IDA

INSTITUTE FOR DEFENSE ANALYSES

Understanding and Managing Military Personnel Quality

Stanley A. Horowitz
Cullen A. Roberts
Julie A. Pechacek

INSTITUTE FOR DEFENSE ANALYSES

Understanding and Managing Military Personnel Quality

Stanley A. Horowitz
Cullen A. Roberts
Julie A. Pechacek

This page is intentionally blank.

Over the last fifty years, military leadership and supporting research organizations have studied the drivers impacting the supply of military personnel from the civilian labor market. However, not much has been learned about what personal attributes of those in uniform contribute to superior performance. This limited understanding has likely hampered opportunities to focus recruiting, make informed choices on recruit attribute trade-offs, improve the assignment of personnel to specialties, improve training, select and develop leaders, and better design the experience mix of teams, units, and the force as a whole.

This paper reviews a selection of the literature on military personnel quality to illustrate patterns in the existing research and challenges posed by historical data and methodological limitations. At present, personnel quality research provides definitive answers to some questions— particularly those concerning the value of training—but limited, inconsistent, or potentially misleading answers to many others. Without definitive results from well-designed research, decision makers commonly rely on correlative observations or anecdotes when developing and executing personnel policies. As military personnel expenditures comprise a large share of the DOD budget, opportunities to improve the efficacy and efficiency of this investment interest all DOD leaders.

What does "quality" mean in a personnel context? Achieving maximal warfighting readiness and making efficient use of scarce human resources require a detailed understanding of how an individual's many attributes contribute to the performance of various tasks. Analytic techniques are sufficiently advanced to enable prediction of performance quality based on individuals' characteristics and other factors, many of which may be observed prior to accession. Diverse personal attributes, including experience, intelligence, education, training, physical health, integrity, attitudes, morale, grit, and mental health, may be relevant. The importance of any one attribute depends on the task evaluated across a diverse set of duties and occupations. Actionable research on quality cannot progress without detailed data on both individuals' attributes and task-specific performance outcomes, and a research environment structured to overcome methodological challenges encumbering prior research. Reaching this goal requires sustained partnership between research analysts and the military Services.

1. Value of Training and Experience

The literature on training represents the best of the research on military personnel quality. Studies in this area generally assess how a well-measured attribute (e.g., a particular training exercise) contributes to performing a relevant task (e.g., sonar equipment operation) using a reliable measurement of performance quality. Most training studies are constructed to produce methodologically sound conclusions (e.g., through the use of intentionally designed controlled experiments). This research demonstrates that training matters, and identifies some features that make certain training programs better than others.

For example, Fletcher's (2010) analysis of the Defence Advanced Research Projects Agency (DARPA) Digital Tutor program provides credible evidence in favor of alternative training

1

regimes.[1] DARPA developed and implemented a computer-administered, self-paced 16-week training course for Navy information technicians built upon instructional principles used by expert human tutors. Digital Tutor graduates performed actual repair tasks considerably better than both recent graduates of the traditional 35-week Information Technology Training Continuum course, and seasoned Navy IT specialists with an average of 9.2 years of fleet experience. This research reliably measures the impact of the Digital Tutor using an intentionally designed experiment.

Similarly, research evaluating the Combat Logistics Force (CLF) provides credible evidence that that smaller, more highly trained and experienced crews can provide better readiness at a lower cost.[2,3] Operated by the Navy's Military Sealift Command (MSC), the CLF implemented a crewing model that relies on task-specific expertise and on-the-job experience. This approach reduced crew size by two-thirds and crew costs by 60 percent, while providing higher availability levels than the similar Navy ships crewed by uniformed sailors. Because such dramatic cost reductions and availability increases followed the policy change, it can be interpreted as effective despite the lack of intentional experimentation. Although the CLF uses civilian crews, similar gains could be realized in the military workforce by mirroring the CLF emphasis on expertise.

Studies of Navy, Marine Corps, and Air Force air crew performance demonstrate large returns to training and experience. Such research has measured performance through expert assessments or objectively measured results. These include kill probabilities in instrumented air-combat maneuvering exercises, bombing accuracy, airdrop accuracy, torpedo exercise scores, the overall results of operational readiness evaluations for carrier-based squadrons, carrier landing grades, and accident rates.[4] Different analyses studied how the experience of those in various roles—pilots, navigators and co-pilots for airdrops, and the entire tactical team (including sensor operators) for torpedo exercises—impacts outcomes. Consistently, this research has found that both recent and total career flying hours predict better performance.

