggested a low starting point, such as 20 or 40 lbs., and small additional weight increments. Moreover, other current operational procedures, such as the starting weight and the small incremental additions, follow the recommendations of McDaniel et al.

The following guidance for how to administer the SAT is provided in AFRS Instruction 36-2001, section 4.21:

> 4.21.1. With the applicant facing the ILD, have him or her grasp the handles with an overhand grip, palms down. Feet should be approximately a shoulder width apart. Have the applicant bend his or her knees slightly and keep the back as erect as possible.
>
> 4.21.2. Have the applicant perform an overhead press, lifting the weights as rapidly and as comfortably as possible and ensuring either they reach the Air Force level that is marked on the machine or to a full arm extension. They will not use their lower body during the press.
>
> 4.21.3. Be sure to start at level ―G (40 pounds) for all applicants. If they are able to lift this weight, go to the next level ―H and so on. Continue the test in this manner until one of the following events occur: (1) the applicant elects to stop, (2) the applicant is unable to raise the weight to the proper level, or (3) the applicant has lifted all the weights up to the 110 pound maximum allowed.
>
> 4.21.4. If the applicant at any time fails at a weight level, the previous lift level will be his or her x-factor.

---

[2] Joe McDaniel is the researcher originally responsible for the development and application of the SAT in the Air Force.

Proper and consistent protocol administration is an important issue when considering the fairness of a test (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). Guidance such as that described here is an important vehicle for establishing consistency in administration. However, simply having appropriate guidance is not sufficient; it is essential that the guidance be put to its intended purpose of standardizing actual implementation. To our knowledge, examination of actual implementation of the above policy and guidance had not been done.

Table 2.2 shows the minimum SAT scores required for admission into each AFS in the Air Force.[3] As can be seen from Table 2.2, about half of the specialties have no restriction beyond the minimum 40-pound requirement for entry into the Air Force. Of those that do have a higher minimum, the requirement of 70 lbs. accounts for the largest number of career fields, followed by 60 and 50 lbs. Only a handful of AFSs require 80, 90, or 100 lbs. Although the SAT is designed to be scored up to 110 lbs., no AFSs currently have that high of a requirement.

[3] Appendix A provides a complete listing of the Air Force Specialty Codes (AFSCs) and the titles of the AFSs for reference purposes.

Table 2.2
Minimum Strength Aptitude Score Required for Entry by AFSC

40 lb.			50 lb.	60 lb.	70 lb.	80 lb.	90 lb.	100 lb.
1A3X1	3D0X3	5J0X1	1W0X1	1A7X1	1A0X1	2A3X1A,B,C	2M0X2	2A5X2A,B
1A4X1	3E5X1	5R0X1	1W0X2*	2A0X1P,S	1A1X1	2A5X1A,B,C	3E0X1	C,D
1A6X1	3H0X1	6C0X1	2A6X2	2A5X3C,D	1A2X1	E,F,G	3E1X1	2A6X3
1A8X1	3N0X1	6F0X1	2A7X1	2A6X1B,C	1C2X1*	3D1X7		3E2X1
1A8X2	3N1X1X	7S0X1	2R1X1	D,E	1C4X1*	3E8X1		3E7X1
1B4X1	3N2X1	8A100	2T1X1	2A6X4	1P0X1			
1C0X2	3S0X1	8A200	3D0X2	2A7X3	1T0X1			
1C1X1	3S1X1	8B000	3D0X4	2A7X5	1T2X1*			
1C3X1	3S2X1	8B100	3D1X1	2P0X1	2A3X2A,B			
1C5X1	3S3X1	8B200	3D1X2	2S0X1	2A3X3A,B,E			
1C6X1	4A0X1	8C000	3D1X4	2T2X1	F,H,J,K			
1C7X1	4C0X1	8D000	3D1X5	2T3X1	2A5X1D,H			
1N0X1	4D0X1	8E000	3D1X6	2W0X1	2A5X3A,B			
1N1X1A,B	4H0X1	8F000	3E9X1	3E4X1	2A6X5			
1N2X1A,C	4J0X2	8G000	3M0X1	3E4X3	2A6X6			
1N3X1	4J0X2A	8P100	3N0X4	3E6X1	2F0X1			
1N4X1	4M0X1	8R000	4A1X1	3N0X2	2M0X1A			
1S0X1	4N0X1	8R200	4A2X1	4B0X1	2M0X3			
1U0X1	4N0X1B,C	8R300	4E0X1		2T0X1			
2A7X2	4N1X1	8T000	4P0X1		2W1X1C,E,F			
2G0X1	4N1X1B,C,D	9C000	4R0X1B,C		J,K,L,N,Z			
2M0X1	4R0X1	9D000	8P000		3D1X3**			
2M0X1B	4R0X1A	9E000	8S000		3E0X2			
2R0X1	4T0X1	9F000			3E3X1			
2T3X2A,C	4T0X2	9G100			3P0X1			
2T3X7	4V0X1	9L000			3P0X1A,B			
2W2X1	4Y0X1				X4N0X1			
3D0X1	4Y0X2				8M000			
					9S100			

SOURCE: AFECD, 30 April 2011.
NOTES: *indicates the AFS was previously closed to women as these jobs are combat positions; ** indicates that women in the AFS were restricted from assignment to units below brigade level whose primary mission is to engage in direct combat on the ground. See Appendix A for the AFS names corresponding to each AFSC.

Gender Differences Among Air Force Applicants

There are differences in how men and women score on the test and, as shown in Table 2.3, those differences are large.

Table 2.3
Gender Differences on the SAT

	Mean	Standard Deviation	Sample Size	Difference in standardized units (Cohen's *d*)
Males	105.9	8.8	10,923	
Females	71.2	16.3	3,195	–2.65

SOURCE: DMDC data on the SAT, 2002–2008.
NOTES: Cohen (1992) defines standardized differences of .20 as small, .50 as medium, and .80 as large. Black, Hispanic and Asian means are compared to the White non-Hispanic mean. The Female mean is compared to the Male mean.

Concerns about such group differences in the employment-testing context are historically linked to cases of employment discrimination against racial/ethnic minorities and women. Although military-specific legislation can override general civilian guidelines, even in the military context civilian guidelines are otherwise considered best practice and hence are relevant to an examination of selection tools used in the military. Particularly applicable in this instance is Title VII of the Civil Rights Act of 1964 (also 1991), which prohibits discrimination on the basis of membership in a protected group (including race, gender, religion, and national origin).[4] Under the *Uniform Guidelines on Employee Selection Procedures* (Equal Employment Opportunity Commission [EEOC], 1978),[5] discrimination claims may be considered under two legal theories in an employment testing context, although only one, adverse impact, is relevant here.[6] Adverse impact occurs when a much larger proportion of one protected group (e.g., men) than another protected group (e.g., women) is selected based on the test results. Because of the differences in physical strength shown in Table 2.3, the SAT will exhibit adverse impact against women (and possibly minority groups) when a career field requires higher cut scores for entry.[7] Although concerns about adverse impact in physical testing typically involve issues of gender differences, as shown in Table 2.3, racial differences may still be relevant (see, e.g., Blakely et al., 1994) and therefore should still be examined.

When adverse impact against a protected group occurs, the EEOC guidelines (1978) state that the test is permissible only if it predicts an important outcome on the job (i.e., has evidence

---

[4] Title VII applies to nonmilitary employers; it does not apply to the military.

[5] The EEOC *Uniform Guidelines* provide interpretation and guidance on what constitutes unlawful discrimination under Title VII.

[6] Disparate treatment is the other. For a review of legal issues in selection, see Gutman (2012).

[7] Although adverse impact is based on selection ratios rather than mean difference, it is highly probable that adverse impact will occur, given mean differences of this size.

of validity in the particular employment context). Thus, tests with adverse impact may still be used, but they must be demonstrably relevant to some important job-related outcome including performance, promotion, separation from the organization, and work injuries.

Legal Context for Strength Testing

The literature on physical ability testing indicates that several different potential factors are important for consideration in jobs that require physical abilities. Gebhardt and Baker (2011, p. 170) describe seven factors.[8]

1. Muscular strength (also called static strength): the ability to apply force to lift, push, pull, or hold objects.
2. Muscular endurance (also called dynamic strength): the ability to apply force "continuously over moderate to long time periods."
3. Aerobic capacity (also called cardiovascular endurance): the ability of the "respiratory and cardiovascular systems to provide oxygen continuously for medium- to high-intensity activities performed over a moderate time period (e.g., > 5 minutes)."
4. Anaerobic power (also called explosive strength): ability to perform activities of high intensity but short duration by using stored energy.
5. Equilibrium (also called balance): ability to keep the "body's center of mass over the base of support (e.g., feet) in the presence of outside forces (e.g., gravity, slipping on ice)."
6. Flexibility: ability to "bend, stoop, rotate, and reach in all directions with the arms and legs through the range of motions at the joints (e.g., knee, shoulders)."
7. Coordination and agility: ability to "perform motor activities in a proficient sequential pattern by using neurosensory cues such as change of direction."

This list highlights the fact that the SAT addresses only one narrow aspect of the domain of physical abilities. The SAT, or any similar lifting test using the ILM, is a measure of upper-body muscular strength. Thus, it has a singular focus.

As described, strength has multiple factors, and the differences between men's performance and women's performance are higher on tests of maximum lifting versus rapid repetitive lifting, upper-body tests versus lower-body tests, and tests with stricter protocols versus tests with less strict protocols (Messing and Stevenson, 1996). Therefore, Messing and Stevenson note that a test that exhibits these characteristics, such as the SAT, would show larger differences than other strength tests. Further, McDaniel, Skandis, and Madole (1983) do note that the SAT ILM protocol is more restrictive than what would likely be found on the job and "is not representative of real-world lifting" (p. 30), although they also note that the protocol improves the safety of the test.

The fact that gender differences exist on the SAT is a concern because it indicates adverse impact. If that impact is not merited by job demands, it could invite legal challenge.[9] As discussed, it seems likely that the requirements for the SAT likely approximate characteristics

---

[8] Other authors suggest slightly different physical ability classifications (e.g., Hogan, 1991; Knapik et al., 2004).

[9] Civilian employment practices that show adverse impact can be challenged under Title VII. However, military challenges to such practices are not governed by Title VII.

that are *more stringent* than those required to meet the minimal qualifications for the job. Jackson (2000) cited recent legal context in the United States that is more supportive of tests and cutoffs that are predictive of the minimal qualification necessary for adequate performance on the job; these make the extant gender differences more problematic. More recently, Gebhardt and Baker (2011) indicated that the courts are currently still more supportive of selection or retention tests that are tied via thorough job analysis to the minimal qualification necessary to perform the job. Hence, validity evidence substantiating the SAT's usefulness for performance on the job is quite desirable, as is an explicit tie to minimal rather than average or above-average qualifications.

A validity argument for a test in the selection or classification context is one in which evidence is accumulated to support different inferences. Validity evidence from multiple sources provides the best overall support for the use of the employment test and support for a variety of the necessary inferences (see, e.g., American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999; Gutman, 2012). Here we discuss three sources:

- consequences of testing
- test content
- relation of test responses to pertinent variables.

The first source of test validity evidence discussed here, *consequences of testing*, refers to the degree to which the consequences of using the test results can be attributed to properties of the test. These consequences, such as different passing rates, are what spur legal challenge, and physical ability tests like the SAT are one of the most common types of employment assessments to be legally challenged by job candidates (Robertson and Smith, 2001), and challenged with relative success (42 percent of challenges are successful; Terpstra, Mohammed, and Kethley, 1999). If group differences in test scores are demonstrated and they reflect differences in characteristics *not relevant* to job performance, then the test's validity can be questioned. Hence, the tie to on-the-job performance and characteristics of the job itself is the essential evidence for validity.

The second source of validity evidence, *test content*, refers to how the test features relate to what the test is trying to measure. Basically, the test should comprehensively measure the content domain it is developed to measure, and not measure things that are irrelevant. Evidence establishing the validity of test content is often drawn from expert judgments about the relationship between test content and the on-the-job behaviors the test is purported to measure. For example, the design of a survey to assess the types of strength-requiring movements needed for a given job could benefit from expert judgment regarding which prototypical work-related movements to include. Other job analysis techniques may also be used to determine what types of physical abilities are needed. Again, the point is to accumulate evidence that the test is relevant to the job.

The third source of evidence, the relation of test responses to other variables, is also key to employment testing, particularly when concerns about adverse impact may be present. This source of evidence usually involves substantiating the inference that the desired qualities or behaviors underlying the selection test are predictive of the desired qualities or behaviors

underlying later performance on the job. Sometimes, the desired qualities or behaviors are hard to define or measure, and proxies are used. With regard to the SAT, traditional measures of task and other types of performance-related work behavior may be an imperfect match. The goal of the SAT is to predict *physical* performance on the job and ensure that enlistees are able to meet the physical demands that will be required. Many of the typical types of performance information collected as part of a performance management system (performance reviews, promotion speed, etc.) may have relatively little relationship to the physical performance of job tasks per se. In a large organization such as the Air Force, there are many jobs, and the need for physical capabilities to do the job is likely to vary widely. The extent to which the typical performance measures collected incorporate consideration of physical performance is likely to vary based on how relevant physical capabilities are to performing the job. Again, a job analysis is recommended to make sure that the physical abilities required on the job are utilized when measuring the performance criterion in the physical performance domain (Gebhardt and Baker, 2011) or other contexts (e.g., Schmitt and Sinha, 2011).

To summarize, in the context of selection for physical capability on the job, it is even more important to gather comprehensive validity evidence. This means ensuring that the measurement of physical capability is a good measure of strength, muscular or cardiovascular endurance, and/or other related physical capabilities that the job analysis evidence suggests will be important; and that the measure of job performance is in turn a good reflection of the physical requirements of performance on the job. When the job is a physical one, seeking the appropriate tool is likely to pay off: Research has demonstrated the usefulness of physical ability testing for a variety of physically demanding jobs (e.g., Blakely et al., 1994; Henderson, 2010; Hogan, 1991). Gebhardt and Baker note that validity coefficients for using physical ability are generally quite acceptable, but for basic ability tests they vary depending on how well the tests mimic actual physical job requirements.

Validity Evidence for the SAT

The original work that led to the Air Force's selection of the ILM is largely undocumented. However, Ayoub et al. (1987) describe several elements of that work in detail.

According to Ayoub et al., the research proceeded in three phases. In Phase I, the researchers collected job task information for a variety of AFSs using three data collection methods: interviews with supervisors, in-person examination of the objects involved, and questionnaires filled out by supervisors. They then identified a set of 13 actions (such as lifting a tool box with one hand, carrying a tool box, lifting a box, pushing or pulling objects) that represented 90 percent of all actions identified as tasks in the AFSs they studied.[10]

In Phase II, they conducted simulations of those 13 actions and also tested people on eight different strength tests, including an incremental lift to knuckle height, to elbow height, and to six feet, a hold at elbow height, a 70-pound hold, a one-handed pull, a hand grip strength test, a

[10] The number of AFSs included in Phase I and II was not reported. However, Bomb-Navigation Systems and Aviation were two that were mentioned by name.

1.25-foot vertical pull, and an elbow-height vertical pull.[11] Each test and each simulation activity was administered to a total of 527 personnel in the Air Force. Each participant completed the simulations of requisite Air Force job movement simulations and tests across a two-and-a-half day time period. From these data, the six-foot incremental lift test was identified as the most predictive for the variety of simulation activities. Regression equations were created for predicting incremental lift scores from the scores in each simulation activity (e.g., the toolbox carry or the box lift).[12]

Phase III of the research described the process that Ayoub et al. used to establish cut scores. The process included first converting the actions from a given AFS to incremental lift scores using the regression equations developed in Phase II. Next, the 25 most physically demanding tasks (based on their ILM conversion score) in that AFS were weighted by frequency, importance, and percentage of people performing them.[13] Then, the weighted ILM scores were averaged to establish the minimum score required for entry into the AFS.

From their research program, Ayoub et al. offered two key recommendations. First, the results from the supervisor interviews and in-person examinations of objects involved described in Phase I were found to be expensive and time consuming, and it was noted that the results often varied significantly from base to base. They also indicated that, for their study, questionnaires filled out by supervisors and their responses in interviews also differed from the results of the in-person examinations by researchers. For this reason, the authors stated that none of the methods they employed to collect information about the physical demands of an AFS was satisfactory and that further research would be needed to identify a better method. We note that, despite the recommendation of Ayoub et al., surveys are actually quite commonly used as part of efficient, large-scale, systematic, and legally defensible job analysis process (see, e.g., Brannick, Levine, and Morgeson, 2007; Morgeson and Dierdorff, 2011; Williams and Crafts, 1997). Common questions on strength-oriented job analysis surveys include questions about tasks performed and questions about perceived demands such as those Ayoub et al. asked, as well as questions about perceived frequency of performance and importance, as are commonly collected as part of the Air Education and Training Command (AETC) Occupational Analysis Division (OAD) survey from which the tasks themselves were (presumably) drawn.

Second, they concluded that the incremental lift to a height of six feet was the single best predictor of a wide variety of tasks that they included in their simulation, and that adding additional tests did not provide much incremental validity in predicting simulation performance. For this reason, the ILM alone was sufficient.

---

[11] No additional information regarding the activities (such as how much weight was pulled, or how many times) was provided.

[12] No information regarding what was measured in the simulation activity or how the activities were scored was provided.

[13] Although not specified in the Ayoub et al. paper, based on current procedures in the Air Force and other unpublished work, the values for frequency, importance, and percentage were drawn from the Occupational Analysis Division task analyses surveys.

Other studies also support Ayoub et al.'s (1987) findings regarding the relationship of the ILM to feet with relevant variables. In a relatively large study of Army recruits, Teves et al. (1985) indicated that out of a test battery administered prior to basic military training (BMT) the best predictor of performance on a lifting task simulation post-BMT was the ILM to six feet, although they indicated that the variables' predictive powers were not of practical significance. However, the validity coefficient reported predicting maximal lifting capability from ILM to six feet is actually quite comparable to what is acceptable more generally (e.g., Blakely et al., 1994; Schmidt and Hunter, 1998).

Myers et al. (1984) administered a test battery, including both the ILM to five and to six feet pre-BMT, compared performance on the battery with performance on several simulations of physically demanding tasks drawn from job analysis of Army jobs, and administered post-Advanced Individual Training (i.e., training that occurs just prior to starting work in their Army jobs). They indicated that the five-foot lift was the best predictor (first entry in stepwise regression, indicating the largest bivariate correlation) for the criterion of job simulation combination score.[14] However, the six-foot lift did make a substantive contribution to an alternative and less directly job-relevant criterion combination made up of physical fitness tests administered during BMT (sit-ups, push-ups, two-mile run). This second finding suggests, at a minimum, that the test has validity in the sense that it is related to physical test variables similar to those used to assess health and fitness for those in the Air Force. Myers et al. also noted that no clear method for establishing occupation-specific cut scores had been developed and suggested further research examining various methodologies was needed.

Sharratt et al. (1984) examined the ILM to six feet as a predictor of performance on a repetitive sandbag lifting task and a stooping sandbag lifting task and found reasonable validity estimates. Despite these reasonable bivariate validity estimates, other predictors did have stronger relationships and, when entered into a stepwise regression equation for sandbag lifting with other tests, the six-feet ILM entered the equations for both men and women at the fifth step. For other criterion tasks (jerry can lifting and tire changing) they indicated that none of the predictive tasks they examined improved prediction of success over having no test at all.

Rayson, Holliman, and Belyavin (2000) investigated the validity of a number of different measures of fitness for predicting performance on strength-related job task simulations that were based on a job analysis of the British Army. Criterion tasks were variations on single lifts, repetitive lift and carry at various weights, and loaded marches at various weights. The fitness battery included a number of measures. In all, over 30 measures of fitness were incorporated. Unsurprisingly, ILM to just under five feet was a good predictor (defined as having one of the five highest bivariate correlations and inclusion in the preferred models[15]) of the criterion task of single lift to just under five feet for both men and women; it was also a good predictor of the

---

[14] The job combination simulation included three tasks (1) maximum weight lift to chest; (2) carry that weight at chest height up to 200 yards; and (3) push four times lift weight on sled. These tasks included procedures to adjust weight as needed. The combination simulation also included a torque task.

[15] Preferred models were stepwise regressions with a maximum of three predictors that also preferenced other criteria, including minimization of standard deviation and classification errors.

criterion task of single lift to just under six feet for women. However, the ILM in either variation did not emerge as one of the three predictors in the preferred models for other criterion tasks, be they lift, repetitive lift and carry, or loaded march.

Knapik et al. (2004) reviewed some of these studies, as well as other literature, in their relatively recent discussion of courses of action the Army could take should the Army desire to more systematically implement pre-enlistment physical fitness testing. Their review noted that there is a fair amount of support for use of the ILM with various protocols as a test of muscular strength, and suggested that the Army may want to incorporate the ILM in conjunction with tests of other physical ability factors (physical endurance tested by push-ups, and cardiovascular fitness tested by a one-mile run), in order to pursue a course of action based on best practice in extant literature. Knapik et al. are silent regarding the height requirement for the lift although various alternatives to the six-foot lift used by the Air Force have been explored and found useful. They did indicate that a full validity study is the optimal course to ensure that such a pre-enlistment battery is job-related and assesses the entirety of potential physical abilities needed in the military (e.g., many studies simply do not include measures that tap into flexibility and balance; hence, there is little evidence for or against the necessity of that factor for Army or other military jobs).

Nevertheless, other factors need to be examined when evaluating the SAT. For example, Vickers, Hodgdon, and Beckett (2009) caution that omission of an important physical ability (e.g., if a relevant physical ability is not examined in the test battery) in a regression equation could lead to false conclusions regarding the importance of that physical ability in predicting later performance. Other studies have, however, shown that physical abilities are highly correlated, and that using a test of one physical ability for prediction may produce quite similar results to that of other physical abilities (see, for example, Blakley et al., 1994).

In addition, studies have also shown that there can be gender differences in the predictive validity of regression equations created using strength tests (see for example, Robertson and Trent, 1985; Arnold et al., 1982). While Myers et al. (1984) did find that the ILM predicts important outcomes for both men and women,[16] Stevenson et al. (1996) showed that using the same cut score on the ILM for men and women can result in a higher number of false negatives for women than for men.

Lastly, more research on the amount of improvement that could be expected during basic training is needed. It is a well-established fact that the physical abilities measured by the tests can be significantly altered through training (see, for example, Vickers and Barnard, 2010; Williams, Rayson, and Jones, 1999; Brock and Legg, 1997; Hogan and Quigley, 1994), and a number of studies have shown that training can have a large impact specifically for women (see, for example, Knapik et al., 1980; Harman et al.,1997; Knapik, 1997). Because the Air Force expends considerable effort toward physical training during the enlisted eight-week basic

[16] Differential prediction (in which the regression slopes and/or intercepts themselves are different for men and women) may be another avenue of inquiry, but relatively few studies have explored this in conjunction with the ILM. Myers et al. is an exception; they found that intercept differences led to only slight overprediction of women's performance on the most job-relevant job combination simulation score.

training course, it is highly likely that substantial gains could be made in ILM performance even though basic training itself does require a certain baseline fitness for success (see, e.g., Knapik et al., 2004). Although there has been some examination of the gains in physical skills during basic training (for a review, see Vickers and Barnard, 2010) and gains on the SAT from basic and technical training have been studied in the Army (see Myers et al. 1984; Teves, Wright, and Vogel, 1985), the amount of gain that occurs during the *Air Force's* basic training needs to be investigated further. The GAO did provide estimates of the gains on the SAT that might result (GAO, 1996). Using data provided by the Air Force, the GAO compared scores from the MEPS station tests to retest scores taken during the second week of Air Force basic training. The average gain for women was 18 pounds. The average gain for men was 15 pounds.

Although a significant amount of research has been conducted on strength testing, much of it since the establishment of the SAT, questions still remain. Areas that need additional attention include examination of the SAT's relationship with on-the-job performance, a reexamination of alternative tests, further examination of adverse impact against women, and a concrete estimate for the amount of improvement that occurs as a result of basic training. Further discussion of the need for additional research is provided at the end of this chapter and in Chapter 6.

How Job-Specific SAT Minimums Are Determined

Currently, the minimum SAT cut points for each AFS are determined according to the following four steps:

1. Identify and select career fields.
2. Resurvey career fields.
3. Produce new cut score estimate.
4. Examine new cut score and adjust if not satisfactory.

Each step is explained in greater detail below.

Identify and Select Career Fields for Reexamination

Career field managers inform the Force Management Division (A1PF) or the Air Force Personnel Center (AFPC) that they would like a reexamination of the SAT minimum for their career field. Career fields are "resurveyed" as time and resources permit. Newly created career fields or recently merged career fields are also considered for examination to establish new cut scores.

We spoke to a number of career field managers throughout this study and discovered that many were unaware that they could request a reexamination of the SAT cut point and many did not know who to contact if they wanted it to be reconsidered.

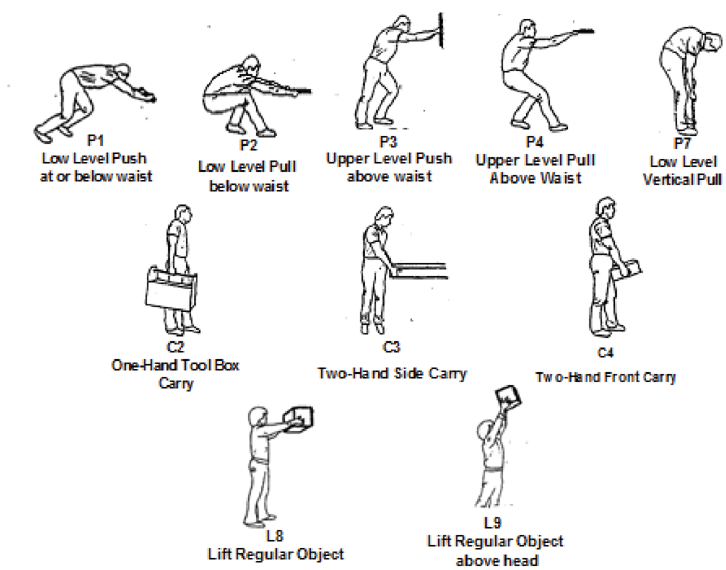
Resurveying Career Fields

The cut score reexamination process is conducted by an Air Force contractor, following a series of prescribed steps developed at the time the SAT was instituted.[17] The process is referred to as the *resurvey process*. The process does not involve a paper-and-pencil or online survey, despite what the name might suggest. Instead, it involves site visits, interviews, observations of key activities, and weighing the objects involved in those activities.

The process starts with the selection of three sites to visit per AFS. Because resources are limited, efforts are made to collect data on more than one AFS during each site visit. During the site visits, the contractor conducts short interviews of workers regarding how physically demanding tasks are performed. The following factors are recorded during the site visits:

- a short description of the task
- how the task is performed (lifting, pulling, carrying, etc., see Figure 2.3 for examples of a few action types and the codes assigned during the resurvey).
- how much the objects involved in the task weigh
- how many people are involved (i.e., total number of people helping).

Figure 2.3
Examples of How Movement Types Are Coded During Site Visits



SOURCE: Unpublished Air Force briefings.

The tasks identified during the site visit are matched to the tasks listed on preexisting occupational analysis reports, which are generated by the OAD about once every three years. Those reports provide the following additional information, which is also recorded:

---

[17] The process was developed by Joe McDaniel, one of the authors of the Ayoub et al. article and the person who was originally responsible for instituting the use of the SAT in the Air Force.

- Frequency of task (i.e., number of times performed per year)
- Percentage of personnel in the AFS who perform the task.

However, it is unclear how the contractor makes the connection between task statements and the activities performed or, for that matter, determines the movement types. For example, to the extent that the tasks in the occupational analysis reports are not directly aligned with the action identified in the site visits, relying on them for estimates of the frequencies and percentage of people performing the task may be inappropriate. Consider, for example, that the occupational analysis report includes a task such as "use a cement mixer." The actions observed by the contractor might include "putting 10 lb. bags of cement ingredients into mixer" or "moving the cement mixer." How should the occupational analysis report's frequency and percentage of people who "use a cement mixer" be applied to the two very different actions associated with it that were observed by the contractor? Moreover, which movement type the contractor decides to use to classify the activity would vary based on what type of physical behavior is envisioned (e.g., "front carry" vs. "lower-level push"). No data are available to confirm that actions identified in site visits and the information reported in the occupational analysis reports can be accurately connected in this way.

Produce New Cut Score Estimate

The data points collected during the site visits and pulled from the occupational analysis reports are fed into a SAS program[18] developed for use with the SAT. The general process applied by the program is as follows:

- Convert each action type (push, pull, carry, lift, etc.) into its equivalent force on the SAT (X1). This is calculated using regression equations created by regressing ILM performance on each action type. The equations were created using data collected during the research that led up to selecting the ILM for use in the Air Force.
- Weight each X1 value by the frequency of the task and the percentage of people in the career field who perform the task.
- Average the weighted X1 values. The result is taken as the average SAT score for the career field.

The exact formulas and procedures used in the program are provided in Appendix B.

The documentation for how the program was developed is limited. A cursory explanation is provided in Ayoub et al., but many important details are left out. Ten of the regressions equations used in the SAS code also appear in the study described in Ayoub et al. Three others that are reported in Ayoub et al. are not consistent with those in the SAS program, and six are not discussed anywhere in Ayoub et al. For the six missing equations and the three that do not match those reported in Ayoub et al., descriptions of the actions and estimates of the R squared values for the regression equations are not available.

---

[18] SAS is a well-known statistical data analysis software package used by many social scientists. Commands for the data analyses are programmed in SAS syntax and can be viewed in a standard text editor.

Examine New Cut Score and Adjust If Not Satisfactory

The number produced by the program is sent to A1PF and the career field managers. If they are satisfied with the cut point it is kept; if not, the cut point may be revised to better reflect the requirements expressed by the career field manager.

There have been multiple instances in which a career field manager has requested that the SAT cut score be revised, yet the above process has yielded the same cut score as was previously in place. This is potentially an indication that important facets of the job may not be accounted for in the above process. For example, the SAS program produces an "average" strength requirement and fails to consider task importance. As a result, if there is one particularly strenuous task that all members of the career field must be able to perform, the SAS would likely underestimate the minimum cut score required.

Conclusions

Our review of the strength training literature and the methodology for calculating SAT cut scores for Air Force career fields points to several areas where further research would be useful in enhancing Air Force strength testing practices.

The Process for Setting Cut Scores Should Be Updated

In the course of this project, we interviewed Dr. Joe McDaniel, the researcher who conducted the work leading to the Air Force's adoption the SAT. From that work, he established the procedures that are still used today in computing the cut scores. His insights were invaluable in helping us understand the research that served as the foundation of the SAT. However, through these interviews we also were able to identify gaps in the documentation of the research that supported the process. We also determined that there were areas needing additional research.

The first gap in documentation concerns the formulas used in the SAS program code. For example, while McDaniel's data were used in creating the regression equations and the other formulas used in the program, creation of the formulas was contracted out to an external statistician. No documentation regarding how the formulas were established or the quality of those formulas was retained. This is unfortunate, because without more information about the formulas we cannot evaluate the appropriateness of their use. For example, the following information is needed:

- R-squared values for the undocumented regression equations and explanations for why three of the regression equations differ from those reported in Ayoub et al.
- Evidence showing that the regression equations do not differ in meaningful ways by gender.[19]

[19] In regression equations that are used for purposes of selection, it is standard practice to examine underprediction as well as differential validity for protected groups (i.e., race and gender). For more on this see AERA, APA, and NCME (1999).

Regardless of the missing information, it is worth noting that converting the force involved in one action to the associated force on the SAT using regression equations may not be the best approach. If the results of simulation data are going to be used for converting one type of action to another, we would suggest considering using some form of equipercentile equating instead of regression equations.[20]

Another gap concerns documentation of the actions included in the regression equations. Although simplistic illustrations are available for some of the actions (such as those shown in Figure 2.3), some actions appear to have no corresponding description or key to identify their meaning. Some are even flagged as unusable in the SAS code because the action is undefined.

A third gap concerns documentation supporting the methodology for setting cut scores. Answers to the following questions are not available:

- Why use the average of the 25 most demanding strength requirements for establishing the cut score? For some jobs, the average across all tasks may overestimate the requirements, particularly if the percentage of people performing the tasks is low. In other AFSs it may underestimate the requirement. In all cases it is unclear that the requirement establishes what is needed for a minimally acceptable person in the job.
- Why treat frequency and proportion of people performing the task equally in weighting the tasks? This is not explained in Ayoub et al.
- Why exclude importance from the weighting calculations? Ayoub et al. includes importance in the calculations, yet importance is not considered in the SAS code. As noted previously, if one task is critical for success on the job but importance is not considered in weighting the requirements, the resulting cut score may be significantly underestimated for some AFSs.
- How should duration of the activity factor into the cut score? Lifting boxes for four hours straight is substantively different from lifting a box once a day. Ayoub et al. does not describe the duration of the activities in their simulations, and it is not clear the extent to which extending the duration of the simulations would have resulted in different findings.

In addition to changing the methodology to better address these questions, we also suggest consideration of other well-established methods for establishing cut scores on selection tests. There is no single best method for setting cutoff scores. Truxillo, Donahue, and Sulzer (1996) note that when the desire is to set a cutoff at the level for a minimally competent person on the job (i.e., criterion-related validity settings), utilizing expert judgments has gained currency due to its track record for defense from legal challenge, though the authors note that multiple pieces of evidence should be gathered to support the cutoff. Sothmann et al. (2004) described in great detail one method for setting cutoff scores that predict minimally acceptable physical performance on the job for firefighters based on physical demands; however, their approach is highly tailored to a specific job and may not work as well in the context of multiple jobs such as the context of selection and classification into Air Force enlisted jobs. Note that best practice for setting cut scores for strength requirements often accords more-frequent activities greater

[20] Regression equations underpredict the force for actions involving object weights above the mean, and overpredict the force for actions involving weights below the mean. For more on methods for equating, see Dorans, 1990.

consideration (as does the current Air Force algorithm). Thus, an algorithm based solely on frequency might exclude from consideration any activities that do not occur at least once a month. However, in practice, a combination of both frequency and importance is often used, such that very or extremely important activities are considered even if they occur with low frequency. An example of this might be a pararescue airman lifting a 180-pound person onto a stretcher for an airlift out of a combat area. While such an activity may occur once a year or less, it is quite important and an essential job activity when it does happen. Other methods of setting cut scores are discussed in Cizek (2001). Regardless of the method chosen, best practice and legal context suggest that thorough documentation of the process and procedures to set cut scores is required. Moreover, a clear tie to the minimal rather than average qualifications necessary for job performance would potentially place the cutoff score process on more secure legal footing.

There are also additional concerns regarding the manner in which the data fed into the SAS program is collected. First, the site-visit methodology may not be obtaining a representative sample of the physical requirements of the job, since the contractor only goes to three base locations and those locations are not randomly selected from all base locations. It is very possible that the weights of the objects and the procedures for handling those objects differ from location to location (especially when personnel are deployed outside the United States), and the current methodology has no way to capture that information. Ayoub et al. expressed a similar concern and therefore suggested that other data collection methods should be explored. In addition, there is a significant leap taken when the contractor identifies a task in the occupational analysis reports and assumes that the data on that task's frequency and percentage of people performing it can also be applied to the actions and objects identified during the site visits. For example, it is possible that while use of a particular object (such as a cement mixer) might occur daily and be reported as such in the occupational analysis (e.g., when asked how often they "use a cement mixer," respondents said "daily"), moving the object may occur much less frequently. To the extent that the tasks in the occupational analysis reports are not directly aligned with the physical action identified in the site visits (as in this example of "moving" the cement mixer versus "using" the cement mixer), relying on it for estimates of the frequencies and percentage of people performing the task may be inappropriate. No data are available to confirm that actions identified in site visits and the information reported in the occupational analysis reports can be connected in this way. Thus, although it is clear that the Air Force process attempts to utilize job analysis data (as is best practice), the actual correspondence of the physical requirements to the elements assessed in the job analysis is suboptimal.

Because occupational analysis surveys are already administered online to every enlisted AFS every three years, adding elements to the survey to collect physical demands information would be a simple solution to the concerns expressed above. Although Ayoub et al. also expressed concern regarding the accuracy of supervisor's questionnaire responses in their study, there is evidence that a paper-and-pencil or online survey of job incumbents could be an effective tool for collecting information about the physical demands of the job (see, for example, Koym, 1975; Blakley et al., 1994; Hughes et al., 1989; Rayson, 1998). Moreover, much of the original research leading to the adoption of the SAT relied in part on questionnaire data (McDaniel, Skandis, and Madole, 1983). For this reason, we set out to design and test a survey of Air Force

officers and administered the survey to six AFSs. The content of that survey, the results, and suggestions for ways to improve the survey are reported in Chapters 4 and 5.

The Strength Aptitude Test Should Be Further Validated

More research on the SAT's predictive validity is needed to fill the gaps in the work from the 1980s, and research on the validity of other strength and stamina measures that assess the multiple potentially relevant types of physical abilities (e.g., muscular and cardiovascular endurance, coordination, and agility) should be included in this assessment as well. The following are some examples of work that should be done.

First, more research is needed to determine whether the SAT is equally valid for both genders and across races. Because there are large gender differences, the use of the SAT excludes women from certain jobs at higher rates than men; if the SAT cut score is set too high, or if the test is not valid predictor of ability to perform the job, it could be excluding them unfairly and unnecessarily.[21] Examination of whether validity holds across races should be explored as well.

Second, research should examine whether there are other measures that are equally valid predictors, or whether the measure should depend on the requirements of the job. For example, in some jobs lifting to six feet may be particularly relevant. In others, actions involving lower body strength (such as pushing objects) may be more important. To the extent that there are greater gender disparities in upper body strength than lower body strength, matching the type of test to job requirements may be important. Studies have shown that other tools (including ILM to five feet rather than six; leg press) could have equal or better validity and some that may have fewer gender disparities. These other measures should be explored further and their ability to predict performance on the job should be evaluated empirically and compared with the SAT.

Third, research should examine how much SAT scores change from the time in which applicants are tested at the MEPS to the time in which they begin performing on the job. Much of the research supporting the SAT has been conducted using artificial simulations in which predictor scores and criterion scores are collected within days of each other. Very little of it has examined the extent to which physical conditioning during basic and technical training serve to increase scores. Such increases need to be accounted for in setting the minimums for scores at the MEPS. For example, if scores increase by ten pounds after completing training (which is approximately the size of the difference reported by Teves, Wright, and Vogel, 1985) the minimum cut scores required at the MEPS should be lowered by ten pounds to account for the expected increase. Similar research on the impact of basic and technical training should be explored with alternative measures as well.

Fourth, examination of the relationship with job performance is critical to showing that strength testing (SAT or otherwise) is necessary for a given AFS. If the SAT or other measures cannot predict the ability to perform the physical requirements of the job, then their use should be discontinued. Other factors that are important to physical performance on the job, such as job-related injury rates, would also be potentially useful to help determine if there is risk involved in

---

[21] Similar concerns were expressed in the 1996 GAO report.

discontinuing strength testing (although events that happen infrequently can be difficult to predict).

Due to the limited resources available for this study, we could not begin to address this recommendation for additional research on the validity of the SAT. However, in Chapter 6, we describe the methodology that would be needed and provide some suggestions for immediate next steps.

More Information Is Needed on How the Test Is Actually Used in Practice

Our review of the established guidelines for administering the SAT (in Air Force Recruiting Service [AFRS] Instruction 36-2001, 2012), and examination of data provided to us from the Defense Manpower Data Center (DMDC) on applicants' scores on the SAT, led to a series of additional questions about the test as it is currently used:

- How are the MEPS stations actually administering the SAT? Does the practice adhere to the guidelines?
- Do applicants know about the test and its purpose ahead of time? Do they try their hardest? Do they prepare?
- Are there differences in how the SAT is presented to or administered to applicants, particularly groups of applicants such as women and men?
- What do the machines look like? Are they all the same? Are they new?

Earlier, we noted that the mere existence of a testing policy is not sufficient. The answers to these questions would help address whether the SAT is a fair and/or unbiased test. If the test is being administered in the same way to all applicants, if all applicants have the same information about the test and its purpose, and if the test administration is consistent with the procedures outlined in AFRS Instruction 36-2001 (2012), then we would have few concerns regarding fairness issues related to the manner in with the test administration is occurring. Therefore, to answer these questions, we conducted a series of observations and interviews with MEPS personnel who administer the tests and with applicants taking the tests. The results of those interviews are presented in Chapter 3.

3. Observations and Interviews at the Military Entrance Processing Stations

There are 65 Military Entrance Processing Stations located primarily within the continental United States where recruits of all branches of the military—Army, Navy, Air Force, Marine Corps, and Coast Guard—are processed for enlistment. At the MEPS, recruits are screened on a number of criteria including scores on the Armed Services Vocational Aptitude Battery (ASVAB); the results of a medical examination; and physical, strength, and/or endurance tests (such as the SAT).

To better understand the operational use of the SAT, we traveled to four medium- to large-sized MEPS locations to observe the SAT administration process, interviewed applicants taking the test, and interviewed the Air Force staff at the observation sites who screen recruits and administer the SAT (i.e., the liaison non-commissioned officers or LNCOs). We also interviewed LNCOs by phone at four other medium- to small-sized MEPS sites.[22]

Specifically, we sought to answer the following questions:

- Are there damaged incremental lift machines or machines in need of repair?
- Are the LNCOs administering the SAT in the same way across locations and in the way in which it was designed?
- What are recruits' reactions to the SAT?

As shown in Table 3.1, 34 recruits and 17 LNCOs were interviewed or participated in our observations. Ten (29 percent) of the participating recruits were women. Interview questions are located in Appendix C.

Table 3.1
Number of Participants at Each MEPS Location

Site	In-Person/Phone	Recruits	LNCOs
MEPS 1	In-Person	3	2
MEPS 2	In-Person	16	4
MEPS 3	In-Person	8	2
MEPS 4	In-Person	7	2
MEPS 5	Phone	NA	2
MEPS 6	Phone	NA	2
MEPS 7	Phone	NA	2
MEPS 8	Phone	NA	1
Total		34	17

[22] MEPS size was measured by the average number of recruits processed annually. Large, medium, and small MEPS processed an average of about 1,200, 800, and 400 recruits per year, respectively.

Observations of the Incremental Lift Machines

In general, the ILMs we observed were in good working order, though we did identify a few machines in need of repair or replacement. Some differences exist in terms of visible information regarding use of the machines and where machines are located at the MEPS, which we describe here.

All of the ILMs we observed displayed a line marking a height of 6 feet on the front of the machine, and each weight was marked with the letter of the corresponding SAT score on the back of the weight stack (i.e., not visible to the recruit taking the test). At one MEPS location, a poster with the same images shown in Figures 2.1 and 2.2 (i.e., showing a test-taker performing lifts using the proper form) along with written instructions regarding proper form was posted next to the ILM. At another MEPS location, LNCOs we interviewed by phone also reported having the same poster. The rest of the locations did not have such a poster. At a couple of locations, a piece of paper next to the machine listed the weight amounts and their corresponding letter (e.g., N=100), positioned where it would be visible to the person being tested. Nearly all machines had a mat positioned where the person stands while taking the test.

Some machines used the original weight stack pin, and others were using a newer pin not specifically designed for the ILM because the original pin had been lost. LNCOs at locations using newer pins noted that they occasionally fall out of the machine or get stuck, and suggested that the pin is something that should be fixed. LNCOs at some locations also mentioned that the track sometimes sticks a little rather than running smoothly, but otherwise reported the machines in good working order. Although none of the machines we observed had any other problems, one LCNO did mention that at another MEPS station, one of the machines was badly damaged and needed to be replaced. The machines we observed were solid and stable when in use.

Location of the machines within the MEPS varied. For example, one was located in a waiting room, one was located next to the base of a stairwell, and one was located in a medical testing area being shared with the medical staff. One reason cited for the varied locations was the height of the machine. The machine is over seven feet tall, and not all of the rooms at the MEPS can accommodate its height. Other LNCOs mentioned that the ILM was moved out of the medical area after the medical personnel refused to continue to administer the test. In at least one case, it was moved into the LNCO office next to the desks and a copier. Most stations reported having sufficient space to operate the machine; however, one LNCO said that while the space was adequate, it would be better if the ILM were in a slightly more open space.

Each location we visited had only one machine; however, one location contacted by phone reported having two working machines. Given this unexpected finding and the discovery that there was reportedly at least one badly damaged machine, we immediately suggested that A1PF conduct a full inventory of the machines by asking all MEPS locations to report the number of machines at their location and any damage to or problems operating the machines. Locations reporting more than one machine would be an obvious source of replacements for other locations reporting problems with their ILMs.

SAT Administration

We observed a total of 34 recruits taking the SAT. Table 3.2 shows summary information about recruits' SAT scores, testing times, and heights. As shown in the table, testing times were comparable for both women and men, whereas SAT scores were not.

Table 3.2
Test Times, Lift Weights, and Recruit Heights Observed During SAT Administrations

	Men				Women			
Location	Average	Min	Max	Number Observed	Average	Min	Max	Number Observed
Test time (in seconds)	63	41	162	21	59	24	120	10
Final lift weight (in pounds)	94	70	100	23	59	50	70	10
Recruit height (in inches)	69	64	78	21	63	58	68	10

Many aspects of the administrations that we observed were fairly consistent across recruits, LNCOs, and locations and were consistent with the way in which the test administration was originally designed. We did, however, discover some variations in administration.

For example, while most LNCOs were adamant about starting the administration at the minimum 40-pound weight limit and increasing it in increments of 10 (as outlined in the original test design), a couple of LNCOs admitted that sometimes they start at 70 pounds and then move straight to 100 pounds when the recruit looks really strong. They added that in those cases, the person is always able to lift 100 pounds (as they suspected). Another variation from the original design of the test is the maximum weight used. Most of the LNCOs interviewed stop the test at 100 and explained they do so because no job requires a higher score. However, consistent with the original intent, LNCOs at a few of the sites continue to 110 pounds and record 110 for a final score if a recruit completes a 110-pound lift.

We also discovered that some recruits take the test individually with only the LNCO watching, while others do so in groups with their peers watching. When administered in groups, the members of the groups offered encouraging words (like "you can do it!") to those struggling to complete the test. From our observations, having an audience cheering for them appeared to lead many recruits to try harder than they might otherwise have, although some particularly shy recruits seemed to be embarrassed by the attention and gave up very quickly.

LNCOs also differed in how they reacted to recruits who were struggling or not trying very hard. Some strongly encouraged them to try as hard possible and allowed them to reposition their feet or try the last lift again, whereas others did not offer strong encouragement or a second try. In discussions with the LNCOs, some said that they occasionally allowed a recruit to try again after everyone was finished or later in the day, whereas other LNCOs allowed no re-do's. For those who offered another chance, it was usually because the recruit wanted a specific job but

had not qualified for it. Sometimes, when recruits were permitted another try, LNCOs allowed the recruit to start at the lift weight where they left off the first time instead of starting again at 40 pounds.

During the interviews, we asked if women wearing skirts or heels had a problem taking the test, and most of the LNCOs told us that there are rules about what is considered appropriate attire at the MEPS. For example, one of our phone interviewees indicated that women are not allowed to wear a tight skirt or open toed shoes. However, some of the LNCOS we spoke to also mentioned that women will sometimes arrive wearing high heels. In those cases, they typically take off their heels to do the lifts, although there is no specific instruction regarding doing so.

In the observed visits and in the interviews with LNCOs, we discovered that the timing of the administration of the SAT might be less than ideal. All of the LNCOs said that it typically occurs after the recruits complete the physical. However, two aspects of the physical, the blood draw and the duck walks (walking while squatting without knees or hands touching the ground), could interfere with a recruit's performance on the SAT, and some of them expressed concerns about this. The blood draw sometimes makes recruits feel faint or weak (perhaps from the sight of the needle or blood), and the duck walks may work the leg muscles of some recruits to exhaustion. Consistent with this possibility, many of the recruits we interviewed said that the duck walks in fact made their legs really tired. Exhaustion from the duck walks or feeling faint from the blood draw could result in a lower SAT score than would have been obtained had a recruit not been exposed to these stressors immediately prior to the test.

Another major difference in test administration was how much information the LNCOs divulged before and during the test. For example, some LNCOs tell the recruits what score they need for a particular job, that the start weight is 40 pounds, and that every subsequent lift is 10 pounds heavier. Other LNCOs do the opposite; they intentionally tell recruits nothing about the amount of weight that they will be lifting or what is required for any particular job. One such LNCO said that knowing the weights could discourage recruits and make them think they cannot do it. Other LNCOs said that recruits learn about the test from the recruiters, so there is no need to further explain the SAT once the recruits arrive at the MEPS. As shown in the next section, the assumption about how much recruits know in advance about the test is not likely to be correct. In a few cases, LNCOs seemed to think they are supposed to mask the information so recruits will not know their scores.

Finally, we did hear that some recruits give up after reaching the level required for the job they wanted. This is perhaps another reason that telling recruits the requirement for their ideal job before they take the test might not be wise.

One aspect of the test that was consistent for everyone we interviewed was what counted as a lift. If recruits make it to the line or fully extend their arms, the lift counts. Although this does seem sensible, it is worth noting that the shorter a recruit is, the harder he or she has to work harder to reach the line. To illustrate, a recruit who is 6' 1" tall only has to lift to the top of his or her head. In contrast, we watched a female recruit who was 4' 11" take the test. She had to fully extend her arms and rise up on her toes and still was barely able to get to the line. Whether this offers an unfair advantage to some test takers remains to be seen. Certainly, lifts to a given

height should be tied to the job—as seen on the survey in Chapter 5, among the jobs we examined, lifts at chest height are common while overhead lifts or other activities are less so.

Interviews with Recruits

We individually interviewed 30 recruits after they completed the SAT to better understand their attitude toward the test (Table 3.3). For example, when asked: "Overall, what do you think about the strength test?" a majority of participants responded with something positive or neutral, like "it was fun," or "it was fine." We also asked if the test was a good measure of their strength, and most recruits agreed that it was. Some of those saying yes added that they thought it was a good measure of upper body strength but not necessarily of endurance.

Table 3.3
Recruits' Answers to Key Interview Questions

Interview Question	Percent	Number
Do you think the SAT is a good measure of your strength? (Percent saying yes)	86%	29
Had you heard about the strength test before arriving at MEPS? (Percent saying yes)	50%	30
Do you know how much weight you lifted on your last trial? (Percent saying yes)	69%	29
What were you told about the purpose of the SAT before you took the test today? (Percent saying it is used to qualify for certain jobs)	38%	26
Do you know how much weight you have to lift to qualify for your preferred Air Force job? (Percent saying yes)	11%	27

We also asked recruits what they knew about the test. As shown in Table 3.3, only half of the participants arrived knowing that they had to take a strength test, and most of them were unsure about what exactly they would have to do for it. Of those that did know about it in advance, some cited their recruiter as the source, while others said that their knowledge came from a friend or family member. Everyone we asked said that they had not attempted to prepare for the test, and a few mentioned that they regularly lift weights anyway. Four of the people who had not heard about the test said that if they had known, they would have tried to prepare for it by practicing or working out.

When asked how much weight they had lifted on their last successful trial, 31 percent had no idea. Of the 29 recruits who said that they knew how much they had lifted (see Table 3.3), three were incorrect about the amount. When asked what the test was used for, only 38 percent stated that it was used to qualify them for certain jobs. The rest seemed unaware of its purpose, other than that it was supposed to measure their strength. When asked if they knew how much they had to lift to qualify for the job that they wanted, only a few said that they did not know what job they wanted. Of the 27 people who did know what job they wanted, only three knew what the required score was.

Interviews with LNCOs

Average tenure of the 17 participating LNCOs is shown in Table 3.4. In the interviews, LNCOs were asked to describe from start to finish how they administer the SAT, including what is said to the recruit during the test, whether they demonstrate the procedures, how many recruits are typically watching while the test is administered, whether recruits are allowed to pause if they need a second to rest, etc.[23] Results for this part of the interview were described in the previous section. Other key findings from the interview are described briefly in this section.

Table 3.4
Average, Minimum, and Maximum Tenure as an LNCO and in the Air Force

	Average	Min	Max
Years in the Air Force	14	5	25
Years as an LNCO	2	.08 (1 month)	6

LNCO opinions about the usefulness of the test were mixed. Many felt it was useful for some jobs; however, this response was most typical of LNCOs who had held a job that required a lot of lifting. In those cases, several added that some people in their career field could not do the lifting and they were not sure how they had made it into the career field in the first place. Other LNCOs said they thought the test was a waste of time.

None of the LNCOs had ever seen someone fail to lift the 40-pound minimum. When asked the typical amount lifted by women and men, nearly all said 70 for women and cited a maximum for men (i.e., 100 or 110 depending on which weight they viewed as the maximum). Many of the LNCOs reported having worked previously as a recruiter. Many also expressed a belief that recruiters usually tell the recruits about the SAT so that they know what to expect when they arrive at the MEPS. This response is in stark contrast to the recruits who typically reported not having been told about the test by their recruiter. It is quite possible that many recruits are in fact told about the SAT in advance, but suffer information overload and promptly forget about it entirely. Regardless, it may be worth noting this inconsistency with the LNCOs and request that recruiters pay special attention to explaining the test when orienting recruits prior to the MEPS visit.