---

[1] J. D. Fletcher, *DARPA Education Dominance Program: April 2010 and November 2010 Digital Tutor Assessments*, IDA Document NS D-4260, (Alexandria, VA: Institute for Defense Analyses, February 2011).

[2] Carol S. Moore et al., *Inside the Black Box: Assessing the Navy's Manpower Requirements Process*, CRM D0005206.A2 (Arlington, VA: Center for Naval Analyses, March 2002), https://www.cna.org/CNA_files/PDF/D0005206.A2.pdf.

[3] Anthony R. DiTrapani and John D. Keenan, *Applying Civilian Ship Manning Practice to USN Ships*, CRM D0011501.A2 (Arlington, VA: Center for Naval Analyses, May 2005), https://www.cna.org/CNA_files/PDF/D0011501.A2.pdf.

[4] Colin P. Hammon and Stanley A. Horowitz, *Flying Hours and Aircrew Performance*, IDA Paper P-2379 (Alexandria, VA: Institute for Defense Analyses, March 1990); Stanley A. Horowitz, Colin P. Hammon, and Paul R. Palmer, *Relating Flying-Hour Activity to the Performance of Aircrews*, IDA Paper P-2085 (Alexandria, VA: Institute for Defense Analyses, December 1987); Colin P. Hammon and Stanley A. Horowitz, *The Relationship between Training and Unit Performance for Naval Patrol Aircraft – Revised*, IDA Paper P-3139 (Revised) (Alexandria, VA: Institute for Defense Analyses, December 1996); Colin P. Hammon and Stanley A. Horowitz, *Relating Flying Hours to Aircrew Performance: Evidence for Attack and Transport Missions*, IDA Paper P-2609 (Alexandria, VA: Institute for Defense Analyses, June 1992); and Thomas E. Cedel, (Lt Col, USAF) and Ronald P. Fuchs (Lt Col, USAF), *An Analysis of Factors Affecting Pilot Proficiency* (Washington, DC: Air Force Center for Studies and Analyses, December 1986).

Despite these successes, efforts to determine the returns to experience and training often encounter methodological challenges that undermine or invalidate statements of *why* the observed outcomes occur. A fundamental question remains: Do personnel who remain have innate characteristics that enable their better performance, or does remaining longer cause improvements in quality? Individuals decide whether to enter and remain in military service in a manner that is likely correlated with their abilities and other features of interest to personnel managers. Selective entry and attrition complicate attempts to estimate the retention impact of attributes correlated with attrition. Because the existing literature on military personnel experience and training rarely accounts for this selection, those studies cannot determine why such a powerful relationship exists between investments, such as experience and training, and performance outcomes. Premature answers to such questions can lead to suboptimal policy.

There are many examples of researchers mistakenly interpreting descriptive research as causal evidence. For instance, RAND documented the higher productivity of more senior enlisted personnel in the Air Force.[5] It found that the most junior personnel take about 2.4 times more personnel hours to perform a fixed amount of sophisticated troubleshooting than personnel in the most senior manpower category. As a descriptive observation, the results cannot be contested. But without the ability to interpret the results in an "if-then" fashion, it is unclear how these observations should inform policy.

In another example, the Center for Naval Analyses (CNA) considered how the returns to experience could be leveraged to improve efficiency, using experience-productivity curves to argue that optimal policies in eight occupational areas would generate savings of from 2 to 18 percent by reducing personnel in the first eight years of service by 8 to 22 percent.[6] Likewise, research by the Congressional Budget Office (CBO)[7] used experience-productivity profiles to argue that transitioning to a high-experience military force would reduce personnel costs by 2.2 percent. However, these studies overstate the effects of increasing the military experience profile if doing so entails retaining more low-productivity individuals who would otherwise attrite.

Research analysing unit performance at the National Training Center found that the amount of training at the home station strongly related to the ratings units received from observer-controllers. Miles driven in train-up predicted performance for both force-on-force offensive missions and live-fire defensive missions.[8] It would be tempting to conclude that similar gains could be achieved in other units by increasing miles driven. However, miles driven in train-up may

[5] S. Craig Moore, *Demand and Supply Integration for Air Force Enlisted Work Force Planning: A Briefing*, N-1724-AF (Santa Monica, CA: The RAND Corporation, 1981).

[6] Ellen Balis, *Balancing Accession and Retention: Cost and Productivity Tradeoffs*, Professional Paper 380 (Arlington, VA: Center for Naval Analyses, 1983).

[7] Congressional Budget Office, *Setting Personnel Strength Levels: Experience and Productivity in the Military*, (Washington, DC: Congressional Budget Office, 1987).