When we asked LNCOs where they learned to administer the test, a few pointed us to a set of written instructions in their official LNCO manuals, others mentioned the posters on the wall near the machine, and nearly all reported being trained by other LNCOs. None of the LNCOs cited section 6.21 of AFRS Instruction 36-2001 as their source for the proper procedure, although one was able to point to AFRS HQ Form 42, which provides procedural guidance consistent with the AFRS Instruction.

[23] See Appendix C for an extended list of interview questions.

Conclusions

From our MEPS observations and interviews, we have several notable findings and recommendations.

Incremental Lift Machines

First, some ILMs may be damaged, in need of parts, or need to be replaced. We also learned that some MEPS may have more than one ILM and any extras could be used as replacements for those that are damaged.

Recommendation: Based on this finding, we recommend that the SAT machines be inventoried on a regular basis (e.g., at least every few years) and damaged machines or those in need of repair be identified and either replaced or repaired.

SAT Administration

Second, we discovered that SAT administration varies from LNCO to LNCO and site to site in some meaningful ways. They also differ from the administration procedures recommended by McDaniel, Skandis, and Madole (1983). These differences in administration mean that one person's score on the SAT is not necessarily equivalent to someone else's. For example, a score of 100 could mean that a recruit cannot lift 110, or it could mean that he or she tested at a location that stops at 100 and never had a chance to lift 110. This variation adds error to the test that could make it less useful in identifying whether a recruit is or is not qualified for a particular job. To maximize the usefulness of the test, individuals should uniformly be allowed to lift to their maximal capacity, at least up to the limit of 110 lbs. Administering the test after the physical and administering it in groups, when necessary, makes some sense. However, whether doing so affects performance on the SAT needs to be investigated. Certainly, the test developers suggested that public administration should be avoided (McDaniel, Skandis, and Madole, 1983). Other sources of variation, such as starting someone at 70 pounds if he or she looks strong, also need to be eliminated. The protocol should explain that eight lifts in a row (40 through 110 pounds) is much harder to do than one 70-pound lift and one 100-pound lift. By skipping the intermediate lifts, LNCOs may be offering recruits who appear strong an unfair advantage over the rest who have to do all eight lifts. These are just a few examples of how additional standardization in the test administration is needed.

Recommendation: To eliminate these potential sources of error, we strongly suggest that new instructions be sent to all MEPS locations and that a standardized training procedure be developed for all LNCOs that outlines what is and is not allowed during administration. These instructions, for example, should address

- whether recruits should be told how much they are lifting before and during the test
- whether or not LNCOs should provide encouragement to recruits to try harder (given that encouragement may be variable, it is easier to implement a restriction on encouragement)
- whether or not recruits should be given an opportunity to try again after a break
- whether recruits are allowed to retest later that day, and if they do retest, whether or not they have to start again at 40 pounds.

Additionally, LNCOs should be retrained every few years and official administration protocols should be redistributed, to help ensure that the proper protocol is maintained over time; every few years, an audit of this implementation should be performed.

Recruits' Knowledge of the SAT

A third finding deals with whether or not recruits have prior knowledge of the test before arriving at the MEPS and the effect that this knowledge, or lack thereof, could have on test scores. While most LNCOs believe that recruiters inform recruits about the strength test, most recruits we interviewed said they had no prior knowledge of the SAT. Given that both general and specific workout regimens can potentially improve test scores (see, e.g., Knapik, 1997, and Sharp et al., 1993), it would only be fair to give the recruits as much advance warning as possible and to advise those who are not familiar with or good at overhead lifts to practice them at the gym so they are prepared when they arrive at the MEPS. This advice to practice would be particularly important for someone who has no experience lifting weights or using weight machines because proper form and technique may make a big difference in how hard it is to complete a lift. In addition, stressing that recruits have to wear tennis or running shoes and clothes in which that they can comfortably squat and lift weights is important, particularly since women sometimes wear shoes with heels, a skirt or low-cut jeans—all of which might hinder their performance on the test.

Recommendation: Ensure that recruiters pay special attention to explaining the SAT and its purpose when orienting recruits prior to MEPS visits and provide insight on the potential value of preparing for the test in advance. Consider creating a pamphlet for recruiters to hand out (perhaps something like the poster described above that is displayed at some of the MEPS) that recruits could refer to later to counter the "information overload" problem.

4. Strength Requirements Survey: Sample and Screener

This chapter describes the web-based survey developed by RAND for defining strength requirements in career fields and how the survey was administered, as well as the results for the item included as a screener to identify jobs with some minimal physical demands. The survey asked respondents in eight AFSs to describe aspects of the job's physical requirements that are vital for defining strength requirements. They are the following:

- **The *types* of physical actions (lifting, pushing, throwing, etc.).** Different actions require different types of strength. For example, lifting an object over one's head requires greater upper body strength than lifting an object from the floor up to a table.
- **The *level* of the action (i.e., how much *weight* is involved and what is the *duration* of the action).** The same action can have very different strength requirements depending on the weight of the object. Lifting a 10-pound object over one's head requires much less strength than lifting a 60-pound object over one's head. Similarly, lifting a 60-pound object into a truck once a day requires different physical ability factors than lifting 60-pound objects into a truck repeatedly for several hours.
- **The *frequency* and *importance* of the actions.** Those activities that occur frequently or are vital to successful performance are central to defining requirements even for the minimally competent person. In contrast, activities that occur rarely and are of little importance are less essential to defining the requirement.

The first step in establishing cut scores on any test involves clearly defining the requirements of the job. In the case of establishing requirements for strength testing, it is critical to have a solid understanding of the type of physically demanding tasks that are required on the job, as well as their importance to the job, the frequency with which they occur, and for how long the physical activity is sustained. This survey was designed specifically to address these key aspects of AFS-specific job demands. The results presented here are intended to illustrate the types of responses we obtained when testing the survey, and determine whether a survey like this one would be a viable alternative to the current method the Air Force is using for collecting information about physical job demands.

Overview of Survey Topics

The Strength Requirements Survey included the six topic areas shown in Table 4.1.[24] (See also Appendix D for a more extensive and consolidated tabular overview of survey content addressed throughout this report.) The survey began with demographic questions to identify paygrade, gender, duty AFS, time in that duty AFS, current skill level, height, and weight. A short screener tool followed these questions. The screener asked participants to check from a list

[24] The survey included other areas not covered in Table 4.1; however, due to resource constraints, we did not analyze those results.

of physical *actions* (e.g., lifting, pushing, pulling, throwing of objects, etc.) any that were required on their job. The purpose of the screener was to identify people who do not typically perform physical activities on their job and prevent them from having to answer any additional survey questions about physical activities when such activities were not applicable to them. One goal was to evaluate the success of this screener at screening out those personnel.

Table 4.1
Summary of Survey Topics and Purpose

Survey Topics	Purpose
Demographics	Identify any skill-level and gender differences.
Strength-requirements screener	Identify the basic type of *actions* (pull, push, lift, lower, carry, hold, throw, support one's body weight, rotate/swing), if any, that are required on the job. Evaluate the functioning of the screener tool.
Action weight/importance/frequency	Provide additional details about the *weight* of the objects involved in the actions, and the *importance* and *frequency* of the actions.
Movement type/duration	Identify *how* the action is performed (e.g., lifting overhead, lifting to chest height) and for what *duration* it is typically performed.
Other strength demands	Determine if the survey items have missed any important aspect of physical activity required on the job.
Final survey comments	

Those who reported at least one type of job-related physical activity on the screener were routed to the next section of the survey, which contained follow-up questions about the activities they reported. The follow up questions asked participants to identify the amount of *weight* involved in the activity, and the *frequency* and *importance* of the activity. After completing the follow-up section on the action weight, importance and frequency, respondents were routed to a second set of follow-up questions that asked them to describe *how the action was performed* (overhead, at waist level, at knee level, on the side with one hand, etc.) and the *duration* of the activity.

All participants (including those who were screened out of the follow-up sections) were routed to an open-ended question asking if there were other types of activities in their job that required physical strength that were not captured in their previous responses and, if so, to describe the activity in detail. This was included to determine whether the survey content was deficient in some way and, if so, what should be added in future versions. (Appendix E contains the findings for that section.) Lastly, we asked participants if they had any additional comments related to the strength requirements of their job.

Sample and Response Rates

We selected eight AFSs to survey (listed in Table 4.2). The AFSs were chosen to cover a range of SAT cutoff scores to allow us to compare career fields with low and high SAT requirements. We also planned to examine differences in physical job requirements by skill level, gender, and AFS. However, for all specialties (except Security Forces), the sub-populations—particularly at the seven and nine skill levels—were small. For this reason, for all career fields except Security Forces we invited the entire population at the three, five, seven, and nine skill levels to participate. Because Security Forces is so large, we opted to select a stratified random sample for some subgroups rather than take a census, and we apply statistical weights to correct for under- or oversampling in any Security Forces analyses that are not broken out by subgroups. Table 4.2 shows the total number of women in each AFS and the total population size by AFS. See Appendix F for further detail on the populations and response rates broken out by gender, skill level, and AFS; and for further explanation of the stratified sampling procedure and the statistical weights.

For simplicity in discussing the results, we have opted to shorten the names of three of the AFSs in Table 4.2 as follows:

- A-10, F-15 & U-2 Avionics Systems will be referred to as *Avionics Systems*
- Manned Aircraft Maintenance–Aircraft Fuel Systems specialty will be referred to as *Aircraft Fuel Systems*
- Aerospace Propulsion–Turboprop/Turboshaft will be referred to as *Aerospace Propulsion.*

In addition, in tables where space is limited, we use the acronyms provided in Table 4.2.

Table 4.2
Population Sizes and SAT Cut Scores for the Air Force Specialties We Surveyed

Air Force Specialty	Air Force Specialty Code	SAT Cut Score	Total Women	Population Total
A-10, F-15 & U-2 Avionics Systems (AFU-AS)	2A3X1	80	57	1,406
Explosive Ordnance Disposal (EOD)	3E8X1	80	64	1,153
Security Forces (SF)	3P0X1	70	4,132	26,202[a]
Aircrew Flight Equipment (AFE)	1P0X1	70	446	2,467
Aerospace Propulsion–Turboprop and Turboshaft (AP-TTP)	2A6X1B	60	22	331
Manned Aircraft Maintenance–Aircraft Fuel Systems (MAFS)	2A6X4	60	128	1,893
Cyber Surety (CS)	3D0X3	40	299	1,175
Surgical Service (SS)	4N1X1	40	354	705

NOTE: See Appendix F for population sizes by skill level.
[a] The sample for Security Forces was significantly smaller than the total population size. Only about 6,000 of the 26,000 members of this AFS were invited to complete the survey.

We emailed invitations to the survey to 14,707 active duty Air Force enlisted personnel. As anticipated, about 9 percent were returned with delivery errors, resulting in a total of approximately 13,400 valid invitations. Of the approximately 13,400 personnel with valid email addresses, 23 percent (3,099 airmen) logged into the survey and 12 percent (1,580) entered and reached the last page of the survey.[25] Table 4.3 shows the number of respondents at various points in the survey process. Of those who reached the end of the survey, the average total time spent was 17.40 minutes (standard deviation = 15.28 minutes).

Table 4.3
Participants Who Viewed Different Sections of the Strength Requirements Survey

Survey Page	Number of Participants Remaining in the Survey [a]	Number That Did Not Continue Any Further	Number of Responses
Informed consent	3,099	31	–
Demographics	3,068	32	3,028
Strength-requirements screener	3,036	336	2,936
Action weight/importance/frequency	2,700	751	2,381 [b]
Movement type	1,949	255	1,510 [b]
Other job demands	1,694	114	398 [c]
Equipment/assistance to reduce physical job demands and final survey comment	1,580	–	1,063

[a] People were counted as remaining participants if they viewed the page in question or if they were branched to a later page, even if branching prevented them from viewing the page in question.
[b] Only a subset of people were branched to these sections; hence, the number of responses is smaller than the number of remaining participants.
[c] This page included two write-in response items seen by all 1,694 remaining survey participants. However, only a subset of people (398) chose to write in a response.

We also evaluated whether there were large differences in self-reported background characteristics and the background characteristics in Air Force personnel files. We found few incongruities between the self-report and personnel-file versions of gender, paygrade, and AFS (i.e., over 89 percent of participants had matches on all three). However, for about 44 percent of participants, self-reported skill levels differed from the skill level on record in their personnel file at the time of the survey. A vast majority of the differences were one skill level higher in the

[25] Due to unexpectedly low response rates to our survey, we sent three waves of reminder emails to those who had not yet responded. In addition, we consulted experts in the AFPC and the Air Force Management Agency (AFMA) who cited survey fatigue, computer server firewalls, and ongoing efforts by leadership to prevent personnel from clicking on dot-com web links because of data security concerns. In an attempt to boost our response rates, we sent one reminder email to "non-completers" (i.e., those entered the survey but did not complete it); career field managers also sent out notices encouraging participation within their career field.

self-report (e.g., three versus five level).[26] Given the discrepancy, we opted to use the information we had available from personnel records rather than the self-report data.

Strength-Requirements Screener

The Strength-Requirements Screener presented respondents with a list of nine actions—support your body, rotate/swing, push/press, pull, carry, hold, lift, lower, and throw/toss—and asked them to check all actions that are required on their job.[27] These actions are consistent with those used in past research (e.g., Ayoub et al., 1987), and were intended to encompass all possible strength-related activities on the job. The screener items are shown in Table 4.4.

The screener served two purposes. First, if the screener is shown to be effective at distinguishing AFSs that have low strength requirements from those that have high strength requirements, we would suggest that it be administered to all AFSs as part of their regular occupational analysis survey.[28] The screener could be used to flag any AFSs whose strength requirements appear to have changed from previous years. Such flagged AFSs would then receive a set of in-depth follow-up questions (such as those described in the next sections) to further evaluate whether a change in the SAT cut point is needed.

Second, using a screener in an online survey can reduce survey burden by allowing for conditional skip-logic. If the screener is successful, it will limit the number of questions seen by participants both in our study and in any future operational surveys using the tool (such as in the occupational analysis survey). Therefore, to reduce burden in the administration of our survey, people only received follow-up questions about actions they checked on the screener.[29]

Table 4.4
The Strength-Requirements Screener

Please indicate whether your job (i.e., your current duty AFSC) REQUIRES the following types of activities. (Check all that apply.)	
☐	**SUPPORTING YOUR BODY** in positions other than normal sitting, standing, or walking. By supporting your body, we mean using your physical strength to support your own body weight in positions other than normal sitting, standing, or walking. Examples include propping yourself up with one arm to drill something with another arm and squatting to access a panel on the underside of a plane.

[26] We examined the relationship between skill-level mismatches and such background characteristics as paygrade; however that failed to explain the mismatch. One plausible explanation for the mismatches is that participants misunderstood the skill-level question, which asked about the skill level associated with one's current *duty* AFS. Control and primary AFSs can also have skill levels attached to them.

[27] From this point forward, we will refer to "rotate/swing" as "rotate," "push/press" as "push," "throw/toss" as "throw," and "support your body" as "support body."

[28] As a reminder, the occupational analysis survey is administered to all enlisted AFSs every three years. OAD is responsible for administration and analysis of the occupational analysis survey.

[29] Follow-up questions are discussed in the next section.

Please indicate whether your job (i.e., your current duty AFSC) REQUIRES the following types of activities.
(Check all that apply.)

- ☐ Continuously or repeatedly **ROTATING or SWINGING** objects or sets of materials of any weight with your hands.
  By rotating or swinging, we mean using your hands and fingers to continuously or repeatedly manipulate objects in a curved pattern. Examples include turning wheels or levers and swinging a hammer several times in a row. This category does NOT include the other actions on this page, even though rotating or swinging objects may be needed to do the other actions (e.g., swinging a line of cable to then throw it).
- ☐ **PUSHING/PRESSING** objects weighing 10 lbs. or more.
  By pushing/pressing, we mean using your hands and/or arms to move objects forward while you either stay in place (e.g., stand) or move your lower body (e.g., walk). Examples include pushing windows closed, pushing a box across the floor, and pressing your hands against a door to keep it from opening.
- ☐ **PULLING** objects weighing 10 lbs. or more.
  By pulling, we mean holding onto an object with your hands to move the object toward you while you either stay in place (e.g., stand) or move with your lower body (e.g., walk). Examples include pulling a door closed, dragging a box across the floor, and dragging a line of cable or a hose.
- ☐ **CARRYING** objects weighing 10 lbs. or more.
  By carrying, we mean holding objects in your arms, hands, or on your back while you move with your lower body (e.g., run). Examples include walking with a box in your arms, running with a backpack on your back, and holding a toolbox at your side while walking. This category does NOT include lifting or lowering objects, even though lifting or lowering is often required to carry objects (e.g., lifting a box off a table to then carry it across a room).
- ☐ **HOLDING** objects weighing 10 lbs. or more.
  By holding, we mean using your upper-body strength to maintain objects in your arms, hands, or on your back while you stay in place (e.g., stand). Examples include sitting with a box in your arms without the box resting on your lap and holding a toolbox at your side while standing in place. This category does NOT include lifting or lowering objects, even though lifting or lowering is often required to hold objects (e.g., lowering a box off a shelf to then hold it).
- ☐ **LIFTING** objects weighing 10 lbs. or more.
  By lifting, we mean using your hands and/or arms to move an object in an upward direction. Examples include moving a box from a lower shelf to a higher shelf and picking up a toolbox off the floor to put it on a table.
- ☐ **LOWERING** objects weighing 10 lbs. or more.
  By lowering, we mean using your hands and/or arms to move an object in a downward direction. Examples include moving a box from a higher shelf to a lower shelf and taking a toolbox off a table to put it on the floor.
- ☐ **THROWING/TOSSING** objects weighing 10 lbs. or more.
  By throwing/tossing, we mean thrusting or propelling an object out of your hands and/or arms, while you either stay in place (e.g., stand) or move with your lower body (e.g., walk). Examples include throwing a line of cable across a room and throwing sand bags into the bed of a truck.
- ☐ My job does not require me to do any of these types of activities.

COMMON OBJECTS that weigh approximately 10 lbs:
metal folding chair
full-sized ironing board
standard two-by-four (approx. 2-in deep, 4-in wide, and 8-ft long; made of pine wood)

Major Findings

Most respondents selected at least one action on the checklist, even in the low-strength career fields. As a result, most were routed to complete at least one set of follow-up questions. In this way, the screener was not successful at screening out those least likely to be engaging in strength-related activities.

The percentage of respondents selecting each action is shown in Table 4.5. Actions are rank-ordered in the table, with the most frequently endorsed action at the top. AFSs are grouped according to their current cut score on the SAT. As shown in the table, Cyber Surety had the largest proportion of people selecting "none required" of all of the AFSs, followed by Surgical Service. This is consistent with our assumption that AFSs with 40-pound cut scores would have fewer physical requirements than AFSs with higher SAT cut scores. In addition, the percentages were smaller for Cyber Surety on any of the specific actions relative to other AFSs.

Table 4.5
Rankings of Strength-Requirements Screener Items Based on Frequency of Endorsement

SAT = 40				**SAT = 60**			
Cyber Surety (sample = 275)		**Surgical Service (sample = 141)**		**Aerospace Propulsion-TTP (sample = 78)**		**Aircraft Fuel Systems (sample = 498)**	
Carry	55%	Lift	76%	Carry	97%	Lift	92%
Lift	53%	Carry	73%	Lift	97%	Support Body	91%
Lower	44%	Push	70%	Lower	96%	Carry	90%
Hold	37%	Pull	67%	Push	96%	Push	89%
Pull	32%	Lower	67%	Pull	95%	Lower	88%
Push	31%	Hold	58%	Hold	91%	Pull	87%
Support Body	13%	Support Body	50%	Support Body	91%	Hold	84%
Throw	9%	Rotate	38%	Rotate	77%	Rotate	69%
Rotate	9%	Throw	6%	Throw	37%	Throw	25%
None required	35%	None required	11%	None required	1%	None required	4%
SAT = 70				**SAT = 80**			
Aircrew Flight Equipment (sample = 652)		**Security Forces (sample = 710)**		**Avionics Systems (sample = 350)**		**Explosive Ordnance Disposal (sample = 308)**	
Carry	88%	Carry	85%	Carry	94%	Carry	99%
Lift	87%	Lift	76%	Lift	94%	Lift	99%
Lower	80%	Hold	72%	Lower	94%	Lower	97%
Pull	71%	Lower	63%	Pull	93%	Hold	95%
Push	71%	Support Body	57%	Push	93%	Pull	95%
Hold	70%	Pull	56%	Hold	92%	Push	93%
Support Body	52%	Push	56%	Support Body	90%	Support Body	92%
Rotate	47%	Rotate	38%	Rotate	69%	Throw	83%
Throw	35%	Throw	36%	Throw	29%	Rotate	67%
None required	5%	None required	7%	None required	3%	None required	1%

NOTE: Some percentages within an AFS are identical because of rounding error. Percentages within columns do not add to 100% because respondents could select more than one option on the checklist. "None required" refers to the item that reads, "My job does not require me to do any of these types of activities."

Surgical Service, in contrast, did not have noticeably smaller percentages for the various actions relative to some of the AFSs with higher existing cut scores. This may suggest that the Surgical Service specialty has greater strength requirements than we originally anticipated, given its low SAT cut score. Identifying such a discrepancy could be one step toward flagging AFSs that need further investigation.

Overall, the results of the screener suggest that setting a higher threshold on the screener, such as 25 lbs. rather than 10 lbs., may be necessary to ensure that those in career fields with low physical demands are not unduly burdened by being required to complete a more in-depth survey of their physical skills. However, the results also show that a screener could be a successful tool for distinguishing those career fields that have physical demands from those that do not. Our results also demonstrate that, at least for these AFSs, carrying and lifting items are quite common. Subsequent sections of the survey examine the frequency and importance of these actions, as well as the location the activities are performed.

5. Survey Results: Actions and Movement Type

This chapter presents the results obtained from two sections of the survey: the Action Section and the Movement Type Section. For each we describe the survey questions particular to that part of the survey and present a portion of the results. Because of the large amount of data collected, it is not feasible to present all available results, but we do provide a sample that is illustrative of the value of the survey as well as areas that need further modification.[30] Overall, we find that the survey could become a viable tool in defining strength requirements of relevant career fields. As shown in the Action section of the survey, strength requirements in jobs considered differentially demanding based on the SAT cut score did indeed differ in terms of frequency of physical task performance as well as in the perceived importance of those physical tasks. Moreover, airmen in jobs classified as having higher demands by the SAT cut score process who engaged in these physical activities were more likely to report that they did so under awkward conditions or in head-height locations in the Movement Type Section. While these findings lend credence to the current SAT cut scores, the results also illustrate the usefulness of these data for delineating job demands within Air Force career fields.

Action Section

We begin with an overview of the Action Section of the survey. As noted in Chapter 2, it is common to ascertain physical job demands via survey items that describe the type of movement or physical task, and quantify the perceived demand with various follow-ups. These sections of the survey did so. (For a consolidated view of the survey sections discussed in the previous chapter and below, see Appendix D.)

Survey Questions

For each action selected on the screener, participants were asked to identify the *weight* of the objects involved with the action, the *importance* of the action for their job, and the *frequency* with which the required action occurs. Exact wording of the questions is shown in Figure 5.1. Note that both aspects are important in determining minimum requirements: Frequent tasks impose frequent physical demands, while important tasks may be those that are key to the job itself. Even a task with very low frequency may be a key physical demand because its performance is essential to performing the job.

---

[30] Given that there were many data elements collected on the survey (144 total in the Action Section alone), the volume of available data is large. Similarly, tables of summary statics that result from our analyses are large as well. As a result, we opted to provide only a snapshot of the results here, for purposes of illustrating the usefulness of the survey. For use in the snapshots, we selected three actions—rotate, carry, and push—because they collectively cover the four types of questions in the Action Section and represent actions that had disparate endorsement rates on the Strength-Requirements Screener. Tables showing results for all actions and all AFSs are available in a separate unpublished report.

Figure 5.1
Action Section—Question Formats

Support Your Body and Rotate

To save time, you can skip any rows that are not applicable to you. You do not have to answer each question to move forward with the survey.

Please indicate:

a) How OFTEN your job requires the following activities

b) How IMPORTANT it is that you perform the following activities for your job

c) For about HOW LONG you typically perform the following activities for your job without taking a break.

CLICK HERE for Definitions of SUPPORT YOUR BODY and ROTATE or SWING and for common objects of different weights.

SUPPORT YOUR BODY in positions other than normal sitting, standing, or walking (for example, squat to access a panel on the underside of a plane).

	REQUIRED how often?	How important?	For how long?
	--	--	--

Continuously or repeatedly ROTATE or SWING an object or sets of materials with your hands (for example, swing a hammer several times in a row) of the following weights:

	REQUIRED how often?	How important?	For how long?
less than 5 lbs	--	--	--
5 to 9 lbs	--	--	--
10 to 24 lbs	--	--	--
25-39 lbs	--	--	--
40-69 lbs	--	--	--
70 lbs or more	--	--	--

Action Section – Question Format for Push

PUSH/PRESS an object weighing approximately

	REQUIRED how often?	How important?	WITHOUT HELP from carts, dollies, etc.?
10-24 lbs	--	--	--
25-39 lbs	--	--	--
40-69 lbs	--	--	--
70-99 lbs	--	--	--
100-199 lbs	--	--	--
200 lbs or more	--	--	--

NOTE: Question format is the same for push and pull.

Action Section – Question Format for Carry

CARRY an object weighing approximately

	REQUIRED how often?	How important?
10-24 lbs	--	--
25-39 lbs	--	--
40-69 lbs	--	--
70-99 lbs	--	--
100-199 lbs	--	--
200 lbs or more	--	--

NOTE: Question format is the same for carry, hold, lift, lower, and throw.

Because we expected that most activities would involve objects weighing less than 100 pounds, we used narrow weight intervals below 100 pounds. For rotate, we used smaller weight categories to account for repetitive activities such as swinging a hammer.

Four of the nine actions also included a third column for responses. For support body and for each weight category of rotate, participants were asked how long they engage in the action at any one time (duration). For push and pull, participants were asked how often they are required not to use mechanical devices such as carts to perform the action at the given weight (no assistance), because many high-weight pushing or pulling activities involve the use of carts, dollies, and other conveyances to push or pull objects. For carry, hold, lift, lower, or throw objects, we instructed participants not to respond about actions for activities that involve mechanical assistance. The response options for the frequency, importance, duration, and no-assistance questions are in Table 5.1.