[8] J. Ward Keesling, et al., *The Determinants of Effective Performance of Combat Units at the National Training Center (*Army Research Institute, June 1992).

reflect equipment reliability or crew skill, which would independently affect National Training Center performance.

Unfortunately, few studies have been able to address the underlying data and study design challenges that impede causal interpretation of their findings. Consequently, the research community supporting DOD has not learned enough from these studies.

2. Value of Innate Abilities

The selection bias pitfalls that complicate the valuation of training and experience also contaminate estimates of how much intelligence and social-emotional skills contribute to the performance of military members. If a feature that affects performance also affects whether a person remains in service or receives a given occupational assignment, then the observed relationship between performance and that feature will entangle the feature selection impact with its pure causal impact. Methodological challenges of this type likely contribute to the inconsistency of existing research findings on the performance value of intelligence and social-emotional skills. See the appendix for an illustration of this selection bias.

a. Intelligence

Following research demonstrating that Armed Forces Qualification Test (AFQT) scores and years of education both predict who completes the first year of service in the Navy,[9] the Navy began using these features for recruitment planning and selection. Subsequent research extended the Success Chances of Recruits Entering the Navy (SCREEN) type of analysis to better understand the determinants of first-term completion in a range of Navy occupations,[10] illuminating subtle differences in which characteristics best predict term completion in various occupations. All four military services currently use similar selection tools. Scores on different components of the AFQT currently determine eligibility for military occupations.

While these innovations have been instrumental in improving the effectiveness and professionalism of the force, they complicate subsequent analyses of the performance value of any attribute used to determine accession or assignment. Research linking AFQT scores to task performance does not account for the assignment mechanism, and has produced inconsistent results. For example, some studies of maintenance personnel find that AFQT scores predict greater readiness of either ships or aircraft, while others do not.[11,12] Likewise, several papers have

---

[9] See, for example, Robert F. Lockman and Patrice L. Gordon, *A Revised SCREEN Model for Recruit Selection and Planning*, CRC 338, (Arlington, VA: Center for Naval Analyses, August 1977).

[10] James S. Thomason, *Rating Assignments to Enhance Retention*, CRC 426, (Arlington, VA: Center for Naval Analyses, February 1980).

[11] A. J. Marcus, *Personnel Substitution and Navy Aviation Readiness*, Professional Paper 363, (Arlington, VA: Center for Naval Analyses, October 1982).

[12] Stanley A. Horowitz and Allan Sherman, *Crew Characteristics and Ship Condition*, CNS 1090, (Arlington, VA: Center for Naval Analyses, March 1977).

examined the interaction between equipment complexity and the AFQT scores of maintainers or operators, finding that higher AFQT matters more in more technical occupations. These considerations are also important to assessing potential interactions between abilities and technologies. Army research found that tank commanders and gunners with higher AFQT scores obtained much higher tank gunnery scores than those with lower AFQT scores when operating older M-60 tanks, and nearly identical gunnery scores when operating newer M-1 tanks.[13] It is likely that the M-1's more intuitive technology reduces cognitive load, which disproportionately benefits personnel with lower AFQT scores. However, it would be premature to conclude that the M-1's design eliminates the benefit of higher AFQT scores. Selection into military occupations will tend to generate negative correlation between AFQT scores and other (often hard to measure) beneficial traits; if these other beneficial traits become relatively more important, then AFQT scores can actually predict worse performance, even if their effect remains positive.

The interplay between intelligence, experience, and performance also deserves further examination. Analysis of National Training Center data indicate that higher AFQT scores predict better performance for inexperienced soldiers, but do not predict better performance among those with at least 20 months of experience.[14] This result might indicate that sufficient experience will substitute for aptitude—or it might indicate that those with lower AFQT scores, who nevertheless persist in the military for more than 20 months, have other attributes that compensate for their lower aptitude. Careful study design can identify and disentangle these contributors, and enable more effective and efficient personnel-resourcing policies.

b. Social-emotional skills

While the military has studied cognitive intelligence and its elements for many decades, personality traits and related social-emotional skills have received considerable attention only recently. The Tailored Adaptive Personality Assessment System (TAPAS) measures such factors, and has enabled a new stream of research. Some research indicates that the TAPAS can help predict who will perform well in an Army Special Operations Forces assessment and selection course.[15] Other research indicates that attributes identified by TAPAS complement AFQT scores in predicting retention[16]. Research assessing the value of these "non-cognitive" skills holds great promise and deserves further exploration. In this nascent research area, researchers and study

---

[13] Barry L. Scribner, et al., *Are Smart Tankers Better Tankers: AFQT and Military Productivity*, (Office of Economic and Manpower Analysis, Department of Social Sciences, United States Military Academy, December 1984).