Table 5.1
Response Options for Action Section Questions

Survey Question	Data Value	Response Options
Frequency:	1	Never
How frequently does your job require it?	2	Once in 1 to 2 years
	3	2 to 4 times a year
	4	Once or twice a month
	5	Once or twice a week
	6	Once or twice a day
	7	Once an hour
	8	Several times an hour
Duration:	1	5 minutes or less
For how long without taking a break?	2	6 to 10 minutes
	3	11 to 30 minutes
	4	31 minutes to 1 hour
	5	2 to 4 hours
	6	5 to 8 hours
	7	More than 8 hours
Importance:	1	Not at all important
How important is it?	2	Slightly important
	3	Moderately important
	4	Very important
	5	Extremely important
No assistance:	1	Never
How often without assistance from carts,	2	Sometimes
dollies, and other conveyances?	3	Always

To help respondents estimate the weights of the objects involved in their work activates, we provided a list of common objects belonging to each of the weight categories (Table 5.2).

Table 5.2
Object Weight Examples

Weight	Common Objects with Approximate Weights
Less than 5 pounds	a hammer with a 12-in wood handle (1 lb) an average clothes iron (3–5 lbs) an average bathroom scale (3–5 lbs)
5 to 9 pounds	a small, table-top ironing board (5–8 lbs) a cordless, 12-volt power drill for home use (5–9 lbs)
10 to 24 pounds	a metal folding chair (10 lbs) a full-sized ironing board (10 lbs) a standard two-by-four (approx. 2 inches deep, 4 inches wide, and 8 feet long; made of pine wood) (10 lbs) a cordless, 18-volt power drill for commercial use (10-12 lbs) a standard, adult-sized bowling ball (12–16 lbs) one passenger car tire, inflated (20 lbs) a 32-inch LCD flat-screen TV (18–25 lbs)
25 to 39 pounds	an average two-year-old child (25 lbs) three metal folding chairs (30 lbs) one mid-sized microwave (35 lbs) a full propane tank for a gas grill (38 lbs)
40 to 69 pounds	a five-gallon plastic water cooler jug filled with water (40 lbs) a small bag of cement mix (50 lbs) a mini window air conditioning unit (40–60 lbs) two large bags of dry dog food (60–69 lbs)
70 to 99 pounds	a punching bag (70–80 lbs) two five-gallon plastic water cooler jugs filled with water (80 lbs) a large bag of cement mix (80–90 lbs) three standard (8 inch by 8 inch by 16 inch) cinder blocks (90–100 lbs)
100 to 199 pounds	a large-sized, adult, male dog, such as a rottweiler or bloodhound (100–130 lbs) a standard, top-loading clothes washing machine (140–150 lbs) an average, adult, American woman (140–160 lbs) an average, adult, American man (170–190 lbs) an average, freestanding kitchen range and oven (185–200 lbs)
200 pounds or more	seven standard (8 inch by 8 inch by 16 inch) cinder blocks (200–230 lbs) two large-sized, adult, male dogs such as rottweilers or bloodhounds (200–260 lbs) an average NFL linebacker (230–270 lbs)

To reduce survey burden, conditional logic was used to branch participants to relevant questions. In the Action Section, participants only received questions for the actions they selected on the screener. Participants who did not select any actions on the screener or only selected "My job does not require me to do any of these types of activities" were branched to the Other Job Demands Section of the survey. (See Appendix E for the results from the Other Job Demands Section.)

Participants who completed questions in the Action Section were eligible to branch to the Movement Types Section (discussed later in this chapter), which included more-detailed questions about the types of movements or positions (e.g., above head) used during physical activities. To reduce survey burden, we limited the number of follow-on Action Section questions sets to only three weight categories per action.[31] The selection of the three weight categories for use in follow-on questions was based on a four-step process. A weight category was only considered if the respondent indicated that he/she does the action at least once or twice per month (for frequency question) or that the action is moderately important to his/her job (for importance question). We do not go into detail here about the process of selecting weight categories; interested readers can find a more detailed description of the process in Appendix F.

Data Adjustments

Before calculating average ratings for each question, we removed respondents who provided inconsistent or questionable responses in the Action Section. These respondents fell into two categories: (1) respondents who were inconsistent between the Strength Requirements Screener and Action Section and (2) respondents who had inconsistent responses within the Action Section. The first type of respondent checked a particular action on the checklist, went forward and completed some questions for that action in the Action Section, but later returned to the Strength-Requirements Screener and unchecked the action. This type of respondent was removed from all analyses for the particular action on which they were inconsistent. The second type of respondent provided inconsistent responses within rows of questions in the Action Section. This type of inconsistency refers to how respondents answered questions within the same row corresponding to a particular *action* and *weight category* (e.g., carry 10–24-pound objects). For any such row, a respondent was considered inconsistent if he/she reported one or both of the following:

- Frequency higher than "Never" but importance and/or duration equal to "Not Applicable"
- Frequency equal to "Never" or left blank (missing) but importance higher than "Not at all important" and/or duration higher than or equal to "5 min or less."

Out of the 2,700 respondents who were branched to the Action Section, a total of 275 respondents had at least one inconsistent row in the Action Section. However, 135 of the 275 respondents (about 49 percent) only had one inconsistent row in the entire section. Of the 135

---

[31] We did not include rotate/swing in the Movement Type Section of the survey because we expected that body locations and positions used to rotate or swing objects would not be critical for determining physical strength requirements.

respondents with only one inconsistent row, 133 of them completed six or more rows of questions in the Action Section. Because so many of the respondents who provided inconsistent responses to one row of questions did not do so for many other rows of questions, we assumed that the one inconsistent row reflected a mistake, not a misunderstanding of the questions or some other systematic error. As such, we decided to retain these respondents in our analyses. However, we removed respondents who had two or more inconsistent rows.

Frequency Ratings for Action

We calculated average frequency ratings two ways.[32] First, we computed average frequency only for respondents who selected the action on the Strength-Requirements Screener (e.g., carry). Second, we expanded the results to include people who did not select the action on the Strength-Requirements Screener. Those who did not select that action on the screener were assigned a frequency response of "Never." This expanded analysis provides a more accurate estimate of the frequency with which the specialty as a whole performs the action. These two frequency ratings for three of the actions are shown in Tables 5.3 and 5.4. We shaded ratings to highlight different ranges of frequency responses. Recall that the current requirements algorithm for the SAT favors frequency in terms of both frequency of occurrence, and percent of people performing in a given career field (i.e., proportion of people performing).

The most obvious difference between Tables 5.3 and 5.4 is that Table 5.3 has higher average frequency ratings than Table 5.4, as reflected by the larger number and types of shaded cells in Table 5.3. This should not be surprising because the additional people included in the analyses reflected in Table 5.4 were coded at the bottom of the frequency scale (i.e., "Never"), which lowers the averages. If we use the average ratings in Table 5.3, we would conclude that there are many actions that airmen perform on a monthly, weekly, or even daily basis. However, if we use the average ratings in Table 5.4, we would conclude that, at most, airmen are required to perform a handful of physical activities on a monthly basis. Both types of information are useful because the information in Table 5.3 tells us how often airmen who perform some minimum level of a given action are required to perform that action, whereas the information in Table 5.4 tells us the AFS base rates of particular actions at given weight categories.

As seen in both Tables 5.3 and 5.4, specialties with higher SAT cut scores generally had higher average frequency ratings than specialties with lower SAT cut scores. For example, the Cyber Surety specialty only had two average frequencies above 4.00 (monthly) in Table 5.3, compared to 11 for each of the two AFSs with 80-lb SAT cut scores (Avionics Systems and Explosive Ordnance Disposal). One deviation from the trend of increasing frequency rates with increasing SAT cut scores concerns Aerospace Propulsion, which has a 60-lb SAT cut score. This AFS has somewhat higher average frequencies than the specialties with a 70-lb SAT cut

---

[32] In both instances, we recorded blank responses as "Never" for respondents who provided at least one response to a question for a particular action. Directions in the Action Section stated that respondents could skip any row of questions that did not pertain to their jobs. Thus, respondents who left a frequency question blank for a given weight category (e.g., the respondent responded about pushing 10–24-pound objects but did not respond for pushing 100–199-pound objects), were assigned scores of 1 for "Never."

score. Skill levels differences by AFS sample might be a factor: Respondents in the Aerospace Propulsion specialty were all 3-level personnel, making the sample the most junior of all the survey samples.[33] Because more-junior personnel might have more physical demands than personnel at higher skill levels, the higher average frequency rates for the Aerospace Propulsion specialty might relate to skill level.

Table 5.3
Frequency for Rotate, Push, and Carry:
Only Those Checking the Action on the Screener

Action	Weight Category (pounds)	SAT=40		SAT=60		SAT=70		SAT=80	
		CS	SS	AP-TTP	MAFS	AFE	SF	AFU-AS	EOD
Rotate	< 5	3.94	6.26	6.15	5.81	5.49	5.43	6.25	5.37
	5-9	3.89	4.57	5.26	4.64	4.35	5.29	4.62	4.82
	10-24	2.56	4.12	4.45	3.84	4.00	4.87	4.18	4.42
	25-39	2.11	2.95	3.77	2.99	3.81	3.68	3.84	3.97
	40-69	2.06	2.24	3.04	2.53	3.41	3.66	3.60	3.64
	70+	1.33	1.95	2.94	2.25	2.50	2.80	3.28	3.36
Sample Size		18	42	47	272	234	188	199	153
Push	10-24	4.32	5.32	4.91	4.71	4.64	4.85	5.22	5.01
	25-39	3.12	3.68	4.07	3.96	3.97	3.64	4.88	4.81
	40-69	2.85	2.90	3.83	3.38	3.55	3.24	5.13	4.60
	70-99	2.00	2.33	3.50	3.06	2.59	2.89	4.39	4.11
	100-199	1.70	2.59	3.00	2.78	2.01	2.25	3.84	3.51
	200+	1.55	2.00	3.67	2.91	1.57	2.10	3.20	2.93
Sample Size		66	73	54	316	314	258	250	182
Carry	10-24	4.48	5.89	5.71	5.26	4.83	5.25	5.87	5.09
	25-39	2.97	3.87	5.40	4.33	4.37	3.87	5.67	4.96
	40-69	2.33	2.37	4.38	3.79	3.91	3.36	5.59	4.81
	70-99	1.47	1.53	3.20	2.80	2.45	2.31	4.69	4.34
	100-199	1.22	1.54	2.18	2.03	1.80	1.57	3.66	3.11
	200+	1.20	1.31	1.69	1.62	1.38	1.32	2.32	2.24
Sample Size		119	70	55	309	388	430	242	193

Weekly frequency or higher (4.5 to 8.0)	Monthly frequency (3.5 to 4.4)	Yearly frequency (2.5 to 3.4)	Never to Once in 1 to 2 years (1.0 to 2.4)

[33] As explained in the last chapter, the Aerospace Propulsion (2A6X1b) specialty is actually a "shred" (i.e., a subspecialty) that is only open to personnel at the one or three skill level. We did not survey any personnel at the one skill level because they are in training. Therefore, all respondents for this AFS were at the three skill level.

Table 5.4
Frequency for Rotate, Push, and Carry:
Including Those Who Did Not Check the Action on the Screener

Action	Weight Category (pounds)	SAT=40 CS	SAT=40 SS	SAT=60 AP-TTP	SAT=60 MAFS	SAT=70 AFE	SAT=70 SF	SAT=80 AFU-AS	SAT=80 EOD
Rotate	< 5	1.20	2.71	4.41	3.87	2.79	2.32	4.20	3.28
	5-9	1.20	2.16	3.82	3.18	2.33	2.28	3.20	2.99
	10-24	1.11	2.02	3.28	2.70	2.19	2.15	2.94	2.78
	25-39	1.08	1.64	2.83	2.19	2.12	1.80	2.73	2.54
	40-69	1.07	1.40	2.35	1.91	1.96	1.79	2.58	2.37
	70+	1.02	1.31	2.28	1.75	1.60	1.54	2.39	2.23
Sample Size		259	129	71	455	588	683-684	326-327	294
Push	10-24	1.85	3.48	3.97	3.57	2.95	2.25	4.24	3.48
	25-39	1.54	2.53	3.34	3.05	2.59	1.86	3.98	3.36
	40-69	1.47	2.08	3.15	2.65	2.36	1.73	4.18	3.23
	70-99	1.26	1.75	2.90	2.43	1.85	1.61	3.61	2.93
	100-199	1.18	1.90	2.52	2.24	1.54	1.41	3.18	2.55
	200+	1.14	1.57	3.03	2.33	1.31	1.36	2.68	2.19
Sample Size		258	127-129	71	455	587	683-684	325-326	294
Carry	10-24	2.60	3.67	4.65	3.90	3.53	3.51	4.60	3.69
	25-39	1.91	2.57	4.41	3.27	3.22	2.69	4.46	3.60
	40-69	1.61	1.74	3.62	2.90	2.92	2.39	4.39	3.50
	70-99	1.22	1.29	2.70	2.23	1.96	1.77	3.73	3.19
	100-199	1.10	1.29	1.92	1.70	1.53	1.34	2.97	2.39
	200+	1.09	1.17	1.54	1.42	1.25	1.19	1.98	1.82
Sample Size		258-259	128-129	71	454	588	683-684	327	294

Weekly frequency or higher (4.5 to 8.0)	Monthly frequency (3.5 to 4.4)	Yearly frequency (2.5 to 3.4)	Never to Once in 1 to 2 years (1.0 to 2.4)

Another trend in Tables 5.3 and 5.4 is that push and carry actions had higher average frequency ratings than rotate/swing for corresponding weight categories. One possible reason for this trend is that many objects that are rotated or swung (e.g., hammers) are of lower weight than objects that typically are pushed or carried. For example, in Table 5.3, none of the average frequency rates for the 25-39-pound weight category is higher than 4.00 (monthly) for rotate/swing but three are higher than 4.00 for push and five are higher than 4.00 for carry. Indeed, for rotating or swinging objects, the only average frequency rates that are 4.00 or greater are for the three lowest weight categories—5 pounds or less, 5–9 pounds, and 10–24 pounds.

Importance, Duration, and No-Assistance Ratings for Action

Our next set of analyses for the Action Section examined the average *importance*, *duration*, and *no-assistance* ratings for each AFS and SAT cut score. Unlike the analysis for the frequency ratings, our analyses for the other types of ratings focused largely on only those respondents who had selected that action on the screener. We also excluded from the analyses those people who indicated that the weight category was “not applicable” to their job. This resulted in a wide range of sample sizes for each question. It also resulted in meaningful differences in the sample sizes

across AFSs.[34] Results for importance ratings are shown in Tables 5.5 and 5.6. Results for average duration ratings and average no-assistance ratings are shown in Tables 5.7 and 5.8.[35]

The results in Table 5.5 show little differentiation in importance ratings by AFS, as most of the importance ratings did not exceed 4.00 ("Very important"). That is, respondents who selected these actions generally felt the physical activities they do are only slightly or moderately important. The one exception is for the Avionics Systems (AFU-AS) specialty, for which most of the ratings averaged around 4.00. This specialty also had higher frequency ratings than most AFSs, including the other 80-pound AFS, Explosive Ordnance Disposal. Note that for setting strength requirements, algorithms vary by job in order to be able to distinguish differences between employees. When most actions are at least minimally important, it is likely that only those task demands that are on average "very important" or higher would be considered as key in setting job requirements for the minimally acceptable person, unless the frequency of performance is high.

Although examination of the importance ratings in Table 5.5 appears to suggest no large differences by AFS, further examination of the data shows that the overall *percentage* of people indicating that the action for a given weight was "moderately important" or higher, shows some meaningful and large differences between the career fields.[36] More precisely, when the people who did not select that action on the screener are taken into consideration in the analyses, the differences are clear. For example, as shown in Table 5.6, between 2 and 6 percent of the Cyber Surety respondents who were routed to the Action Section selected "moderately important" or higher. In contrast, the percentage of Explosive Ordinance Disposal respondents that indicated the action was "moderately important" or higher, was much larger, ranging from 39 to 45 percent. This suggests that those in physically demanding jobs consider these demanding tasks to be a more important part of the job, as would be expected.

---

[34] We caution readers to keep the sample size differences in mind when viewing the remaining results. While some of the average ratings may not differ much across AFSs, the proportion that engages in the activity does. When establishing a physical strength requirement for an AFS, the differences in proportion engaging in the activity should be factored into the interpretation of the average importance ratings, duration ratings and no-assistance ratings. To help illustrate this point, we report average importance ratings in Table 5.5 along with the proportion reporting moderate or higher importance (including those who did not select the action on the screener) in Table 5.6. Note that for duration and no assistance, we do not report companion tables showing the results including those who did not check the action on the screener; nevertheless, the same caveats still apply.

[35] As a reminder to the reader, the question for no-assistance reads, "Please indicate…how often you are required to do this [push or pull] WITHOUT HELP from carts, dollies, hand trucks, or other mechanical devices?" The response scale for this question is 1 (Never), 2 (Sometimes), and 3 (Always). Therefore, higher scores for no-assistance indicate that responds are more frequently required to push or pull objects without help from mechanical devices.

[36] These percentages reflect the number that endorsed "moderately important" or higher out of the total number of people in the AFS who answered at least one question in the entire Action Section. For example, 118 EOD people endorsed "moderately important" or higher on rotate < 5 lbs. The 45 percent reported in Table 5.6 is calculated as (118/261)*100.

Table 5.5
Importance Ratings for Rotate, Push, and Carry:
Only Those Checking the Action on the Screener

Action	Weight Category (pounds)	SAT = 40 CS	SAT = 40 SS	SAT = 60 AP-TTP	SAT = 60 MAFS	SAT = 70 AFE	SAT = 70 SF	SAT = 80 AFU-AS	SAT = 80 EOD
Rotate	< 5	3.00	3.74	3.95	3.95	3.76	3.73	4.11	3.67
	5-9	3.42	3.81	3.79	3.76	3.70	3.71	3.97	3.71
	10-24	3.38	3.77	3.94	3.65	3.78	3.71	3.99	3.70
	25-39	3.43	3.56	3.80	3.58	3.80	3.69	4.06	3.82
	40-69	2.71	3.29	3.88	3.51	3.80	3.61	4.13	3.81
	70+	–	3.13	3.81	3.53	3.52	3.54	4.18	3.96
Sample Size		7-13	14-38	24-43	111-249	93-189	102-138	99-181	105-137
Push	10-24	3.26	3.66	3.83	3.68	3.73	3.46	4.04	3.79
	25-39	3.16	3.79	3.77	3.63	3.72	3.43	4.07	3.81
	40-69	3.09	3.76	3.83	3.59	3.67	3.26	4.07	3.78
	70-99	3.00	3.75	3.72	3.55	3.61	3.32	4.05	3.80
	100-199	3.04	3.82	3.60	3.48	3.74	3.22	3.99	3.82
	200+	2.74	3.80	3.97	3.69	3.68	3.37	4.04	3.63
Sample Size		19-53	20-59	30-41	157-248	60-227	83-195	128-206	126-158
Carry	10-24	3.32	3.82	3.92	3.78	3.82	4.07	4.22	3.94
	25-39	3.20	3.97	3.96	3.73	3.84	3.77	4.20	3.98
	40-69	3.05	3.74	3.89	3.65	3.80	3.80	4.14	3.98
	70-99	2.84	3.11	3.69	3.61	3.64	3.54	4.06	3.96
	100-199	2.87	3.50	3.55	3.45	3.62	3.19	4.08	3.99
	200+	2.79	3.50	3.62	3.47	3.55	2.95	4.14	3.77
Sample Size		14-104	6-62	13-48	76-269	56-284	70-325	93-214	102-167

Very important or higher (4.00 to 5.00)	Very Important (3.50 to 3.99)	Moderately important (3.00 to 3.49)	Slightly to moderately important (2.00 to 2.99)

NOTE: – indicates fewer than five respondents.

Table 5.6
Proportions for "Moderately Important" or Higher Ratings:
Including Those Who Did Not Check the Action on the Screener—CS and EOD Comparison

		SAT = 40	SAT = 80
Action	**Weight Category (pounds)**	**Cyber Surety (n = 151)**	**Explosive Ordinance Disposal (n = 261)**
Rotate	< 5	6%	45%
	5-9	7%	44%
	10-24	4%	45%
	25-39	4%	40%
	40-69	3%	39%
	70+	2%	39%
Push	10-24	27%	54%
	25-39	21%	55%
	40-69	20%	57%
	70-99	11%	55%
	100-199	11%	49%
	200+	7%	40%
Carry	10-24	52%	57%
	25-39	35%	59%
	40-69	24%	62%
	70-99	9%	60%
	100-199	5%	50%
	200+	5%	34%

Results for average duration ratings (Table 5.7) also do not differentiate by SAT cut score, when we examine the results for only those who selected the action. However, there are some differences by AFSs. In particular, respondents in Aircraft Fuel Systems and Security Forces, on average, report higher duration ratings than respondents in other specialties. Security Force respondents even report that they continuously rotate or swing light objects (i.e., objects weighing less than 10 pounds) anywhere from 31 minutes to an hour. Such repetitive activity can contribute to fatigue and injury even when lighter-weight objects are involved. Similar to the importance ratings, however, when we examine the total proportion of respondents selecting even a modest duration level (such as 6 to 10 minutes or more) we again see differences across career fields.

The no-assistance ratings (Table 5.8) exhibit some differences by SAT cut score when we examine the results for only those who selected the action. The two specialties with 80-pound cut scores more frequently performed the activity without assistance—indicated by an average no-assistance rating above 2.00 (i.e., sometimes required to push without help) than the other AFSs. Also, the average no-assistance ratings were higher for the specialties with 60-pound SAT cut scores than for those with 70-pound SAT cut scores. One explanation for this outcome is that the two 60-pound specialties are maintenance occupations, which require routine work in confined spaces, such as the inside of a fuel tank on a plane (AFECD, 2011, p. 107). Work in confined spaces would limit the ability to use carts or other mechanical devices for assistance.

Table 5.7
Duration Ratings for Rotate:
Only Those Checking the Action on the Screener

Weight Category (pounds)	SAT = 40		SAT = 60		SAT = 70		SAT = 80	
	CS	SS	AP-TTP	MAFS	AFE	SF	AFU-AS	EOD
< 5	2.23	3.08	2.67	3.54	2.57	4.10	2.72	2.92
5-9	1.91	2.96	2.59	3.38	2.76	4.14	2.74	2.95
10-24	2.43	2.46	2.59	3.12	2.79	3.90	2.75	2.87
25-39	2.57	2.21	2.43	2.80	2.72	3.21	2.79	3.03
40-69	2.43	2.18	2.08	2.58	2.69	2.86	2.66	2.88
70+	1.80	1.92	2.15	2.40	2.62	2.74	2.75	2.63
Sample Size	5-13	11-36	24-43	112-250	90-181	99-136	95-177	101-134

31 minutes to one hour (4.00 to 4.99)	11 to 30 minutes (3.00 to 3.99)	6 to 10 minutes (2.00 to 2.99)	5 minutes or less (1.00 to 1.99)

Table 5.8
No-Assistance Ratings for Push:
Only Those Checking the Action on the Screener

Weight Category (pounds)	SAT= 40		SAT = 60		SAT=70		SAT=80	
	CS	SS	AP-TTP	MAFS	AFE	SF	AFU-AS	EOD
10-24	2.08	2.02	2.21	2.12	2.09	1.97	2.22	2.18
25-39	1.93	1.90	2.14	2.09	2.08	1.92	2.20	2.12
40-69	1.80	1.94	2.05	1.98	1.90	1.93	2.12	2.11
70-99	1.67	1.85	1.82	2.00	1.83	1.81	2.05	2.04
100-199	1.61	1.77	1.80	1.86	1.75	1.82	1.93	2.05
200+	1.52	1.75	1.65	1.91	1.51	1.83	1.84	1.91
Sample Size	27-51	24-57	30-42	176-251	100-221	113-197	146-204	129-155

Between Never & Sometimes (1.0 – 1.89)	Sometimes (1.90 – 2.10)	Between Sometimes & Always (2.11 – 3.00)

NOTE: As a reminder, the No-Assistance survey question reads, "Please indicate…how often you are required to do this [i.e., push] WITHOUT HELP from carts, dollies, hand trucks, or other mechanical devices?"

Note again, however, that for both Tables 5.7 and 5.8, these are average ratings by those people that selected the action. While the average ratings may not differ markedly for some of these career fields, the proportions that selected the action do (i.e., there are large differences in the Table 5.7 and 5.8 sample sizes by AFS). Keeping this in mind leads to very different conclusions regarding the strength requirements of the AFS. For example, we can say that when those in Cyber Surety have to engage in a particular activity, they may have to do so for similar durations of time as those in more demanding career fields (like Explosive Ordinance Disposal); however, that activity is rare for Cyber Surety but it is commonplace for an AFS like Explosive Ordinance Disposal.

Summary of Results for Action

Overall, the average ratings of frequency, importance, duration, and no-assistance in the Action Section revealed some differences by AFS. Respondents in the specialties with 80-pound SAT cut scores reported more-frequent rotate, push, and carry activities—particularly for higher-

weight categories—than respondents in most of the other specialties. Respondents in the AFSs with 80-pound SAT cut scores also reported more requirements to push objects without assistance than respondents in other specialties. A comparison of a 40-lb AFS with an 80-lb AFS also indicated the overall career field percentage of those indicating that their physical tasks were at least moderately important was far lower for the 40-lb career field. Finally, respondents in the 40-pound AFSs tended to report less-frequent and lower-weight physical demands. Even among those who checked the action on the screener and hence reported that their job involved physical activity, most of the tasks involved that were at higher weights than 40 pounds were reported to occur "never" to "once in 1 to 2 years."

Movement Types

Because handling an object over one's head involves distinctly different strength requirements than handling it at waist height, and because the SAT specifically tests overhead lifting capacity, we also asked about the height at which respondents typically handle objects and other important object locations/positions with respect to the body. We also tried to determine if the positioning of the action was awkward, which might impose a greater strength requirement in order to handle the object without injury. In other words, we attempted to understand the type of movement typically involved in a particular action. Again, we assessed frequency and importance as well as the duration of the action in order to gather sufficient detail regarding physically demanding activities, while using skip patterns to reduce survey burden.