[14] Keesling, et al., *The Determinants of Effective Performance of Combat Units at the National Training Center.*

[15] Christopher D. Nye, et al., *Assessing the Tailored Adaptive Personality Assessment System for Army Special Operations Forces Personnel*, (U. S. Army Research Institute, January 2014).

[16] White, Leonard A., Michael G. Rumsey, Heather M. Mullins, Christopher D. Nye, and Kate A. LaPort., *Toward a New Attrition Screening Paradigm: Latest Army Advances,* Military Psychology 26, no. 3 (May 1, 2014): 138–52.

sponsors have the opportunity to carefully design their studies to overcome ability-correlated selection bias.

3. Data Challenges in Assessing and Studying Performance Drivers

The lack of direct, consistent, and generalizable measures of Service member performance quality significantly hinders efforts to identify and quantify performance drivers. Some research substitutes speed of promotion for performance quality.[17] While expedient and analytically tractable, this strategy permits few strong conclusions for two reasons. First, it cannot evaluate whether the military is promoting or retaining the right people; doing so would result in circular reasoning. Second, because promotion depends on more than personnel quality—for example, the military's need to fill higher-ranking positions, or changes in the force structure—it cannot indicate how aggregate quality differs across units or over time. Ultimately, research identifying drivers of time-of-promotion illuminate what current exams and promotion boards *actually* prioritize— which is not necessarily what they *should* prioritize. Development and sharing of direct, consistent, and generalizable measures of Service member performance quality would enable significant progress toward identifying and quantifying performance drivers.

4. Final Remarks and Suggested Next Steps

Synthesizing the current literature on Service personnel attributes, we observe the following:

- Individual and unit characteristics matter to performance.
- AFQT, high school graduation, and attitudinal variables predict early attrition.
- Entry test scores predict promotion speed, which is not a satisfying performance metric.
- Experience and AFQT matter to performance, but it is unclear how they relate.
- Modern individual training can make a big difference. So can better collective training.
- Much of the relevant research is old. The nature of key relationships is likely similar today, but specifics need re-examination.

While notable progress has been made, significant portions of the existing research on the impact of the individual attributes, training, and experience of military personnel suffers from methodological challenges that limit interpretation. The best examples of this literature are studies testing experimental training regimes against the status quo. In other cases, the existing studies can be interpreted predictively: experience strongly predicts better performance; AFQT predicts faster promotion; AFQT and TAPAS scores predict lower first-term attrition; and AFQT sometimes predicts better performance of some tasks. Although prediction tells us what is likely to occur, it cannot explain why, and is therefore of limited value to military leadership. The following questions illuminate areas where future research and policy initiatives might bear significant fruit:

---

[17] Hosek, James and Michael G. Mattock, *Learning About Quality: How the Quality of Military Personnel Is Revealed Over Time*, (Santa Monica, CA: The RAND Corporation, 2003) and Peggy Golfin, Warren Sutton, and David Gregory, *Sailor Quality Metrics for Billet-Based Distribution,*(Arlington, Virginia, CNA, July 2015), 4.

- What are the true values of cognitive intelligence, social-emotional skills, and experience in producing Service-relevant performance (i.e., force readiness and lethality)? How do these individual attributes interrelate? How can they be enhanced or substituted?
- How can we better screen recruits and assign them to occupations?
- Can we develop better mechanisms to identify and promote the most promising people?
- How can we better leverage technologies in training and to improve performance?

To provide the military reliable answers to these questions and others, research on military personnel quality must demonstrate the relative return to various personal attributes using methodologies that can support causal interpretation. Accomplishing this requires comprehensive data, deep institutional knowledge, and a research environment structured to overcome methodological challenges, including the willingness to engage in carefully-designed experimentation when necessary. Strong commitment at the Office of the Secretary of Defense and Service levels can create conditions needed to overcome these methodological challenges.

References

Balis, Ellen. *Balancing Accession and Retention: Cost and Productivity Tradeoffs*. Professional Paper 380. Arlington, VA: Center for Naval Analyses, 1983.

Cedel, Thomas E. (Lt Col, USAF) and Ronald P. Fuchs (Lt Col, USAF). *An Analysis of Factors Affecting Pilot Proficiency*. Washington, DC: Air Force Center for Studies and Analyses, December 1986.

Congressional Budget Office. *Setting Personnel Strength Levels: Experience and Productivity in the Military*. Washington, DC: Congressional Budget Office, 1987.

DiTrapani, Anthony R. and John D. Keenan. *Applying Civilian Ship Manning Practice to USN Ships*. CRM D0011501.A2. Arlington, VA: Center for Naval Analyses, May 2005. https://www.cna.org/CNA_files/PDF/D0011501.A2.pdf.