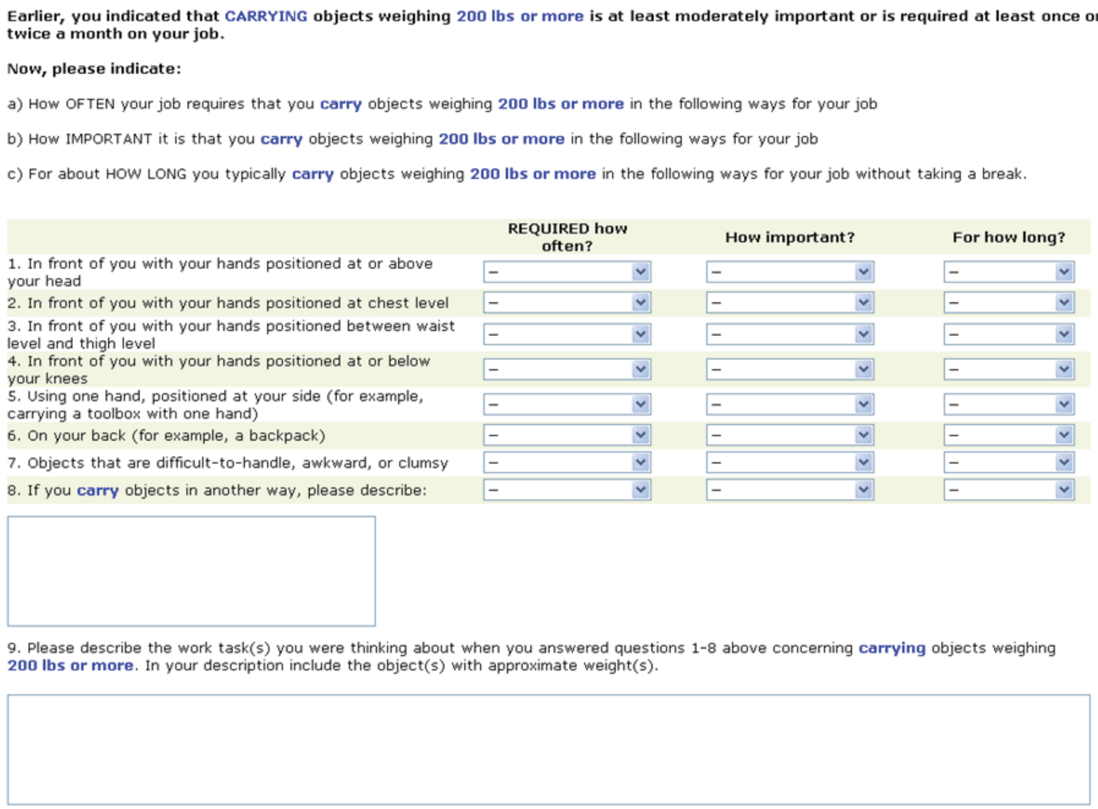
The same movement type questions were repeated for up to three different weight categories per action. For example, a participant might receive the same questions for lifting 10–24-pound weights, 40–69-pound weights, and 100–199-pound weights.[37] For each action-by-weight category, participants were asked about the frequency, importance, and duration for each movement type (e.g., carry a 40–69-pound object on the back). For each action-by-weight category, participants could write in a movement type not on the list and rate its frequency, importance, and duration. Respondents were also asked to provide a written description of the work tasks involved in that set of actions and weights, in order to provide context for interpreting responses. Figure 5.2 illustrates sample movement type questions.

Response rates for the Movement Type Section were lower than for the Action Section, perhaps because of survey length. So we focused on a subset of specialties, actions, and weight categories for our discussion. Specifically, we selected one AFS to represent three of the four SAT cut scores in our data: SAT = 60 (Aircraft Fuel Systems), SAT = 70 (Aircrew Flight Equipment), and SAT = 80 (Avionics Systems). Regrettably, we lacked sufficient sample sizes to include an AFS with a 40-lb cut score. We also limited our focus to five actions: push, pull, carry, hold, and lift. Many respondents did not complete most of the questions concerning throw—a sample size constraint—and we felt that the action of lowering was sufficiently similar

[37] We gave respondents question sets for only to those actions that they complete on a frequent basis (at least once a month) or that they rated as at least moderately important to perform, in order to target these questions at physical demands that would be important for overall determination of the physical demands of a career field.

to lifting that it was not necessary to include both in the discussion. Finally, we focused our discussion on two weight categories: 40–69 pounds and 70–99 pounds. We selected these weight categories not only because they had sufficient sample sizes for most questions but also for substantive reasons: the minimum SAT cut score is 40 pounds, and two of the three AFS cut scores represented in our data are 70 pounds or above. Comparisons between handling 40-69 pound objects and handling 70-99 pound objects could provide useful information about differences by SAT cut scores.

Figure 5.2
Sample Movement Type Questions—Carrying 200 Pounds or More



Earlier, you indicated that CARRYING objects weighing 200 lbs or more is at least moderately important or is required at least once or twice a month on your job.

Now, please indicate:

a) How OFTEN your job requires that you **carry** objects weighing **200 lbs or more** in the following ways for your job

b) How IMPORTANT it is that you **carry** objects weighing **200 lbs or more** in the following ways for your job

c) For about HOW LONG you typically **carry** objects weighing **200 lbs or more** in the following ways for your job without taking a break.

	REQUIRED how often?	How important?	For how long?
1. In front of you with your hands positioned at or above your head	–	–	–
2. In front of you with your hands positioned at chest level	–	–	–
3. In front of you with your hands positioned between waist level and thigh level	–	–	–
4. In front of you with your hands positioned at or below your knees	–	–	–
5. Using one hand, positioned at your side (for example, carrying a toolbox with one hand)	–	–	–
6. On your back (for example, a backpack)	–	–	–
7. Objects that are difficult-to-handle, awkward, or clumsy	–	–	–
8. If you **carry** objects in another way, please describe:	–	–	–

9. Please describe the work task(s) you were thinking about when you answered questions 1-8 above concerning **carrying** objects weighing **200 lbs or more**. In your description include the object(s) with approximate weight(s).

We took a different approach to analyzing the results for the Movement Type Section of the survey than we did for the Action Section. Instead of presenting tables of average ratings, we rank-ordered the top three movement types according to average ratings for the 40–69-pound weight category. Using this approach, we are able to identify the most popular movement type for an important weight category.

Frequency Ratings for Movement Type

We began our analysis of movement type with frequency. We calculated average frequency ratings using a counting rule akin to the first counting rule in the Action Section: only respondents who were correctly branched to the particular action-by-weight category were

counted toward the frequency averages for that action-by-weight category. As in the Action Section, we directed respondents to leave blank any rows of questions that did not apply to them. Thus, we coded missing frequency questions to Never (= 1) if a respondent completed at least one other question for that action-by-weight category. Table 5.9 provides the top three movement types for five actions (push, pull, carry, hold, and lift) based on average frequency ratings.

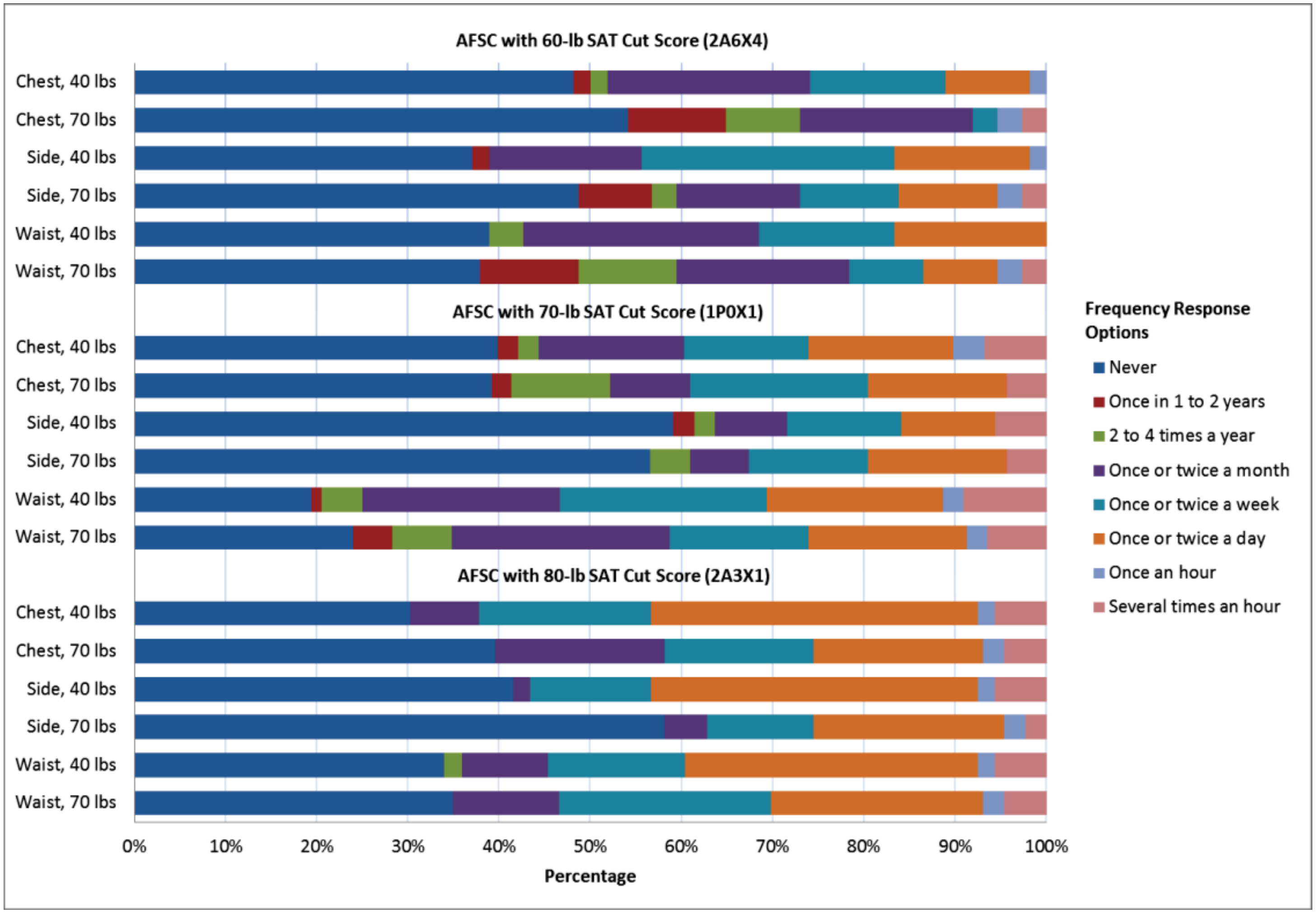
The most frequent movement type was fairly consistent across specialties and actions, particularly for carry and hold. Movements at waist level, on one's side, and at chest level figure prominently in Table 5.9. Other movement types, such as above the head, at knee level, and handling awkward objects also appeared in the top three but only for push and pull. Note here that Avionics Systems (the AFS with the 80-pound cut score) reported both pushing and pulling at or above chest level, on average, about once or twice a week. Based on the overall consistency in results, we decided to look further at waist, side, and chest movement type for one of the actions: lift. We focused on the action of lift because the SAT is mainly a measure of lifting ability. Figure 5.3 maps the percentages of respondents who selected different frequency response options for lift, using both the 40–69-pound and the 70–99-pound weight categories.

Table 5.9
Top Three Most-Frequent Movement Types at the 40–69-Pound Weight Category

	Aircraft Fuel Systems (SAT = 60)		Aircrew Flight Equipment (SAT = 70)		Avionics Systems (SAT = 80)	
Action	**Movement Type**	**Average Frequency**	**Movement Type**	**Average Frequency**	**Movement Type**	**Average Frequency**
Push	1. Waist	4.06	1. Waist	4.25	1. Chest	5.26
	2. Chest	4.02	2. Chest	4.22	2. Head	5.00
	3. Head	3.23	3. Knee	3.75	3. Awkward	4.05
Pull	1. Chest	4.23	1. Waist	4.14	1. Chest	5.05
	2. Waist	4.15	2. Chest	4.07	2. Head	4.98
	3. Knee	3.55	3. Knee	3.45	3. Waist	3.82
Carry	1. Waist	3.74	1. Waist	4.06	1. Waist	4.56
	2. Side	3.54	2. Chest	3.64	2. Side	4.24
	3. Chest	3.04	3. Side	2.88	3. Chest	3.41
Hold	1. Waist	3.31	1. Waist	3.96	1. Waist	4.31
	2. Side	3.06	2. Chest	3.65	2. Chest	3.98
	3. Chest	2.40	3. Side	2.98	3. Side	3.71
Lift	1. Side	3.48	1. Waist	4.40	1. Chest	4.28
	2. Waist	3.28	2. Chest	3.57	2. Waist	4.04
	3. Chest	2.89	3. Side	2.72	3. Side	3.89

NOTES: Frequency ratings ranged from 1 (Never) to 8 (Several times an hour). Sample size ranged from 40 to 70 for Aircraft Fuel Systems, 46 to 94 for Aircrew Flight Equipment, and 53 to 80 for Avionics Systems.

Figure 5.3
Frequency Ratings of Three Movement Types for Lift, for Three AFSs



The most prominent patterns in Figure 5.3 involve the differences by movement type within a particular specialty. For Aircraft Fuel Systems (SAT=60), the waist-level movement type was the only one that had similar endorsement rates of "Never" for both the 40–69-pound and 70–99-pound weight categories, although respondents who endorsed an option other than "Never" for waist-level lifting had higher frequency ratings for the 40–69-pound category than for the 70–99-pound category. For the other two movement types, respondents from the Aircraft Fuel Systems specialty reported more lifting at chest-level or at the side with 40–69-pound objects than with 70–99-pound objects.

For Aircrew Flight Equipment (SAT=70), differences by movement type were the largest of all three AFS samples. Lifting objects at waist level far exceeded the frequency ratings for lifting objects at chest-level or at the side of one's body. For example, over 80 percent of respondents from the Aircrew Flight Equipment specialty reported that they lift 40–69 pound objects at waist level, compared to only 60 percent who report doing so at chest level, and only about 40 percent who report doing so at their sides. These large differences by movement type were not mirrored by large differences by weight category for this specialty.

Finally, respondents in the Avionics Systems specialty—the AFS with an 80-pound SAT cut score—selected chest-level lifts with 40–69 pound objects as their most-frequent type of lift at that weight category. Moreover, respondents in this specialty tended to have higher-frequency ratings. For example, none of the respondents from Avionics Systems selected the option "Once in 1 to 2 years," whereas respondents from the other specialties selected that option for more than one question.

Overall, Figure 5.3 reveals important differences in frequency ratings by movement type. Moreover, the movement type differences vary by AFS but less so by weight category. We next look at whether movement type makes a difference when importance and duration ratings are used.

Importance and Duration Ratings for Movement Type

We conducted the same analysis for importance and duration of movement type as we did for frequency. We begin with importance ratings.

The rankings of the top three most-important movement types at the 40–69-pound weight category are shown in Table 5.10. Unlike the frequency-based rankings in Table 5.9, the importance-based rankings show more variability in movement type. In addition to chest-level, on the side, and waist-level, the most important movement type included above-the-head, knee-level, on one's back, and handling awkward objects. Small sample sizes partly explain why the rank ordering of movement types varies so much across AFSs. For example, fewer than ten respondents from the Aircraft Fuel Systems specialty completed one of the importance questions for hold. The small sample sizes for importance ratings also precluded an analysis similar to that illustrated in Figure 5.3. To the extent that the information in Table 5.10 reflects the true importance of certain movement types involving 40–69-pound objects, handling awkward objects is a moderately or very important movement type for different actions across all three specialties represented in the table.

Table 5.10
Top Three Most-Important Movement Types at the 40–69-Pound Weight Category

Action	Aircraft Fuel Systems (SAT=60) Movement Type	Average Importance	Aircrew Flight Equipment (SAT=70) Movement Type	Average Importance	Avionics Systems (SAT=80) Movement Type	Average Importance
Push	1. Head	3.68	1. Awkward	3.72	1. Head	4.31
	2. Knee	3.56	2. Waist	3.69	2. Waist	4.19
	3. Waist	3.42	3. Knee	3.69	3. Awkward	4.15
Pull	1. Head	3.89	1. Awkward	3.96	1. Head	4.31
	2. Awkward	3.80	2. Side	3.90	2. Knee	4.22
	3. Knee	3.69	3. Head	3.83	3. Chest	4.15
Carry	1. Side	3.64	1. Back	3.80	1. Awkward	4.43
	2. Awkward	3.55	2. Waist	3.73	2. Head	4.43
	3. Waist	3.46	3. Chest	3.73	3. Knee	4.21
Hold	1. Side	3.26	1. Waist	3.63	1. Awkward	4.52
	2. Waist	3.14	2. Head	3.62	2. Side	4.31
	3. Awkward	3.10	3. Side	3.55	3. Waist	4.30
Lift	1. Awkward	3.72	1. Awkward	3.70	1. Awkward	4.45
	2. Side	3.61	2. Knee	3.69	2. Knee	4.24
	3. Chest	3.52	3. Chest	3.63	3. Chest	4.19

NOTES: Importance ratings ranged from 1 (Not at all important) to 5 (Extremely important). Sample sizes ranged from 10 to 52 for Aircraft Fuel Systems, 13 to 66 for Aircrew Flight Equipment, and 19 to 72 for Avionics Systems.

In Table 5.11, we rank movement types based on duration. None of the average duration ratings are above 4.00 (31 minutes to 1 hour), suggesting actions that are of relatively short duration (30 minutes or less). The patterns in Table 5.11 somewhat mirror those in Table 5.10, which ranks movement type based on importance. First, the importance and duration ratings produced a greater variety of movement type for the top three spots than did the frequency ratings. However, these differences might be partly explained by sampling variability. Second, the movement type of handling awkward objects appears throughout Table 5.11 as it did in Table 5.10. As is the case with repetitive or continuous movements, research shows that handling awkward objects, if requiring awkward postures or body movements, relates to an increased risk of injury (Bernard, 1997).

Table 5.11
Top Three Longest-Duration Movement Types at the 40–69-Pound Weight Category

	Aircraft Fuel Systems (SAT=60)		Aircrew Flight Equipment (SAT=70)		Avionics Systems (SAT=80)	
Action	**Movement Type**	**Average Duration**	**Movement Type**	**Average Duration**	**Movement Type**	**Average Duration**
Push	1. Knee	2.96	1. Awkward	2.15	1. Knee	2.00
	2. Awkward	2.81	2. Knee	2.09	2. Waist	1.87
	3. Head	2.79	3. Waist	1.96	3. Awkward	1.82
Pull	1. Side	3.31	1. Side	2.67	1. Knee	1.78
	2. Head	3.11	2. Awkward	2.35	2. Awkward	1.71
	3. Awkward	3.00	3. Knee	2.28	3. Side	1.71
Carry	1. Back	3.09	1. Awkward	1.97	1. Back	2.43
	2. Head	2.56	2. Head	1.93	2. Awkward	2.13
	3. Knee	2.52	3. Waist	1.91	3. Side	2.08
Hold	1. Head	2.70	1. Knee	2.06	1. Side	1.72
	2. Knee	2.60	2. Head	1.93	2. Knee	1.68
	3. Awkward	2.33	3. Chest	1.92	3. Awkward	1.67
Lift	1. Head	2.24	1. Awkward	2.00	1. Side	1.35
	2. Awkward	2.06	2. Side	1.94	2. Awkward	1.25
	3. Chest	2.06	3. Knee	1.93	3. Head	1.21

NOTES: Duration ratings ranged from 1 (5 minutes or less) to 7 (more than 8 hours). Sample sizes for duration ranged from 16 to 51 for Aircraft Fuel Systems, 14 to 64 for Aircrew Flight Equipment, and 14 to 68 for Avionics Systems.

Summary of Results for Movement Type

Although small sample sizes restricted our analysis of movement types, interesting patterns were found in the results that we were able to produce. The top-ranked movement types based on average frequency ratings were fairly consistent across specialties and actions, with waist-level, chest-level, and on-the-side appearing across all AFSs and most of the actions. However, further analysis of those three movement types revealed important differences within specialties, particularly within the Aircrew Flight Equipment specialty.

In contrast to the frequency-based rankings, rankings based on importance or duration ratings varied considerably more. Again, low sample sizes could explain this. Nevertheless, despite the variability in rankings, one movement type featured prominently throughout importance and duration: objects that are awkward to handle. If personnel have important tasks involving handling awkward objects, the risk for fatigue and injury would need to be considered when determining job demands and assigning them to particular jobs.

Overall Summary

Overall, the average ratings of frequency, importance, duration, and no-assistance in the Action Section revealed some differences by AFS. Respondents in the specialties with 80-pound

SAT cut scores reported more-frequent rotate, push, and carry activities—particularly for higher-weight categories—than respondents in most of the other specialties. Moreover, objects were more commonly pushed and carried at higher weights than they were rotated. In general, there was a trend such that AFSs with higher SAT cut scores reported higher frequencies of physical demands; the exception to this trend was the Aerospace Propulsion AFS, a career field that includes only junior personnel. Respondents in the AFSs with 80-pound SAT cut scores also reported more requirements to push objects without assistance than respondents in other specialties. Finally, respondents in the 40-pound AFSs tended to report that activities involving weights more than 40 pounds occurred "never" to "once in 1 to 2 years," though among those performing such actions, Surgical Services tended to report these activities as more important ("moderately" to "very important") than did Cyber Surety.

When examining in more detail the conditions of performance for important and/or frequent tasks, the top-ranked movement types based on average frequency ratings were fairly consistent across specialties and actions, with waist-level, chest-level, and on-the-side appearing across all AFSs and most of the actions. Avionics Systems, one of the AFSs with an 80-pound cut score, reported both pushing and pulling at or above chest level, on average, about once or twice a week. Further analysis of the commonly reported waist-level, chest-level, and on-the-side movement types revealed important differences within specialties. For example, the 70-pound cut score AFS Aircrew Flight Equipment specialty participants reported that lifting objects at waist level far exceeded the frequency for lifting objects at chest-level or at the side of one's body. Respondents in Avionics (80-pound cut score) also tended to report higher frequencies of lifting than the other career fields examined. An important caveat to the findings in this section is the low sample sizes, which may produce greater variability in results.

These findings suggest that the SAT cut score may indeed distinguish career fields with greater physical demands. However, our survey offers far more detail about the frequency, importance, and nature of those demands. The low frequency and importance ratings for some of the career fields with lower cut scores may suggest that physical demands are simply not a substantive part of their job performance.

Although the screener we used did not do a sufficient job of screening participants out of more-detailed question sets, the screener with a higher minimum weight (perhaps of 25 pounds rather than 10) could help distinguish career fields that indeed have physical demands substantive enough to warrant career field entry requirements. Using follow-on questions to determine perceived weight demands in conjunction with frequency and importance may signal a need to reevaluate a career field's physical demands if the pattern is not characteristic of lower- versus higher-demand cut scores. If a baseline is taken for a given career field, a change in the pattern that indicates a change in physical demand (much more frequent performance, or an infrequent action at a high weight suddenly increasing in importance rating to "very" or "extremely" important) might trigger a more in-depth audit and potential revision of the cut score. When career fields separate or combine, or a piece of equipment that is much heavier is introduced, the survey may be deployed and compared to the baseline ratings to determine if an adjustment to the cut score may be necessary (for example, if importance or frequency ratings of actions using the new, heavier equipment do not change, it may be an indication that the cut

score should not alter). Finally, the type of information collected in our survey would be a valuable addition to the process of setting cut scores if a new method is chosen. For example, information regarding the frequency and importance of actions, and movement location, would be helpful in determining the expectations for a minimally acceptable performer in a given career field. This information could be used to help subject matter experts (SMEs) determine what the job demands for a given career field really are, particularly in cases where there is variance at different geographic locations with which the SME may not have direct experience.

OAD’s process of career field task development, or an independent series of discussions with career field SMEs, could also be leveraged to ensure more targeted survey content, although it would be optimal to ensure that a uniform baseline be taken prior to narrowing survey content. If a given type of movement or physical demand seems likely to be essential for job performance based on detailed physical demand information obtained from SMEs, survey content could cover that particular area more thoroughly and evaluate other demands in a more cursory fashion in order to reduce response burden.

A response rate at least equivalent to that obtained by other types of surveys would increase confidence that survey results are indeed representative. Sending survey invitation emails from an Air Force address would potentially help survey response rates, given some technical difficulties we experienced. In addition to getting supportive endorsements from career field managers, it would also be helpful to get endorsements from MAJCOM commanders and to provide explicit duty time to the collection of these data. Given that survey information would be most useful with a baseline for each career field, it would be important to ensure higher response rates and/or coverage of an appropriate sample of bases and skill levels to be able to regard these findings as representative for that baseline. The survey discussed here exemplifies the type of surveys that are characteristic of organizations attempting to set physical demands. Thus, to the extent that our results are characteristic of the known demands of the career fields and of the current SAT cut score, as well as being in line with current best practice, the survey shows promise.

6. Conclusions and Recommendations

Over the course of this project, we closely examined the procedures that were used to develop the SAT, that are currently used to administer the SAT, and that are used to establish minimum cut points for entry into various AFSs. From that examination, we have a number of recommendations regarding the Air Force's use of strength tests going forward.

Continue to Use Strength Testing to Screen for Certain AFSs

It is clear from job descriptions and from the confirmatory information provided by our survey that there are AFSs in the Air Force that require high levels of strength. For those AFSs, failure to screen for strength capability could have negative consequences. Personnel who are not strong enough to handle the objects involved in the job could be injured while attempting the work. Additionally, their inability to properly control or stabilize heavy objects could cause others to be injured around them. Injuries are problematic not only because of the potential immediate and long-term medical costs, but also because of the downtime associated with having personnel out on medical leave. In addition to injuries, it is also likely that those who are not strong enough to accomplish the job will not be relied on to do the work. Not only would they not be able to accomplish the work, they might also be taking up a billet of someone who would do better.

For these reasons, we caution against entirely abandoning the idea of strength testing or eliminating the use of the SAT without finding a suitable replacement. On the other hand, we do think that alternative tests should be pursued, and the existing cut scores should be reexamined to make sure that they are not set too high or too low for a given AFS. These suggestions are discussed further below.

Enforce Proper Administration of the Strength Aptitude Test

Until an alternative measure is identified, the use of the SAT should continue. While the SAT is in use, administration of the test should adhere to specific guidelines to ensure the fairness and effectiveness of the scores.

Our examination of how the SAT is currently being administered at the MEPS showed that it is administered in a generally consistent manner and in line with current guidelines. There is also general agreement that the test is useful in assigning recruits to career fields. But our investigation did uncover some inconsistencies in test administration that could have a meaningful impact on test scores and ultimately on the career fields for which recruits are eligible. We offer several recommendations that could improve test administration and, in turn, the assignment of career fields, over the long run.

Recommendation: *Conduct a full inventory of SAT machines on a regular basis (every few years).*

SAT machines may be damaged, in need of parts, or need to be replaced. MEPS locations should be contacted regularly to identify if they have more than one machine at that location, and to determine whether any of their machines are damaged and in need of repair. When machines are identified as damaged, they should either be repaired or replaced.

Recommendation: *Send new instructions to all MEPS locations, develop a standardized training procedure for all LNCOs, and audit this implementation.*

Extra effort should be taken to ensure that MEPS station personnel adhere to the test administration guidelines. Through our interviews we learned that there are deviations from the established protocol. SAT administration varies to some degree from LNCO to LNCO and from site to site. Though the variances are small, they could impact test scores in a meaningful way—that is, result in test scores that are neither consistent nor comparable across recruits, thus making the test less useful in determining whether a recruit is qualified for a career field.

To remedy this, we suggest that the Air Force issue new guidance outlining what is and is not allowed during test administration. In addition, current MEPS station personnel should be retrained and new personnel should be trained in the proper procedures, and explanations for why there should not be deviations from the protocol should be included in the training. The aim of this effort would be to eliminate any variation in test administration that would diminish the usefulness and fairness of SAT scores. Audits of implementation would ensure that consistency is maintained over time, as well.

Recommendation: *Require recruiters to inform recruits about the SAT and encourage preparation.*

Having prior knowledge and understanding of the SAT and having the opportunity to prepare could significantly impact test scores. For this reason, new guidance should be issued to recruiters requiring them to fully inform recruits about the test before they send the recruits to the MEPS. Specifically, recruiters should make sure that recruits understand the nature of the test and how it relates to career field assignments. They should also explain that the test requires them to do a series of six-foot lifts on a machine with increasing weights, and that if they want to do well on the test, they should go into a gym and prepare for the test. Recruiters should also make sure to communicate to recruits the proper attire that should be worn for the test (such as tennis shoes and shorts/pants). They should especially communicate that they will need to squat with feet apart to properly grip the handle bars and initiate the lift; therefore, skirts, tight-fitting or low-cut pants/shorts, and high heels should not be worn to the MEPS. An explanatory pamphlet would help ensure the information provided is standardized.

Change the Methods for Establishing Career Field Cut Scores

Recommendation: *Establish a new method for converting job demands information into an SAT cut score.*

Many elements in the SAS program are unsupported (including some of the regression equations), and other key elements that should be considered in establishing cut scores are absent (including duration and importance of various tasks). For these reasons, the process for converting job tasks into a suitable AFS-specific cut score should be changed.

Factors that should be explored in developing a new process include

- using well-established approaches for setting standards (examples can be found in Cizek, 2001)
- compensating for gains expected from basic training
- considering task importance and duration in addition to frequency and percentage of people performing the task
- considering a wider variety of physical demands, such as those that may emphasize endurance in addition to those that require strength
- using score crosswalks instead of regression equations for converting information about the force associated with one action to another.

In all cases, the method and process for setting cut scores should be comprehensively documented. Given Ayoub et al.'s (1987) finding of variation in physical task performance by location, this should also be taken into consideration in the process of setting cut scores. In some cases, if a particular subset of personnel in an AFS need to undertake demanding tasks, a shred (career subfield) may be established with a different cut score. Alternately, if a given highly demanding physical task is only performed at one location and the job itself cannot be re-engineered to reduce the demand, it may be worth considering a physical training requirement prior to personnel rotating into the job. Note that both of these scenarios presume agreement within a career field that a given physical demand is indeed required. If there is no agreement on a demand, it should not be considered for an AFS-wide or even shred-wide cut score. Finally, if a given career field has low physical demands (as evidenced by a high proportion indicating they do not perform tasks requiring manipulating a 25-pound weight or higher), cut scores and physical demands testing may not be necessary.

Recommendation*: Add items addressing physical demands to OAD's occupational analysis survey.*

In 1996, the GAO reported that:

> Each of the services has ongoing processes through which they can identify occupational tasks in each specialty in order to revise training curriculums and which they use for other reasons. However, the services do not collect data on the physical demands of jobs with these processes.

Today, we draw the same conclusion regarding the Air Force's occupational analysis reports. AETC's OAD collects extensive data about the tasks involved in every enlisted job via an online survey, conducted every three years. Unfortunately, the data OAD collects do not include sufficient information to ascertain the strength requirements in the job. The results of this report illustrate that OAD could include survey items, like those on our Strength Requirements Survey, to address this.

Our survey findings also exemplify another potential use of job analysis information focusing on physical demands of the job. In light of the concerns discussed in Chapter 2 regarding the process currently used to define the job demands (i.e., the fact that examination of only three bases may not be sufficient to represent a career field and the process may rely on faulty inferences), conducting an online survey should be considered as an alternative or at least as a supplement. Therefore, we strongly recommend that the Air Force add questions about physical job requirements to OAD's occupational analysis survey and make them a permanent component of the survey that is administered every three years. OAD also does endeavor to ensure appropriate representativeness and comprehensiveness of its sample, and, because its surveys come from an Air Force system, it would not encounter some of the technical difficulties that our survey did.

Our survey findings also lead to suggestions regarding how such a survey tool should be improved before it (or something like it) is added to the occupational analysis survey.

- **Refine the screening tool.** Our results suggest that the Screener did not perform as we had hoped. Many people still ended up having to complete the long form of the survey. So we suggest adjusting the screener to prevent false positives. The following are three ways to do this:
  - Change the instructions. For example, ask respondents to select the action only if it involves objects weighing more than 25 pounds (our survey screener said "greater than 10 pounds").
  - Set the threshold for triggering the supplemental strength survey to be higher. For example, in the case of Cyber Surety, 35 percent of the respondents selected "none required" and fewer than 60 percent selected the highest-ranked action (carry). Taking this as a baseline for rates of false positives, a minimum of 70 percent for the actions or fewer than 20 percent selecting "none required" could serve as the trigger for a follow-on survey.[38]
  - Incorporate a task-specific screener as part of the regular occupational analysis survey. On the regular survey, OAD could ask respondents to indicate which tasks are physically demanding. Or, during the stage of survey development, prior to fielding a survey for a given career field, OAD could ask its career field managers and subject matter experts to flag the demanding tasks. For those tasks that are flagged, a follow-on survey regarding weight, frequency, and importance could be administered as part of the regular survey effort.
- **Add questions about other types of physical job demands.** Although the Strength Requirements survey covered a wide variety of strength-related actions, other types of physical activities were not included. Cardiovascular endurance or stamina may be important aspects of many occupations. Also, some military jobs may require activities such as swimming, marching for long distances wearing heavy equipment, etc. Future surveys of the physical demands of airmen's jobs should add items addressing these other

---

[38] We do not know what the appropriate baseline is. One way to establish it would be to administer the screener to all AFSs and set the bar at a level that excludes the lowest scoring career fields.

types of physical requirements. In some cases, the level of fitness required to accomplish these tasks is commensurate with those to achieve acceptable performance on Tier I physical fitness testing for health purposes; however, for other cases it may not be.

- **Items could be tailored to be task-specific.** OAD currently maintains a comprehensive list of tasks in a given career field for use in the occupational analysis survey that is administered every three years. Items could be created that ask about the physical demands associated with each of those tasks, and the frequency, importance, and duration of those physical demands. A comparison of a task-specific questionnaire to that of a generic survey (like the one used in our study) would be worthwhile.
- **Compare survey responses to other evaluations of job demands, and use specific evaluations of physical job demands to target questions.** Invariably, the results of the survey may lead to additional questions about the demands in some career fields. Following up the results by interviewing career field managers, conducting focus groups or interviews with job incumbents, and conducting in-person site visits may be warranted to better understand the demands in certain AFSs. This is a process that is consistent with the work already done by OAD in preparing for and interpreting the results of its occupational analysis surveys, and it would be vital for ensuring that the minimum score for entry into some AFSs is set properly. This could also be used at the front end as a mechanism to reduce survey burden and fatigue: If a given type of movement or physical demand seems likely to be essential for job performance, survey content could cover that particular area more thoroughly and evaluate other demands in a more cursory fashion.

We also offer the following suggestions regarding analysis of the results:

- **Compare responses by gender and skill level.** Gender and skill-level differences in survey responses should be compared when measuring job demands. If gender differences are identified on the survey, further examination of why perceptions of the job requirements might differ by gender should be explored before setting the minimum cut point for the AFS. For example, it is possible that women have devised a less physically demanding, yet equally effective, way of performing the requirements. If that is the case, and the alternatives are safe, these alternatives should be promulgated more generally. On the other hand, it is possible that women, because of differences in physical strength (whether perceived or actual), are not being allowed to perform important aspects of the job. In the former case, the cut point should be adjusted to reflect less demanding alternative ways of doing the job. In the latter, the cut score should not be adjusted if the tasks are critical to the job. Skill-level differences are also important. It may be the case that the most physically demanding work is done by the lower skill levels. If so, the cut score should pay more attention to the lower skill-level responses. However, measurement of higher skill levels is also important if those levels have substantial physical demands that are not represented in earlier skill levels.
- **Ensure that survey questions are analyzed properly**. For example, while frequency ratings for actions revealed consistent mean score differences by AFS—in that specialties with higher SAT cut scores reported more-frequent actions—the same difference was not observed for other types of ratings (such as importance and duration ratings). However, when we examined the total proportion of people rating them as moderately important or higher (including those who indicated that they did not do the action at all in the group

that rated it as unimportant), we discovered that there were meaningful differences. Thus, examination of data that are absent as a result of important skip patterns is critical.

Lastly, we recommend that OAD continually refine its survey items to ensure that they adequately capture AFS specific physical demands, and conduct periodic checks (e.g., by meeting with career field managers and AFS subject matter experts) to ensure that the results are accurate.

Conduct a Predictive Validity Study Using the SAT and Other Alternative Measures

The link between test performance and on-the-job performance is critical for determining the overall effectiveness of a test. However, research on the SAT has not adequately explored this issue.[39] This work is necessary for justifying the continued use of the SAT or the use of any other physical ability tests. As a corollary, conducting and adequately documenting this work would also put the Air Force in line with best practice.

Recommendation: *Begin collecting data on the SAT and other alternative tools before and after basic training. The link between test performance and on-the-job performance is critical for determining the overall effectiveness of a test for use in future validation studies.*

We suggest proceeding with the research in two stages. In Stage 1, data on the SAT and other measures should be collected both prior to and after basic training. The purpose of the two measurements is to establish the amount of improvement that could be expected on each of the potential measures resulting from basic training. The types of measures considered should cover a wide range of physical skills, including upper-body strength, lower-body strength, cardiovascular endurance, anaerobic power, etc. The practicality of the tests should also be considered in deciding which to include in the study (i.e., the tests should not be time consuming, require substantial space or equipment, expensive, or difficult to administer at a MEPS). A thorough examination of the tests that have been considered by other researchers (such as a lift to chest or waist height, a leg press, a step test of cardiovascular endurance, a Wingate test,[40] or others cited in this report) would be good starting points for identifying a broad range of viable measures.

In this initial phase of the research, sample sizes should include sufficient numbers of women and minorities to allow for an examination of gender and race differences. In addition, the sample sizes for these groups should be as large as possible in anticipation of their use in the subsequent phases of the study.

Collection of this Stage 1 data would allow for some immediate findings, including comparisons of scores by race and gender for all tests, estimation of the amount of improvement

---

[39] The 1996 GAO report also suggested that the link between test performance and on-the-job performance was a concern and strongly advised that this link be examined empirically.

[40] The Wingate test for lower- and upper-body anaerobic power requires pedaling or using an arm crank at maximum speed with resistance.

caused by basic training, and a comparison of the SAT scores obtained in a controlled testing environment versus that of the MEPS.

Stage 2 of the research would involve collecting data on performance outcomes. We suggest collecting those data in two ways. First, participants from Stage 1 could be asked to come back and participate in a laboratory work simulation activity several months after they have started in their AFS. The simulations could be similar to those described in Ayoub et al. and others (e.g., Rayson et al., 2000) to develop the SAT, utilizing similar movement patterns (box lifting and lowering to a simulated truck bed height, jerry can or box carrying activities, sled pushing; ideally, these patterns would be related to movement patterns also required on the job by important or frequent tasks similar to those a survey such as ours would detect, were it deployed broadly). However, making improvements to the methodology, such as varying the durations involved in the simulations and recording that information, would be important. In addition, adding a significant time gap between initial testing and the subsequent simulation activity would better approximate the actual predictive validity of the test because participants would not be fatigued.

We suggest that the second method of collecting performance data be supervisor's ratings of performance relating to the physical aspects of the job. Direct supervisors of the participants from Stage 1 could be contacted and asked to evaluate their performance in certain physically demanding but important aspects of the job. The aspects of the job on which members of each AFS would be rated could be identified a priori through meetings with career field managers or other members of the AFS using simple rating scales. Ideally, this would be done for every AFS unless a deliberate determination was made that the job did not in fact have strength demands that necessitated testing (perhaps by using our screener, modified to ask about 25-pound actions, or by interviewing a sample of career field SMEs and examining career field documentation).

The results of the Stage 2 data collection efforts would be vital for demonstrating the predictive validity of the tests examined in Stage 1. A study involving both stages of data collection would provide a solid ground for determining which tests are most predictive, which tests show the least amount of predictive bias against key subgroups (i.e., race and gender), and whether one test should be used for certain AFSs and another test used for a different group of AFSs.

Appendix A. AFSC Codes and Career Field Specialty Names

Table A.1
List of AFSC Codes and Corresponding Specialty Names

1A0X1 In-Flight Refueling	2A0X1 Avionics Test Station and Components
1A1X1 Flight Engineer	2A3X1 A-10, F-15, & U-2 Avionics Systems
1A2X1 Aircraft Loadmaster	2A3X2 Integrated Avionics Systems (Attack/Special)
1A3X1 Airborne Mission Systems	2A3X3 Tactical Aircraft Maintenance
1A4X1 Airborne Operations	2A5X1 Aerospace Maintenance
1A6X1 Flight Attendant	2A5X2 Helicopter/Tiltrotor Maintenance
1A7X1 Aerial Gunner	2A5X3 Integrated Avionics Systems (Heavy)
1A8X1 Airborne Cryptologic Language Analyst	2A6X1 Aerospace Propulsion
1A8X2 Airborne ISR Operator	2A6X2 Aerospace Ground Equipment
1B4X1 Cyberspace Defense Operations	2A6X3 Aircrew Egress Systems
1C0X2 Aviation Resource Management	2A6X4 Aircraft Fuel Systems
1C1X1 Air Traffic Control	2A6X5 Aircraft Hydraulic Systems
1C2X1 Combat Control	2A6X6 Aircraft Electrical and Environmental Systems
1C3X1 Command Post	2A7X1 Aircraft Metals Technology
1C4X1 Tactical Air Control Party	2A7X2 Nondestructive Inspection
1C5X1 Command & Control Battle Management Ops	2A7X3 Aircraft Structural Maintenance
1C6X1 Space Systems Operations	2A7X5 Low Observable Aircraft Structural Maintenance
1C7X1 Airfield Management	2F0X1 Fuels
1N0X1 Operations Intelligence	2G0X1 Logistics Plans
1N1X1 Geospatial Intelligence	2M0X1 Missile and Space Systems Elect Maintenance
1N2X1 Signals Intelligence Analyst	2M0X2 Missile and Space Systems Maintenance
1N3X1 Cryptologic Language Analyst	2M0X3 Missile and Space Facilities
1N4X1 Network Intelligence Analyst	2P0X1 Precision Measurement Equipment Laboratory
1P0X1 Aircrew Flight Equipment	2R0X1 Maintenance Management Analysis
1S0X1 Safety	2R1X1 Maintenance Management Production
1T0X1 Survival, Evasion, Resistance, and Escape	2S0X1 Materiel Management
1T2X1 Pararescue	2T0X1 Traffic Management
1U0X1 Remotely Piloted Aircraft Sensor Op	2T1X1 Vehicle Operations
1W0X1 Weather	2T2X1 Air Transportation
1W0X2 Special Operations Weather	2T3X1 Vehicle and Vehicular Equipment Maintenance

2T3X2 Special Vehicle Maintenance
2T3X7 Vehicle Management & Analysis
2W0X1 Munitions Systems
2W1X1 Aircraft Armament Systems
2W2X1 Nuclear Weapons
3D0X1 Knowledge Operations Management
3D0X2 Cyber Systems Operations
3D0X3 Cyber Surety
3D0X4 Computer Systems Programming
3D1X1 Client Systems
3D1X2 Cyber Transport Systems
3D1X3 RF Transmission Systems
3D1X4 Spectrum Operations
3D1X5 Ground Radar Systems
3D1X6 Airfield Systems
3D1X7 Cable and Antenna Systems
3E0X1 Electrical Systems
3E0X2 Electrical Power Production
3E1X1 Heating, Ventilation, AC, & Refrigeration
3E2X1 Pavements and Construction Equipment
3E3X1 Structural
3E4X1 Water and Fuel Systems Maintenance
3E4X3 Pest Management
3E5X1 Engineering
3E6X1 Operations Management
3E7X1 Fire Protection
3E8X1 Explosive Ordnance Disposal
3E9X1 Emergency Management
3H0X1 Historian
3M0X1 Services
3N0X1 Public Affairs
3N0X2 Broadcast Journalist
3N0X4 Still Photography
3N1X1 Regional Band
3N2X1 Premier Band
3P0X1 Security Forces
3S0X1 Personnel
3S1X1 Equal Opportunity
3S2X1 Education and Training
3S3X1 Manpower
4A0X1 Health Services Management
4A1X1 Medical Materiel
4A2X1 Biomedical Equipment
4B0X1 Bioenvironmental Engineering
4C0X1 Mental Health Service
4D0X1 Diet Therapy
4E0X1 Public Health
4H0X1 Cardiopulmonary Laboratory
4J0X2 Physical Medicine
4M0X1 Aerospace and Operational Physiology
4N0X1 Aerospace Medical Service
4N1X1 Surgical Service
4P0X1 Pharmacy
4R0X1 Diagnostic Imaging
4T0X1 Medical Laboratory
4T0X2 Histopathology
4V0X1 Ophthalmic
4Y0X1 Dental Assistant
4Y0X2 Dental Laboratory
5J0X1 Paralegal
5R0X1 Chaplain Assistant
6C0X1 Contracting
6F0X1 Financial Management & Comptroller
7S0X1 Special Investigations
8A100 Career Assistance Advisor
8A200 Enlisted Aide
8B000 Military Training Instructor
8B100 Military Training Leader
8B200 Academy Military Training NCO
8C000 Airman & Family Readiness Center RNCO
8D000 Linguist Debriefer
8E000 Research, Analysis and Lessons Learned
8F000 First Sergeant
8G000 Honor Guard
8H000 Airmen Dorm Leader
8M000 Postal
8P000 Courier
8P100 Defense Attaché
8R000 Enlisted Accessions Recruiter
8R200 Second-Tier Recruiter

8R300 Third-Tier Recruiter

8S000 Missile Facility Manager

8T000 Professional Military Education Instructor

9A000 Awaiting Retraining–Reasons Beyond Control

9A100 Awaiting Retraining–Reasons Within Control

9A200 Awaiting Discharge/Separation/ Retirement

9A300 Awaiting Discharge/ Separation/Retirement for Reasons Beyond Their Control

9A400 Disqualified Airman, Return to Duty Program

9C000 CMSgt of the Air Force

9E000 Command Chief Master Sergeant

9F000 First Term Airmen Center

9G100 Group Superintendent

9J000 Prisoner

9L000 Interpreter/Translator

9P000 Patient

9R000 Civil Air Patrol (CAP)-USAF Reserve Assistance

9S100 Scientific Applications Specialist

9T000 Basic Enlisted Airman

9T100 Officer Trainee

9T200 Pre-Cadet Assignee

9U000 Enlisted Airman Ineligible for Local Utilization

9U100 Unallotted Enlisted Authorization

9W000 Potential Wounded Warrior

9W100 Reserved for Future Use

9W200 Wounded Warrior

9W300 Wounded Warrior–Returned to Duty

9W400 Wounded Warrior–Limited Assignment Status (LAS)

9W500 Wounded Warrior–Retired/Discharged

9W600 Reserved for Future Use

9W700 Reserved for Future Use

9W800 Reserved for Future Use

9W900 Reserved for Future Use

SOURCE: AFECD, April 30, 2011.

Appendix B. Additional Details on the Process Currently Used to Establish SAT Cut Scores

This appendix provides technical details on the information that is collected during the resurvey process that had been used to establish cut points since the inception of the SAT in 1987. Table B.1 illustrates the content that is collected during site visits and from occupational analysis reports and provides a sample of the excel spreadsheet that is produced from the job resurvey process. The information in Table B.1 is then fed into the SAS program that crunches the numbers and produces the final cut score. A description of what the SAS program does with the numbers is explained in the section below.

How the Final Cut Score Is Produced (Explanation of the SAS Program Code)[41]

The first step in the SAS code involves reading and recoding the data provided by the contractor that does the resurveying. The SAS program reads and records the following info:

- Object Description—description of the target object to be weighted
- Object Weight (TSKFORCE)
- Task Movement Type (SMTSK) The physical action performed during the task. These actions are identified during interviews with workers. Each task was analyzed to determine the type of physical demand required, such as lifting, carrying, pushing, pulling, etc.
- Number of people (NOPEO)—Number of people required to perform each task
- Occupational Analysis Report Task Number (LNO)
- Task Frequency (FREQ)—Frequency of each task
- Percentage (Perc)—Percentage of 1st term airmen.

Next, the SAS code converts the frequency from the number of times per year to a letter code. Codes range from Yearly, to Daily as follows:

- 1 time per year = Yearly (Y)
- 2–3 times per year = Semiannually (S)
- 4–7 times per year = Quarterly (Q)
- 8–25 times per year = Monthly (M)
- 26–99 times per year = Weekly (W)
- 100 or more times per year = Daily (D)

[41] The interpretation of the SAS code was gleaned from our meetings with Greg Zehner and HyegJoo Choi (711th HPW/RHPA), an unpublished briefing summarizing the interpretation of the SAS code, and our review of the actual SAS code itself.

Table B.1
Example Information Collected and Calculated When Reestablishing Cut Scores

	Collected During Site Visit				Pulled from Occupational Analysis Reports			Produced by SAS Code			
Site Visit Task #	Object Description	Object Weight (TSKFORCE)	Task Movement Type (SMTSK)	Number of People Involved (NOPEO)	Occupational Analysis Report Task Number (LNO)	Task Frequency (FREQ)	Percent of Personnel Doing Task (Perc)	Per-Person Weight (Force)	Equivalent Vertical Lift Weight (X1)	Weighted Perc (WTP)	Weight for FREQ (WTF)
1	Generator - Miller Big Blue 500D	237	P2	1	S569	Daily	18	237	134.36	0.82	3
2	Portland cement - bag	90	L8	1	B 82	Daily	55	90	131.73	2.01	3
3	Softcut saw - Norton clipper	242	C4	2	A 19	Daily	74	121	111.31	2.62	3
4	Tire - Tractor (24" rim/16.9" wide)	241	C4	2	A 38	Monthly	77	121	111.04	2.71	1
5	Target econoline concrete saw	117	P4	1	A 19	Weekly	74	117	107.03	2.62	1.75
6	Geotextile - 1 roll	400	C3	4	T607	Weekly	20	100	105.75	0.88	1.75
7	Water barrier-empty (8', 200gal)	209	C4	2	A 25	Quarterly	62	105	102.02	2.23	.2
17	Jackhammer (with bit)	89	C4	1	A 2	Weekly	86	89	92.6	3.00	1.75
22	Wooden ramps - transport material	88	C4	1	T629	Semi-annually	8	88	91.96	0.5	0.1

This portion of the SAS code is written as follows:

```
IF FREQN = 1 THEN FREQ='Y';
IF FREQN = 2 OR FREQN = 3 THEN FREQ='S';
IF FREQN GE 4 AND FREQN LT 8 THEN FREQ='Q';
IF FREQN GE 8 AND FREQN LT 26 THEN FREQ='M';
IF FREQN GE 26 AND FREQN LT 100 THEN FREQ='W';
IF FREQN GE 100 THEN FREQ='D';
IF FREQN = . THEN FREQ=' ';
```

Next, the SAS code calculates the per-person weight (called FORCE) for each action type:

```
FORCE=TSKFORCE/NOPEO;
```

and calculates X1 using regression equations from the original research used to select the SAT as the Air Force's strength test. X1 is intended to represent the amount of vertical lift on the SAT that corresponds to the force exerted when engaging in a different movement (e.g., P1 – low-level push; P2 – low-level pull, etc.).

This section of the SAS code is written as follows:

```
FORCE = TSKFORCE/NOPEO;
IF SMTSK = 'L1' THEN DO; * PATIENT HANDLING;
if tskforce le 170 and nopeo lt 2 then nopeo=2;
if tskforce gt 170 and nopeo lt 3 then nopeo=3; * force=tskforce/nopeo;
X1=6.89661+0.3783366*((tskforce-20)/nopeo);
END;
IF SMTSK='L2' THEN DO;
X1=-53.83552+18.08275657*FORCE**0.5;
END;
IF SMTSK='L6' THEN DO;
X1=-31.648093+12.08225934*FORCE**0.5;
END;
IF SMTSK='L7' THEN DO;
X1=-17.284023+11.50576248*FORCE**0.5;
END;
IF SMTSK='L8' THEN DO;
X1=-56.929896+19.88653984*FORCE**0.5;
END;
IF SMTSK='L9' THEN DO;
X1=-31.2655535+18.91308746*FORCE**0.5;
```

```
END;
IF SMTSK=‘C2’ THEN DO;
X1=-50.66176+15.99146875*FORCE**0.5;
END;
IF SMTSK=‘C3’ THEN DO;
X1=-27.995348+13.37477065*FORCE**0.5;
END;
IF SMTSK=‘C4’ THEN DO;
X1=-20.136857+11.94973478*FORCE**0.5;
END;
IF SMTSK=‘P1’ THEN DO;
X1=-9.396+0.40390704*FORCE;
END;
IF SMTSK=‘P2’ THEN DO;
X1=-9.330+0.60628811*FORCE;
END;
IF SMTSK=‘P3’ THEN DO;
X1=-14.205+0.6067755*FORCE;
END;
IF SMTSK=‘P7’ THEN DO; * 38CM LIFT;
X1=-22.578377+8.781529*FORCE**0.5;
END;
IF SMTSK=‘P4’ THEN DO; * HIGH PULL;
X1=-24.185084+12.131223*FORCE**0.5;
END;
IF SMTSK=‘H3’ THEN DO;
X1=-55.2871+16.41555677*FORCE**0.5;
END;
IF SMTSK=‘H4’ THEN DO;
X1=-55.66845+16.93856394*FORCE**0.5;
END;
IF SMTSK=‘S2’ THEN DO; * DO NOT USE - DO NOT KNOW WHAT S2
REPRESENTS;
X1=0.96073032+1.76294185*FORCE;
END;
```

```
IF SMTSK='S3' THEN DO; * DO NOT USE - DO NOT KNOW WHAT S3
REPRESENTS;
X1=-11.2258826+3.27271401*FORCE;
END;
IF SMTSK='T1' THEN DO; * DO NOT USE - DO KNOW KNOW WHAT T1
REPRESENTS;
X1=8.967205+0.925833*FORCE;
END;
```

Several of the regression equations used in the above SAS code also appear in the study described in Ayoub et al. More specifically, ten of the equations are approximate linear transformations of the regression equations reported in Ayoub et al. to convert the equations from kilograms to pounds. The transformed equations do not match those reported in Ayoub et al. exactly, perhaps due to some sort of rounding error in the transformation. Unfortunately, not all of the regression equations in the SAS code were reported in Ayoub et al. The following action types are missing from the published article: L1, P4, P7, S2, S3, and T1. Therefore, for these actions, we have no description of the action and no estimate of the R-squared value for the regression equation. In addition, equations for P1, P2, and P3 do not appear to be linear transformations of those reported in Ayoub et al., and no explanation for these equations is provided elsewhere. We were referred to Gibbons (1989), who in discussion of the CREW CHIEF legacy software program, mentions benchmarking strength tests such as the ILM lift to represent Air Force strength requirements via regressions utilizing the data they summarize, but a citation for these regressions is not provided.

The formulas from the above code generally adhere to one of the two following patterns:

$$y = \beta+\beta x^{1/2} \text{ or } y = \beta+\beta x \text{ .}$$

In Ayoub et al., they provide both formulas (those where they enter x into the regression equation and those where they use the square root of x in the formula instead). In that article, the authors argue that the formulas taking the square root of x are superior because the other formulas exhibited heteroscedasticity, even though the R-squared values were lower. We note, however, that while some of the formulas in the SAS code above are those that take the square root of X, others are not. There is no explanation available for why.

As noted in the above SAS code, it appears that the Air Force does not use S2, S3, and T1 because it does not know what action those codes represent.

The final step in the process involves translating all of the information from all of the tasks into a single SAT cut score by weighting the X1 values by their frequency and the percentage of people performing the task. The process defined in the SAS code is as follows:

First FREQ is recoded as a numerical weight ranging from .05 to 3 and named WTF as follows:

```
IF FREQ='D' THEN WTF=3.0;
IF FREQ='W' THEN WTF=1.75;
```

```
IF FREQ='M' THEN WTF=1.0;
IF FREQ='Q' THEN WTF=0.2;
IF FREQ='S' THEN WTF=0.1;
IF FREQ='Y' THEN WTF=0.05;
```

Next, the SAS code identifies the task with the lowest percentage of people engaging in it (Perc) and names the percentage of people doing that task as MINP and the task with the highest percentage and names it MAXP across all tasks in the data. So, in Table B.1 this would belong to Task 17 and Task 22 at 86 percent and 8 percent respectively. Next the SAS program calculates the weighted percentage (WTP) for each task as:

```
WTP = (Perc–MINP)/(MAXP–MINP)*2.5+0.5
```

The above formula is a simple linear transformation that serves to rescale Perc from a percentage ranging from 0 to 100 to a scale that ranges from .5 to 3.0. Then WTF and WTP are averaged for each task to create an average weight for the task (WTAVG2):

```
WTAVG2 = (WTP+WTF)/2
```

In other words, the frequency of the task and the percentage of people performing the task are intended to contribute equally in computing the final cut score.

The SAS code next calculates the size of each task's new weight relative to the sum of the new weights across all tasks and calls it PWT:

```
PWT = WTAVG2/SUM
```

The final step involves calculating weighted X1 values (PX1) as

```
PX1 = X1*PWT
```

The weighted X1 values are then summed to create the final number for the career field, called: ADJX1. It is this final sum that is then used to determine the final SAT cut score. It is rounded to create RNDX1 as follows:

```
IF ADJX1 < 200 THEN RNDX1=130;
IF ADJX1 < 126 THEN RNDX1=120;
IF ADJX1 < 116 THEN RNDX1=110;
IF ADJX1 < 106 THEN RNDX1=100;
IF ADJX1 < 96 THEN RNDX1=90;
IF ADJX1 < 86 THEN RNDX1=80;
IF ADJX1 < 76 THEN RNDX1=70;
IF ADJX1 < 66 THEN RNDX1=60;
IF ADJX1 < 56 THEN RNDX1=50;
IF ADJX1 < 46 THEN RNDX1=40;
```

RNDX1 is the final SAT cut score required by the career field. It is this number that is submitted to AFPC/A1PF and to the career field managers for final review and approval.

No justification for the weighting and averaging process explained above is provided. Ayoub et al. describe a process that is quite similar; however, no explanation for why they adopted that process was provided in their report. In addition, there are a few differences between the process in Ayoub et al. and the process applied in the SAS code above, including the fact that the SAS code omits task importance as a weighting factor; does not adjust the cut score based on the size of the AFS; and has some unexplained differences in the regression equations (as noted above). We interviewed Dr. Joe McDaniel (one of the authors of Ayoub et al. and the person responsible for establishing and managing the SAT cut score process at the Air Force), who confirmed that there is no other documentation on why or how these formulas were chosen. He could not provide us with any further information to support their use.

Appendix C. LNCO and Recruit Interview Questions

LNCO Interview Questions

1. Describe the normal test protocol.
2. Describe any other ways that the SAT is administered.
3. Describe any needed repairs or problems with ILMs that could be affecting the SAT.
    - Have you had any problems with the ILM?
    - Do you know of any needed repairs to the ILM? Describe them.
    - Do you think there is anything about the ILM's state of repair that could be affecting SAT scores?
4. Describe from start to finish how you administer the SAT. Leave nothing out, even if it seems trivial or you think everyone probably knows.
5. Who explains and/or demonstrates the SAT to recruits?
    - What do you say to recruits before they begin?
    - What do you say to recruits during the test?
6. Do you always start with the carriage alone (40 pounds)? Is there any reason the starting weight would be higher than 40 pounds? Lower than 40 pounds?
7. How do you determine if the lifting carriage is six feet high? If a recruit lifts close to six feet would you let them continue?
8. What is the weight increment per attempt (ten pounds)? Is there any reason the weight increment would be higher than ten pounds? Lower than ten pounds?
9. Do you pause for a certain amount of time before the recruit's next attempt? How long? Is there any time or reason this would be longer or shorter than usual (i.e., really strong recruit trying for a heavy weight needs a few extra seconds to get ready)?
10. How long does it take to administer the SAT to one recruit?
11. What is the maximum weight you allow recruits to attempt (110 or 200 pounds)?
12. Is there any circumstance where you would record a number higher than the weight the recruit actually lifted to six feet? If a recruit lifts less than 40 pounds, how is the score recorded? If a recruit lifts more than 110 pounds, how is the score recorded?
13. Does lifting protocol or starting position matter?
    - If a recruit is lifting unsafely, how do you know?
    - What do you look for that would make you stop the test because it may be unsafe for the recruit?
    - When the recruit lowers the carriage from six feet to the floor, must it be lowered slowly and in a controlled motion?
    - Is there any reason a recruit would get a second chance to lift a weight attempt they previously failed to lift to six feet?

14. How many other recruits are watching when one recruit is taking the SAT? How many other recruits taking the SAT would one recruit observe before it's his/her turn? Does it vary by position in line, or is it a continuous line?
15. When in the MEPS timeline is the SAT administered? Any physical or tiring components before the SAT?
16. Is there anything else about the SAT we haven't talked about?

Recruit Interview Questions

1. What did you hear about the SAT before you arrived at the MEPS?
2. Did you try to practice for the SAT in any way?
3. What were you told about the purpose of the SAT before you took the test today?
4. Do you know how much the carriage weighs? (on your first lift)
5. Do you know how much weight you lifted on your last trial? (on your last lift)
6. Do you know how much weight you have to lift to qualify for your preferred Air Force job?
7. Did you experience any obstacles during the test? Were you tired, bored, distracted, etc.?
8. Overall, what do you think of the SAT? Do you think the SAT is a good measure of your strength?

Appendix D. Tabular Overview of Survey

Below, we provide an overview of the entirety of the survey sections, followed by discussion of the specific sections addressed in this document.

Table D.1
Summary of Survey Topics and Purpose

Survey Topics	Purpose
Demographics	Identify any skill-level and gender differences.
Strength-requirements screener	Identify the basic type of *actions* (pull, push, lift, lower, carry, hold, throw, support one's body weight, rotate/swing), if any, that are required on the job. Evaluate the functioning of the screener tool.
Action weight/importance/frequency	Provide additional details about the *weight* of the objects involved in the actions, and the *importance* and *frequency* of the actions.
Movement type/duration	Identify how the action is performed (e.g., lifting overhead, lifting to chest height, etc.) and for what *duration* it is typically performed.
Other strength demands	Determine if the survey items have missed any important aspect of physical activity required on the job.
Final survey comments	

The survey began with a screening tool designed to exclude participants that did not have job strength requirements:

Table D.2
The Strength-Requirements Screener

Please indicate whether your job (i.e., your current duty AFSC) REQUIRES the following types of activities. (Check all that apply.)	
☐	**SUPPORTING YOUR BODY** in positions other than normal sitting, standing, or walking. By supporting your body, we mean using your physical strength to support your own body weight in positions other than normal sitting, standing, or walking. Examples include propping yourself up with one arm to drill something with another arm and squatting to access a panel on the underside of a plane.
☐	Continuously or repeatedly **ROTATING or SWINGING** objects or sets of materials of any weight with your hands. By rotating or swinging, we mean using your hands and fingers to continuously or repeatedly manipulate objects in a curved pattern. Examples include turning wheels or levers and swinging a hammer several times in a row. This category does NOT include the other actions on this page, even though rotating or swinging objects may be needed to do the other actions (e.g., swinging a line of cable to then throw it).

Please indicate whether your job (i.e., your current duty AFSC) REQUIRES the following types of activities.
(Check all that apply.)

- ☐ **PUSHING/PRESSING** objects weighing 10 lbs. or more.
  By pushing/pressing, we mean using your hands and/or arms to move objects forward while you either stay in place (e.g., stand) or move your lower body (e.g., walk). Examples include pushing windows closed, pushing a box across the floor, and pressing your hands against a door to keep it from opening.
- ☐ **PULLING** objects weighing 10 lbs. or more.
  By pulling, we mean holding onto an object with your hands to move the object toward you while you either stay in place (e.g., stand) or move with your lower body (e.g., walk). Examples include pulling a door closed, dragging a box across the floor, and dragging a line of cable or a hose.
- ☐ **CARRYING** objects weighing 10 lbs. or more.
  By carrying, we mean holding objects in your arms, hands, or on your back while you move with your lower body (e.g., run). Examples include walking with a box in your arms, running with a backpack on your back, and holding a toolbox at your side while walking. This category does NOT include lifting or lowering objects, even though lifting or lowering is often required to carry objects (e.g., lifting a box off a table to then carry it across a room).
- ☐ **HOLDING** objects weighing 10 lbs. or more.
  By holding, we mean using your upper-body strength to maintain objects in your arms, hands, or on your back while you stay in place (e.g., stand). Examples include sitting with a box in your arms without the box resting on your lap and holding a toolbox at your side while standing in place. This category does NOT include lifting or lowering objects, even though lifting or lowering is often required to hold objects (e.g., lowering a box off a shelf to then hold it).
- ☐ **LIFTING** objects weighing 10 lbs. or more.
  By lifting, we mean using your hands and/or arms to move an object in an upward direction. Examples include moving a box from a lower shelf to a higher shelf and picking up a toolbox off the floor to put it on a table.
- ☐ **LOWERING** objects weighing 10 lbs. or more.
  By lowering, we mean using your hands and/or arms to move an object in a downward direction. Examples include moving a box from a higher shelf to a lower shelf and taking a toolbox off a table to put it on the floor.
- ☐ **THROWING/TOSSING** objects weighing 10 lbs. or more.
  By throwing/tossing, we mean thrusting or propelling an object out of your hands and/or arms, while you either stay in place (e.g., stand) or move with your lower body (e.g., walk). Examples include throwing a line of cable across a room and throwing sand bags into the bed of a truck.
- ☐ My job does not require me to do any of these types of activities.

COMMON OBJECTS that weigh approximately 10 lbs:
metal folding chair
full-sized ironing board
standard two-by-four (approx. 2-in deep, 4-in wide, and 8-ft long; made of pine wood)

The next section of the survey is the Action Section. This section takes the basic type of actions (pull, push, lift, lower, carry, hold, throw, support one's body weight, rotate/swing), identified as required on the job from the screener and asks for additional details about the weight of the objects involved in the actions, and the importance and frequency of the actions. For some actions we also asked about duration and performance of actions without assistance.

Figure D.1
Action Section—Question Formats

Support Your Body and Rotate

To save time, you can skip any rows that are not applicable to you. You do not have to answer each question to move forward with the survey.

Please indicate:

a) How OFTEN your job requires the following activities

b) How IMPORTANT it is that you perform the following activities for your job

c) For about HOW LONG you typically perform the following activities for your job without taking a break.

CLICK HERE for Definitions of SUPPORT YOUR BODY and ROTATE or SWING and for common objects of different weights.

SUPPORT YOUR BODY in positions other than normal sitting, standing, or walking (for example, squat to access a panel on the underside of a plane).

	REQUIRED how often?	How important?	For how long?
	–	–	–

Continuously or repeatedly **ROTATE or SWING** an object or sets of materials with your hands (for example, swing a hammer several times in a row) of the following weights:

	REQUIRED how often?	How important?	For how long?
less than 5 lbs	–	–	–
5 to 9 lbs	–	–	–
10 to 24 lbs	–	–	–
25-39 lbs	–	–	–
40-69 lbs	–	–	–
70 lbs or more	–	–	–

Action Section – Question Format for Push

PUSH/PRESS an object weighing approximately

	REQUIRED how often?	How important?	WITHOUT HELP from carts, dollies, etc.?
10-24 lbs	–	–	–
25-39 lbs	–	–	–
40-69 lbs	–	–	–
70-99 lbs	–	–	–
100-199 lbs	–	–	–
200 lbs or more	–	–	–

NOTE: Question format is the same for push and pull.

Action Section –Question Format for Carry

CARRY an object weighing approximately

	REQUIRED how often?	How important?
10-24 lbs	–	–
25-39 lbs	–	–
40-69 lbs	–	–
70-99 lbs	–	–
100-199 lbs	–	–
200 lbs or more	–	–

NOTE: Question format is the same for carry, hold, lift, lower, and throw.

Below are the item stems and response options used in the Action Section:

Table D.3
Response Options for Action Section Questions

Survey Question	Data Value	Response Options
Frequency: How frequently does your job require it?	1 2 3 4 5 6 7 8	Never Once in 1 to 2 years 2 to 4 times a year Once or twice a month Once or twice a week Once or twice a day Once an hour Several times an hour
Duration: For how long without taking a break?	1 2 3 4 5 6 7	5 minutes or less 6 to 10 minutes 11 to 30 minutes 31 minutes to 1 hour 2 to 4 hours 5 to 8 hours More than 8 hours
Importance: How important is it?	1 2 3 4 5	Not at all important Slightly important Moderately important Very important Extremely important
No assistance: How often without assistance from carts, dollies, and other conveyances?	1 2 3	Never Sometimes Always

To help clarify the questions in the Action Section regarding the weight of objects, participants were able to link to a screen that provided examples of object weights that participants might encounter in everyday life.

Table D.4
Object Weight Examples

Weight	Common Objects with Approximate Weights
Less than 5 pounds	a hammer with a 12-in wood handle (1 lb) an average clothes iron (3-5 lbs) an average bathroom scale (3-5 lbs)
5 to 9 pounds	a small, table-top ironing board (5-8 lbs) a cordless, 12-volt power drill for home use (5-9 lbs)
10 to 24 pounds	a metal folding chair (10 lbs) a full-sized ironing board (10 lbs) a standard two-by-four (approx. 2- inches deep, 4-in inches wide, and 8-ft feet long; made of pine wood) (10 lbs) a cordless, 18-volt power drill for commercial use (10-–12 lbs) a standard, adult-sized bowling ball (12–16 lbs) one passenger car tire, inflated (20 lbs) a 32-inch LCD flat-screen TV (18–25 lbs)
25 to 39 pounds	an average two-year-old child (25 lbs) three metal folding chairs (30 lbs) one mid-sized microwave (35 lbs) a full propane tank for a gas grill (38 lbs)
40 to 69 pounds	a five-gallon plastic water cooler jug filled with water (40 lbs) a small bag of cement mix (50 lbs) a mini window air conditioning unit (40–60 lbs) two large bags of dry dog food (60–69 lbs)
70 to 99 pounds	a punching bag (70–80 lbs) two five-gallon plastic water cooler jugs filled with water (80 lbs) a large bag of cement mix (80–90 lbs) three standard (8-in by 8-in by 16-in) cinder blocks (90–100 lbs)
100 to 199 pounds	a large-sized, adult, male dog, such as a rottweiler or bloodhound (100–130 lbs) a standard, top-loading clothes washing machine (140–150 lbs) an average, adult, American woman (140–160 lbs) an average, adult, American man (170–190 lbs) an average, freestanding kitchen range and oven (185-–200 lbs)
200 pounds or more	seven standard (8-in by 8-in by 16-in) cinder blocks (200–230 lbs) two large-sized, adult, male dogs such as rottweilers or bloodhounds (200–260 lbs) an average NFL linebacker (230-270 lbs)

Participants who completed questions in the Action Section were eligible to branch to the Movement Types Section. Because handling an object over one's head involves distinctly different strength requirements than handling it at waist height, we asked about the height at which respondents typically handle objects and other important object locations/positions with respect to the body. We also tried to determine if the positioning of the action was awkward which might impose a greater strength requirement. To reduce survey burden, we limited the number of follow-on Action Section questions sets to only three weight categories per action, for activities that met requirements for importance and frequency. In the Movement Types section, we asked about duration, importance, and frequency of activities at given locations.

Figure D.2
Sample Movement Type Questions—Carrying 200 Pounds or More

Earlier, you indicated that CARRYING objects weighing 200 lbs or more is at least moderately important or is required at least once or twice a month on your job.

Now, please indicate:

a) How OFTEN your job requires that you **carry** objects weighing **200 lbs or more** in the following ways for your job

b) How IMPORTANT it is that you **carry** objects weighing **200 lbs or more** in the following ways for your job

c) For about HOW LONG you typically **carry** objects weighing **200 lbs or more** in the following ways for your job without taking a break.

	REQUIRED how often?	How important?	For how long?
1. In front of you with your hands positioned at or above your head	–	–	–
2. In front of you with your hands positioned at chest level	–	–	–
3. In front of you with your hands positioned between waist level and thigh level	–	–	–
4. In front of you with your hands positioned at or below your knees	–	–	–
5. Using one hand, positioned at your side (for example, carrying a toolbox with one hand)	–	–	–
6. On your back (for example, a backpack)	–	–	–
7. Objects that are difficult-to-handle, awkward, or clumsy	–	–	–
8. If you **carry** objects in another way, please describe:	–	–	–

9. Please describe the work task(s) you were thinking about when you answered questions 1-8 above concerning **carrying** objects weighing **200 lbs or more**. In your description include the object(s) with approximate weight(s).

Below are the item stems and response options used in the Movement Type Section:

Table D.5
Response Options for Movement Type Section Questions

Survey Question	Data Value	Response Options
Frequency: How frequently does your job require it?	1	Never
	2	Once in 1 to 2 years
	3	2 to 4 times a year
	4	Once or twice a month
	5	Once or twice a week
	6	Once or twice a day
	7	Once an hour
	8	Several times an hour
Duration: For how long without taking a break?	1	5 minutes or less
	2	6 to 10 minutes
	3	11 to 30 minutes
	4	31 minutes to 1 hour
	5	2 to 4 hours
	6	5 to 8 hours
	7	More than 8 hours
Importance: How important is it?	1	Not at all important
	2	Slightly important
	3	Moderately important
	4	Very important
	5	Extremely important

Appendix E. Responses to Open-Ended Survey Questions

The Strength Requirements Survey ended with three types of open-ended questions: (1) descriptions of work tasks and physical job demands, (2) suggestions on ways to reduce (or not reduce) physical job demands, and (3) general comments at the end of the survey. We describe the questions asked and responses in this appendix.

Work Tasks and Physical Job Demands

Respondents were asked to give additional detail about their work tasks in three ways. First, the survey asked respondents to describe the activities that they were thinking of when responding to the items in the Movement Type Section of the survey. Of the 1,769 respondents for the Movement Type Section, 819 provided at least one description. Results by AFS are described below and sample sizes by action and AFS are shown in Table E.1.

Second, in the Movement Type Section we also offered an "other, please specify" option in the list of movement types, in case a movement type was not adequately represented in our list. Few respondents utilized this "other" option. Of the few that did, most provided descriptions that matched other existing categories in the Movement Type Section.

Third, we asked all respondents (including those who did not get branched to the Action Section) to describe any other job demands not already addressed in the survey that required physical strength. A total of 302 respondents provided a description; however, most descriptions were similar to those made in the Movement Type Section. Sample sizes by AFS are shown in Table E.1. Cases where comments offered additional insights are noted in the discussion below.

Table E.1
Number of Comments by AFS

	Write-In Explanations for Movement Type							
	SS	**CS**	**AP-TTP**	**MAFS**	**AFE**	**SF**	**AFU-AS**	**EOD**
Push	29	--	22	--	113	72	111	62
Pull	22	--	21	--	--	--	82	--
Cary	19	57	20	--	188	120	77	--
Hold	--	--	--	--	--	--	64	--
Lift	17	45	--	--	96	46	61	--
Lower	--	--	--	--	--	--	56	--
Throw	--	3	--	--	--	--	8	34
Total, all actions	31	68	23	144	175	161	121	99
Other job demands not addressed elsewhere								
Total	14	26	--	43	37	84	25	--

Cyber Surety. Across all actions, the weight category with the most comments was 10–24 pounds, followed by 25–39 pounds. Low-weight activities included pushing or pulling carts with boxes of materials (e.g., office supplies or manuals), carrying computer equipment (e.g., a

computer or a monitor) from one area of the office to another, holding a server to align it with a rack-mount, and lifting and lowering boxes of computer equipment from and to tables and the ground. At higher weights, activities involved moving around heavy safes or the occasional heavy box. The following are examples:

> Push and pull a 3-4 feet cart up and down the hall in order to transport these FLIP [Flight Publications] to the proper shelving units after building and updating the bags.
>
> Different types of computer equipment (varying weights between 10 and 39 pounds), including computer cases, monitors, laptops, display televisions, etc. Some items are two man carry due to awkwardness or size.
>
> Boxes being brought out for customer service have to be lowered to the ground or table and boxes received from DCS/Registered Mail are stored on the ground.

Surgical Service. Most comments were in the two lowest weight categories (10–24 and 25–39 pounds). Actions at the lower weight categories involved holding, lifting, lowering, or carrying surgical instrument sets or supplies and pushing or pulling such materials on carts. At higher weights, work tasks usually involved moving patients onto and off of gurneys or, in rare cases, pulling injured airmen out of hostile areas. The following are examples:

> Pulling the sterilizer rack out of the sterilizer onto the cart and then pulling the cart away from the heat of the sterilizer to cool
>
> Carrying sets to the OR [operating room] from CSS [central sterile supply] or carrying it from the cooling rack to put it on the shelf.
>
> While scrubbed into surgical procedures surgical techs (SS) frequently aid the surgeon by holding retractors. The strength, tension, grip, position and duration very widely but it is rarely ergonomic. To provide visualization for the surgery, isometric tension must be used and this results in muscle and joint fatigue. Surgeries vary greatly in length but it's not uncommon for a tech to spend 2–4 hours in one case.

In response to our question about strength requirements we missed, respondents noted the need to stand in place for hours during surgery.

Aerospace Propulsion. The first four weight categories (i.e., up to and including 70–99 pounds) had the largest number of comments for all actions except push and pull. Comments indicate that Aerospace Propulsion personnel carry, hold, lift, and lower tools, toolboxes, engine parts, and other aircraft parts (e.g., propeller components). Such objects are moved around the workshop, out of trucks, and into and out of the aircraft during installation and repair. Some of these tasks involve more than one person, for example:

> Extremely important that certain engine parts be held while another mechanic starts bolts as to not damage seals or surfaces. Power turbine 80–90 lbs

For push and pull, 200 pounds or more had the most responses with descriptions of pushing or pulling aircraft engines (weighing hundreds or even thousands of pounds depending on the type of engine) on engine stands. This confirms our expectation that pushing and pulling actions for 200 pounds or more would require assistance of carts or other mechanical conveyances, such as engine stands.

In general, the comments confirm that personnel in this specialty handle objects above 60 pounds; however, in many cases respondents did not indicate whether they received assistance from others when moving such objects.

Like Aerospace Propulsion personnel, Aircraft Fuel Systems specialists had more push and pull comments at 200 pounds or more than at lower weights, and many reported having to hold a part in place while another person works on it. The following is an example:

> Holding parts in place while installing or having parts/tools ready to hand someone who is about to install them.

However, many of the weight-related comments were unique to this specialty, such as working with fuel systems, such as pulling fire suppression foam inside a fuel tank or maneuvering (i.e., pushing, pulling, lifting, and lowering) heavy fuel cell bladders. One respondent described such activities in detail:

> Pushing rubber fuel cells straight up into a fuel cell cavity or pulling them up into the cavity from the inside of the aircraft. This takes two people and the cells weigh over 100 lbs. Pushing our maxi tool kit into position in the hangar when preparing for maintenance. This tool kit weighs over 1000 lbs but rolls somewhat easily once it gets going. Pushing/pulling maintenance stands into position around the aircraft. These stands vary in weight but all are a great deal over 199 lbs. I would approximate between 500 and 100 lbs. We do this quite often.

In response to our question about strength requirements we missed, Aircraft Fuel Systems specialists described the awkwardness of working inside fuel tanks, including having to crawl through the tanks.

Aircrew Flight Equipment. For Aircrew Flight Equipment personnel, comments were most common in the first four weight categories (i.e., up to and including 70–99 pounds). Respondents described work tasks related to inspecting, packing, and loading or unloading parachutes, flight gear, life-support equipment and supplies, test equipment, and life rafts, among other things. Specific actions included lifting and lowering parachutes and gear on and off of airplanes; tossing bags of survival gear into trucks; pulling supplies off of shelves and carrying them to trucks for transport; and pushing carts with equipment. For the higher-weight categories (i.e., 70–99 pounds and above) several comments involved moving large rafts:

> If stairs are not available, which is frequent, then to remove life rafts (70–80 lbs) from the aircraft, you have to lower it down a ladder to a person or people on the ground. The floor of the aircraft is about 10 ft off the ground, so the raft must be lowered by a person laying on the floor of the aircraft to 5–6 feet off the ground where the people below can catch it.

Security Forces. For Security Forces the four lowest-weight categories had the most comments. Descriptions included everything from moving (e.g., carrying, lifting, and holding) weapons, ammunition crates, and personal gear to handling dogs and people, for example:

> There are times when I lift/carry a military working dog weighing close to 100 lbs in a vari kennel weighing 20 lbs.

Comments about working dogs were uncommon; however, comments about wearing or carrying bags of personal gear, especially while on guard duty, were common. The following is an example:

> WE have to carry all required equipment to any post we are at. Required equipment includes flak vest, helmet, second chance vest, cold weather gear, wet weather gear, gloves, eye pro, ear pro, whistle, M9 holster, M9 with 30 rounds of ammo, M4 with 90-120 rounds of ammo, OC Pepper spray, Monadnock collapsable baton, AFMAN, handcuffs with key, and gas mask with canister.

In response to our question about strength requirements we missed, wearing heavy personal gear—sometimes for long stretches of time (e.g., 12 hours)—was a major theme.

Avionics Systems. For Avionics Systems and Explosive Ordnance Disposal specialties, the 200-pounds-or-more weight category had many descriptions (more than in any other specialty). Activities included replacing line replaceable units (LRUs), carrying, lowering, and lifting various tools and tool boxes (sometimes while on ladders), holding amplifiers in place while they are being installed, and moving (e.g., lifting and lowering) hydraulic test stands, engine trailers, engines, and aircraft cooling air units, and aircraft ground equipment (AGE). Lower-weight categories usually involved objects like handling tools or toolboxes and LRUs, as one respondent describes below:

> When you replace a LRU you actually pull one out and push a new one in. A lot of them are awkward in shape also. They are scattered throughout the aircraft, ranging from ankle high when on top of the aircraft, and chest to overhead level when standing on the ground.

Higher-weight categories involved objects like radar transmitters, some types of LRUs, amplifiers, and AGE, which one respondent claimed can weigh "as much as a small car." Objects that are heavy are handled by more than one person:

> Holding the band 3 aft amp up so they can connect wires an things, or holding the band 3 fwd amp up so they can push it into place (97 pounds)

Other actions, while optional, often are needed to be efficient or most effective at one's job:

> avionics parts, support equipment, tool boxes, age equipment…Avionics airmen carry, lift, lower, raise, all items noted above. We upload and download these items from vehicles and or carts etc. Many times we carry these items several hundred yards if need be. For example if I was at my 11 1/2 hour mark of work, working outside in the weather on a jet that is 200 yards away from the building and I did not expect to get a ride anytime soon, I would carry what I could inside. I would say on a daily basis. Airman who are faced with carrying in a part to fix a jet or wait for a ride typically hike it.

Explosive Ordinance Disposal. Comments included moving boxes, bags, or bins of gear, explosives, and equipment on and off shelves and into and out of trucks or other vehicles; wearing bomb suits and other gear; removing explosives from the ground or clearing spent ordnance during range operations; pulling injured team members to cover; and handling (pulling, lifting, and carrying) robots and damaged vehicles. Several respondents commented that certain activities, like handling robots, are common in deployed environments. For example:

> 65-pound robot being pulled by a rope into the back of a large armored vehicle approximately 5’ off the ground. Pulling explosive devices from hardened earth with rope or cord weighing 25–80 lbs. Daily occurrence in Afghanistan.

Explosive Ordinance Disposal personnel often wear protective equipment or a bomb suit, and removing injured personnel wearing the bomb suits was a common response for the two highest weight categories (100–199 pounds and 200 pounds or more). For example:

> The effort require to pull a fully loaded EOD Operator up a wall or out of a ditch frequently exceeds 200 lbs. Additionally, pulling an injured or unconscious team member out of a hazard zone requires significant strength as these members will most likely exceed 200 lbs. If clad in the EOD 9 Bomb Suit weights well over 250 lbs would not be uncommon. The EOD Operator carries a full combat load and mission essential tools and gear. This ensemble makes for an awkwardly shaped package that becomes increasingly difficult to manipulate with fatigue or increased weight.

Compared to the other AFSs, Explosive Ordnance Disposal had the most comments about throwing objects. Most involved range clearance operations as described below:

> Clearing ranges of dud-ordnance requires repeated collection of 25 lb practice bombs, by picking up those bombs and tossing them into a front-end-loader bucket or the back of a dump-truck. The work is highly repetitive and could last for hours.

Overall, the comments indicate that Explosive Ordnance Disposal is a highly physical specialty, requiring everything from repetitive throwing actions with lighter-weight (25 pound) objects to activities pulling an injured teammate to safety.

Final Survey Comments

The last question of the survey allowed respondents to provide any additional comments about the physical requirements of their jobs. A total of 121 airmen took the opportunity to submit final comments to the survey. Of those 121 airmen, 11 wrote “none” or otherwise, indicated the question was not applicable. We thus analyzed the comments from the remaining 110 respondents.

As one might expect, airmen touched upon a variety of topics in the final survey question. As with the previous open-ended questions, we did our best to code responses and put them into categories.

The comment categories in Table E.2 fall into two groups. The first group includes injury and strains working with parachutes. A lot of the descriptions in this group were related to repetitive motions or working in awkward positions, such as having to repeatedly pack parachutes on the ground. Other comments reflected the long-term impacts of having a career in a physically demanding job, as seen by the lengthy comment in the Injury category in Table E.2.

The second group includes height and weight restrictions and increased strength requirements. Comments in this group were less about complaints or concerns and more about recommending changes to the type of personnel allowed to enter the specialty. Comments about height and weight restrictions were interesting because the comments made by Aircrew Flight

Equipment personnel were more concerned about people not being tall enough to do the job (e.g., load parachute packs on high shelves), whereas the comments made by Aircraft Fuel Systems personnel focused on people being *too* tall or too heavy to fit into and maneuver inside fuel tanks. Other comments were less about size and more about strength. Respondents from three different specialties stated that higher strength standards are needed. Arguments focused on the need to reduce the risk of injury associated with weaker individuals and/or to improve performance in the AFS.

Table E.2
Categories of Final Survey Comments That Do Not Overlap with Comment Categories from Previous Questions

Category	Category Description	Number of Responses	AFSs Associated with Category	Typical Comment
Injury	Respondent has injury related to work or thinks job generally causes injuries	13	Security Forces, Aircrew Flight Equipment, and Avionics Systems	Regardless of your physical conditioning and strength over the years with even the proper techniques this job takes an effect on your body. One bad move and you can throw something out or tear a muscle…It has definitely taken a toll on my body and now I'm paying the price with bad ankles, knees, shoulders, and back. There are constant repetitions required to perform the job and everything you need to work on is either above shoulder level or below your waist so your are constantly reaching or on your knees. 20 years of flight-line aircraft maintenance will definitely take a toll on your body no matter who you are.
Strains working with parachutes	Descriptions of awkward positions and other issues when handling parachutes	7	Aircrew Flight Equipment	Packing personnel parachutes requires a lot of awkward positions and also requires a lot of moving after you are done packing.
Height and weight restrictions	Recommendations about the required height and/or weight of personnel working in the AFS	6	Aircrew Flight Equipment and Aircraft Fuel Systems	For AFS MAFS there should be height and weight requirements for the job. i.e. anyone over 5'10 inches tall should NOT be allowed to be in this career field. Also anyone under that height and over 200 lbs should NOT be MAFS. Climbing through fuel tanks when you are too tall or overweight seems to make the possibility of getting stuck in a fuel tank much greater.
Increase strength requirements	Recommendation to increase entry strength requirements for the AFS	6	Explosive Ordnance Disposal, Aircraft Fuel Systems, and Aircrew Flight Equipment	Our job does not need [to] reduce the physical strength requirements. It needs to increase them. People coming into EOD should be able to lift the bomb suit and wear the bomb suit without a problem. The work place could be made safer by increasing the strength requirements of new airmen and ensuring that proper assistive equipment is available.

NOTES: AFSs are listed in descending order based on the number of responses they contributed. Only AFSs that made up at least 10 percent of comments for the category are listed.

As with the other open-ended questions on the survey, some respondents provided comments that were either unclear or did not fit well into any specific category. For the final survey question, 27 responses were placed into this miscellaneous category. Topics ranged from comments about the survey itself (7 responses) to comments that men should only be allowed in the AFS (2 responses). Other than a few complaints about the length of the survey, one type of survey critique provided a specific suggestion worth noting. More than one respondent stated that the survey did not address physical endurance (e.g., needing to walk for hours on a mission), which is an important aspect of their jobs. Indeed, this survey was focused on moving and handling objects, rather than aerobic or other types of physical stamina. For this reason, future surveys of the physical demands of airmen's jobs should consider adding specific questions about other lower-body strength.

Lastly, many responses to this final survey item touched on a topic that was explored in greater detail in an unanalyzed section of the survey asking about how strength demands could be reduced. For example, respondents suggested the following general ways to reduce the strength requirements:

- Technology (e.g., using lighter, newer gear)
- Workload (e.g., reducing work hours and/or increase manning levels)
- Personal gear (e.g., reducing the amount of gear)
- Workspace improvement.

or offered the following explanations for why physical demands cannot be reduced:

- Having sufficient strength is part of job qualifications; manual labor is needed.
- Strength requirements are already low for the job.
- Heavy protective gear/using heavy tools is a necessary part of the job or certain missions.
- Equipment/assistance (e.g., cart) is already in use.
- Confined workspace (e.g., fuel tank) precludes use of assistance.

Although we did not analyze that section of the survey (due to resource constraints), we provide the following example comments that illustrate the types of responses that were provided.

> give us different gear. There is absolutely NO REASON to be wearing 60+lbs of gear here. By giving us different gear (i.e lighter/less restricting) we will have less back and joint problems and MORALE would go WAY up.

> Newer equipment, that's not made in the 60's and 70's, with more modern technology which would reduce the size and weight of the equipment.

> Rolling Stock to move equipment. Ramps to get equipment in and out of ISUs.

> Need more manning! Low manned and requires a lot of work from a few people. When they cut manning they did not factor in ORM. People are worn out and working 12+ hours daily to meet mission requirements.

> Stage deployment gear and equipment at forward deployed locations to avoid one person having to transport 3-5 bags weighing 60-110 lbs to a deployment.

> Everything is done appropriately already. If something needs to be lifted by two people, then two people will lift it.

The hazards presented in the environment we operate preclude more personnel on-scene. Individuals must be able to handle equipment and ordnance solely, as there will be no one else within the hazard zone.

Our job requires moving communications equipment worldwide to include flightlines, parking lots, and FOBs. Extra equipment to do the job would require more time handling the equipment than we currently use.

There is no way to substitute the human factor when moving or lifting items. There will always be some injuries somewhere and we will always try our best to prevent them, but we are human.

Working with and around aircraft, not much you can do, if you require more than one person for more parts, requires more people for a job and less gets done.

Appendix F. Population and Sample Characteristics for Strength Requirements Survey

Table F.1 shows the survey population, invited sample (including individuals whose emails were not delivered), and number of respondents (i.e., the number of participants who completed demographic questions – pay grade, gender, AFS, and skill level). The number of respondents totals 3,012. For all career fields, except Security Forces, we included everyone in each of the four skill levels who had email addresses on record in the personnel files. Discrepancies between the Invited Sample and the Population Size listed in Table F.1 for all career fields except Security Forces occurred because some members of the career field did not have an email address listed on file.

For Security Forces, however, we used disproportionate random stratified sampling because the career field was so large that a census would not be necessary. We stratified by skill level (3, 5, 7, and 9) and gender (male and female). The sample size needed per stratified subgroup was estimated as 1,100 people for all subgroups. To arrive at this number, we first estimated that approximately 200 respondents per stratified subgroup would be needed for 80 percent power to detect a difference of 20 percentage points. Next, we considered our expected response rates. Recent Air Force surveys have achieved response rates around 30 percent. However, we expected that a survey of strength requirements would be less intrinsically interesting than many Air Force surveys (covering topics such as language learning or sexual harassment), and we planned for a 20 percent response rate. Therefore, to achieve a response rate of 200, we would need to invite a total of 1,000 people to participate. Lastly, we expected a small proportion of the email accounts we have on record would no longer exist and would result in bounceback messages. We estimated the percentages of bouncebacks at 7 to 10 percent. This brought our total invitation sample to approximately 1,100 people per subgroup to ensure that we would ultimately have 200 respondents per subgroup. In any subgroup with fewer than 1,100 people, we invited everyone.

Table F.1
Population, Invited Sample, and Response Sizes by Subgroup

		Men			Women			Total		
AFS	Skill Level	Population Size	Invited Sample	Response n	Population Size	Invited Sample	Response n	Population Size	Invited Sample	Response n
AFE	3	698	560	121	128	125	40	826	685	161
	5	735	734	185	217	217	64	952	951	249
	7	540	539	192	98	98	28	638	637	220
	9	48	48	20	3	3	2	51	51	22
AFU-AS[a]	3	521	508	73	22	21	6	543	529	79
	5	511	511	144	13	13	6	524	524	150
	7	317	317	113	22	22	8	339	339	121
AP-TTP[b]	3	309	292	72	22	22	6	331	314	78
MAFS[a]	3	492	468	107	49	49	11	541	517	118
	5	795	795	187	61	61	21	856	859	208
	7	478	477	164	18	18	8	496	495	172
CS[a]	3	146	142	25	49	48	16	195	190	41
	5	373	373	75	131	131	21	504	504	96
	7	357	357	97	119	119	41	476	476	138
EOD	3	352	180	44	25	12	2	377	192	46
	5	431	431	127	31	31	8	462	462	135
	7	285	285	113	8	8	4	293	293	117
	9	21	21	10	0	0	0	21	21	10
SF	3	11,068	1,100	80	2,457	1,100	71	13,525	2,200	151
	5	7,866	1,100	117	1,384	1,100	115	9,250	2,200	232
	7	2,938	1,100	218	275	275	59	3,213	1,375	277
	9	198	198	48	16	16	2	214	214	50
SS	3	109	96	10	114	105	19	223	201	29
	5	135	134	23	156	156	31	291	290	54
	7	98	98	24	80	80	29	178	178	53
	9	9	9	3	4	4	2	13	13	5

NOTE: Response n is the size of the largest sample of respondents used in any of our analyses.
[a] The AFU-AS, MAFS, and CS AFSs did not have any nine skill level personnel.
[b] The AP-TTP AFS subspecialty (shred) is only open to personnel in the one and three skill levels.

Statistical Weights for the Security Forces Sample

As described previously, we used a disproportionate stratified random sampling procedure to select active duty personnel in the Security Forces specialty to invite for the survey. Because of the survey's low participation and completion rates, we examined whether differences in completion rates in the two major sections of the survey—the Action Section and Movement Type Section—had any impact on results. After we discovered that survey completion rates affected survey results for those two sections, we decided to compute two sets of statistical weights for the Security Forces sample. The two weight sets correspond to survey participation at two critical points in the survey. This means that the weight groups are nested such that the second group is a subset of the first group. The statistical weight sets are defined as follows:

5. Survey Start Set. Participants who reached the Background Characteristics Section (i.e., the first page with questions). This group had 710 participants and was used for analyses involving the Strength-Requirements Screener and Action Section.
6. Movement Type Section Set. Participants available as of the Movement Type Section. This group had 544 participants and was used only for analyses in the Movement Type Section.

We computed the weights as follows: [population N for the stratified group] / [number of respondents in the stratified group within the weight set]. For example, 83 of the 117 male 5-levels in Security Forces who participated in the survey made it to the Movement Type Section. The population of male 5-levels in Security Forces was 7,866. Therefore, the sampling weight for each male 5-level in the Movement Type Section Group was approximately 94.77 (i.e., calculated as 7,866/83).

Unless otherwise noted, we use statistical weighting for all analyses involving Security Forces respondents and no statistical weights for the other AFS samples.

References

Air Force Enlisted Classification Directory (AFECD) (2011). *The Official Guide to the Air Force Enlisted Classification Codes* (corrected copy as of April 18, 2011).

Air Force Recruiting Service Instruction 36-2001 (2012). *Personnel Recruiting Procedures for the Air Force.*

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing,* Washington DC: American Psychological Association.

Arnold, J. D., J. M. Rauschenberger, W. G. Soubel, and R. M. Guion (1982). Validation and utility of a strength test for selecting steelworkers. *Journal of Applied Psychology*, 67(5), 588–604.

Aume, Nilss M. (1984). *A Machine for Weight-Lift Testing,* Wright-Patterson Air Force Base, OH: Air Force Aerospace Medical Research Laboratory, AFAMRL-TR-84-040.

Ayoub, M. M., B. C. Jiang, J. L. Smith, J. L. Selan, and J. W. McDaniel (1987). Establishing a physical criterion for assigning personnel to U.S. Air Force jobs, *American Industrial Hygiene Association Journal,* 48(5), 464–470.

Bernard, B. P. (ed.) (1997). *Musculoskeletal Disorders and Workplace Factors*, Cincinnati, OH: National Institute for Occupational Safety and Health (NIOSH).

Blakley, B. R., M. A. Quiñones, M. S. Crawford, and I. A. Jago (1994). The validity of isometric strength tests, *Personnel Psychology*, 47(2), 247–274.

Brannick, Michael T., Edward L. Levine, and Frederick P. Morgeson (2007). *Job and Work Analysis: Methods, Research, and Applications for Human Resource Management*, Thousand Oaks, CA: Sage Publications, Inc.

Brock, J. R., and S. J. Legg (1997). The effects of 6 weeks training on the physical fitness of female recruits to the British Army, *Ergonomics,* 40(3), 400–411.

Cizek, G. J. (ed.), (2001). *Setting Performance Standards: Concepts, Methods, and Perspectives*, Mahwah NJ: Lawrence Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.

Dorans, N. J. (1990). Equating methods and sampling designs. *Applied Measurement in Education,* 3(1), 3.

Equal Employment Opportunity Commission (1978). Civil Service Commission, Department of Labor, and Department of Justice, *Uniform Guidelines on Employee Selection Procedures*, Federal Register, Vol. 43, pp. 38290–38315.

Gebhardt, D. L., and T. A. Baker (2011). Physical performance. In John C. Scott, Douglas H. Reynolds, Allan H. Church, eds., *Handbook of Workplace Assessment: Evidence-Based Practices for Selecting and Developing Organizational Talent*, San Francisco, CA: Jossey-Bass, pp. 165–196.

Gibbons, L. E. (1989). *Summary of Ergonomics Research for the Crew Chief Model Development.* Dayton, OH: University of Dayton Research Institute, Report No. AAMRL-TR-90-038.

Government Accountability Office (1996). Physically demanding jobs: Services have little data on ability of personnel to perform. *Report to the Chairman, Subcommittee on Military Personnel, Committee on National Security, House of Representatives.* July, GAO/NSIAD-96-169.

Gutman, A. (2012). Legal constraints on personnel selection decisions. In Neal Schmitt (ed.), *The Oxford Handbook of Personnel Assessment and Selection,* New York: Oxford University Press, 686–720.

Harman, E., P. Frykman, C. Palmer, E. Lammi, K. Reynolds, and V. Backus (1997). *Effects of a Specifically Designed Physical Conditioning Program on the Load Carriage and Lifting Performance Of Female Soldiers.* (Technical Report T98-24). Natick, MA: U.S. Army Research Institute of Environmental Medicine, November.

Henderson, N. D. (2010). Predicting long-term firefighter performance from cognitive and physical ability measures. *Personnel Psychology,* 63, 999–1039.

Hogan J., and A. Quigley (1994). Effects of preparing for physical ability tests. *Public Personnel Management,* 23(1): 85–104.

Hogan, J. C. (1991). Physical abilities. In Marvin D. Dunnette and Leaetta M. Hough, eds., *Handbook of Industrial and Organizational Psychology*, Vol. 2 (2nd ed.), Palo Alto, CA: Consulting Psychologists Press, , 753–831.

Hughes, M., Ratliff, R., Purswell, J., & Hadwiger, J. (1989). A content validation methodology for job related physical performance tests. *Public Personnel Management,* 18(4), 487.

Jackson, A. S., (2000). Physical test validation for job selection. In S. Constable and B. Palmer, (eds.), *The Process of Physical Fitness Standards Development,* Wright-Patterson AFB, OH: Human Systems Information Analysis Center, 139–177.

Knapik, J. J. (1997). The influence of physical fitness training on the manual material handling capacity of women, *Applied Ergonomics,* 28, 339–345.

Knapik, J. J., S. Darakjy, S. Scott, K. G. Hauret, S. Canada, R. Marin, F. Palkoska, S. VanCamp, E. Piskator, W. Rieger, and B. H. Jones (2004). *Evaluation of Two Army Fitness Programs: The TRADOC Standardized Physical Training Program for Basic Combat Training and the Fitness Assessment Program*, Aberdeen Proving Ground, MD: U.S. Army Center for Health Promotion and Preventive Medicine, 12-HF-5772B-04.

Knapik, J. J., B. H. Jones, M. A. Sharp, S. Darakjy, S. Jones, K. G. Hauret, and G. Piskator (2004). *The Case for Pre-Enlistment Physical Fitness Testing: Research and Recommendations,* Aberdeen Proving Ground, MD: US Army Center for Health Promotion and Preventive Medicine, 12-HF-01Q9D-04.

Knapik, J. J., J. E. Wright, D. M. Kowal, and J. A. Vogel (1980). The influence of US Army basic initial entry training on the muscular strength of men and women, *Aviation, Space and Environmental Medicine,* 51(10), 603–606.

Koym, K. G. (1975). *Development of Physical Demand Profiles for Four Airman Career Ladders,* Occupational and Manpower Research Division: Lackland Air Force Base, November 1975. AFHRL-TR-75-67.

McDaniel, J. W., R. J. Skandis, and S. W. Madole (1983). *Weight Lift Capabilities of Air Force Basic Trainees,* AFAMRL Report TR-83-0001 Wright-Patterson Air Force Base: Air Force Aerospace Medical Research Laboratory.

Messing, K., and J. Stevenson, (1996). Women in procrustean beds: Strength testing and the workplace. *Gender, Work & Organization,* 3(3), 156–167.

Morgeson, Frederick P., and Erich C. Dierdorff (2011). Work analysis: From technique to theory," in Zedeck, Sheldon, ed., *APA Handbook of Industrial and Organization Psychology, Vol. 2: Selecting and Developing Members of the Organization*, Washington, DC: American Psychological Association, pp. 1–41.

Myers, D. C., D. L. Gebhardt, C. E. Crump, and E. A. Fleishman (1984). *Validation of the Military Entrance Physical Strength Capacity Test,* Technical Report 610, Natick MA: U.S. Army Research Institute of Environmental Medicine.

Rayson, M. P. (1998). The development of physical selection procedures for the British Army. Phase 1: Job analysis and identification of criterion tasks. In M. A Hanson (ed.), *Contemporary Ergonomics* (London: Taylor & Francis), 393–397.

Rayson, M., D. Holliman, and A. Belyavin (2000). Development of physical selection procedures for the British Army. Phase 2: Relationship between physical performance tests and criterion tasks, *Ergonomics,* 43(1), 73–105.

Robertson, D. W., & T. T. Trent (1985). *Documentation of Muscularly Demanding Job Tasks and Validation of an Occupational Strength Test Battery* (*STB*). (NPRDC Technical Report No. 86-1). San Diego, CA: Navy Personnel Research and Development Center.

Robertson, I. T., and M. Smith (2001). Personnel selection, *Journal of Occupational and Organizational Psychology,* 74(4), 441–472.

Schmidt, F. L., and J. D. Hunter (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin,* 124(2), 262–274.

Schmitt, N., and R. Sinha (2011). Validation Support for Selection Procedures. In Sheldon Zedeck, ed., *APA Handbook of Industrial and Organizational Psychology,* Vol. 2: Selecting

and developing members for the organization, Washington, DC: American Psychological Association, 399–420.

Sharp, M. A., E. A. Harman, B. E. Boutilier, M. W. Bovee, and W. J. Kraemer (1993). Progressive resistance training program for improving manual materials handling performance. *Work,* 3(3), 62.

Sharratt, M. J., N. J. Ashton, and R.W. Norman (1984). *Validity of the Incremental Lifting Machine.* Waterloo, Ontario: Department of Kinesiology, University of Waterloo, DCIEM Report 84-R-71.

Society for Industrial and Organizational Psychology, Inc. (2003). *Principles for the Validation and Use of Personnel Selection Procedures* (4th ed.). Bowling Green, OH.

Sothmann, M. S., D. L. Gebhardt, T. A. Baker, G. Kastello, and V. A. Sheppard (2004). Performance requirements of physically strenuous occupations: validating minimum standards for muscular strength and endurance. *Ergonomics,* 47(8), 864-875.

Stevenson, J. M., D. R. Greenhorn, J. T. Bryant, J. M. Deakin, and J. T. Smith (1996). Selection test fairness and the incremental lifting machine. *Applied Ergonomics,* 27, 45–52.

Terpstra, D., A. A. Mohamed, and R. B. Kethley (1999). An analysis of federal court cases involving nine selection devices, *International Journal of Selection and Assessment*, 7(1), 26–34.

Teves, M., J. E. Wright, and J. A. Vogel (1985). *Performance on Selected Candidate Screening Test Procedures Before and After Army Basic and Advanced Individual Training.* Natick, MA: Army Medical Research and Development Command, Report ADA162805.

Truxillo, D. M., L. M. Donahue, and J. L. Sulzer (1996). Setting cutoff scores for personnel selection tests: Issues, illustrations, and recommendations, *Human Performance*, 9, 275–295.

Vickers, R. R., Jr., J. A. Hodgdon, and M. B. Beckett (2009). *Physical Ability-Task Performance Models: Assessing the Risk of Omitted Variable Bias* (Report No. 09-04). San Diego, CA: Naval Health Research Center.

Vickers, R. R. Jr., and Amanda C. Barnard (2010). *Effects of Physical Training in Military Populations: A Meta-Analytic Summary* (Report No. 11-17). San Diego, CA: Naval Health Research Center.

Williams, Kristen M. and Jennifer L. Crafts (1997). Inductive job analysis: The job/task inventory method." In Deborah L. Whetzel and George R. Wheaton, eds., *Applied Measurement Methods in Industrial Psychology*, Mountain View, CA: Davies-Black Publications, , pp. 51–88.

Williams, A. G., M. P. Rayson, and D. A. Jones (1999). Effects of basic training on material handling ability and physical fitness of British Army recruits, *Ergonomics,* 42(8), 1114–1124.

RAND PROJECT AIR FORCE

www.rand.org

$29.95

ISBN-10 0-8330-8435-6
ISBN-13 978-0-8330-8435-4
52995
9 780833 084354

RR-471-AF