Fletcher, J. D. *DARPA Education Dominance Program: April 2010 and November 2010 Digital Tutor Assessments*. IDA Document NS D-4260. Alexandria, VA: Institute for Defense Analyses, February 2011.

Golfin, Peggy, Warren Sutton, and David Gregory. *Sailor Quality Metrics for Billet-Based Distribution*. Arlington, Virginia: CNA. July 2015.

Hammon, Colin P. and Stanley A. Horowitz, *The Relationship between Training and Unit Performance for Naval Patrol Aircraft – Revised*. IDA Paper P-3139 (Revised). Alexandria, VA: Institute for Defense Analyses, December 1996.

Hammon, Colin P. and Stanley A. Horowitz. *Flying Hours and Aircrew Performance*. IDA Paper P-2379. Alexandria, VA: Institute for Defense Analyses, March 1990.

Hammon, Colin P. and Stanley A. Horowitz. *Relating Flying Hours to Aircrew Performance: Evidence for Attack and Transport Missions*. IDA Paper P-2609. Alexandria, VA: Institute for Defense Analyses, June 1992.

Horowitz Stanley A. and Allan Sherman. *Crew Characteristics and Ship Condition*. CNS 1090. Arlington, VA: Center for Naval Analyses, March 1977.

Horowitz, Stanley A., Colin P. Hammon, and Paul R. Palmer. *Relating Flying-Hour Activity to the Performance of Aircrews*. IDA Paper P-2085. Alexandria, VA: Institute for Defense Analyses, December 1987.

Hosek, James and Michael G. Mattock. *Learning About Quality: How the Quality of Military Personnel Is Revealed Over Time*. Santa Monica, CA: The RAND Corporation, 2003.

Keesling, J. Ward, Patrick Ford, Howard McFann, and Robert Holz. *The Determinants of Effective Performance of Combat Units at the National Training Center*. Army Research Institute, June 1992.

Lockman, Robert F. and Patrice L. Gordon. *A Revised SCREEN Model for Recruit Selection and Planning*. CRC 338. Arlington, VA: Center for Naval Analyses, August 1977.

Marcus, A. J. *Personnel Substitution and Navy Aviation Readiness*. Professional Paper 363. Arlington, VA: Center for Naval Analyses, October 1982.

Moore Carol S., Anita U. Hattiangadi, G. Thomas Sicilia, and James L. Gasch. *Inside the Black Box: Assessing the Navy's Manpower Requirements Process*. CRM D0005206.A2. Arlington, VA: Center for Naval Analyses, March 2002. https://www.cna.org/CNA_files/PDF/D0005206.A2.pdf.

Moore, S. Craig. *Demand and Supply Integration for Air Force Enlisted Work Force Planning: A Briefing*. N-1724-AF. Santa Monica, CA: The RAND Corporation, 1981.

Nye, Christopher D., Scott A. Beal, Fritz Drasgow, J. Douglas Dressel, Leonard A. White, and Stephen Stark. *Assessing the Tailored Adaptive Personality Assessment System for Army Special Operations Forces Personnel*. U. S. Army Research Institute, January 2014.

Scribner, Barry L., D. Alton Smith, Robert H. Baldwin, and Robert W. Phillips. *Are Smart Tankers Better Tankers: AFQT and Military Productivity*. Office of Economic and Manpower Analysis, Department of Social Sciences, United States Military Academy, December 1984.

Thomason, James S. *Rating Assignments to Enhance Retention*. CRC 426. Arlington, VA: Center for Naval Analyses, February 1980.

White, Leonard A., Michael G. Rumsey, Heather M. Mullins, Christopher D. Nye, and Kate A. LaPort, *Toward a New Attrition Screening Paradigm: Latest Army Advances*. Military Psychology 26, no. 3, May 2014.

Abbreviations

AFQT	Armed Forces Qualification Test
CBO	Congressional Budget Office
CLF	Combat Logistics Force
CNA	Center for Naval Analyses
DARPA	Defense Advanced Research Projects Agency
MSC	Military Sealift Command
OSD	Office of the Secretary of Defense
SCREEN	Success Chances of Recruits Entering the Navy
TAPAS	Tailored Adaptive Personality Assessment System
USAF	United States Air Force

Appendix

Notional figures 1 and 2 illustrate how selection can confound causal estimation. To highlight selection effects, we present a simplified model, abstracting from many details.

In Figure 1, some combination of grit and intelligence determines occupational assignment. Those low in both features are not accepted into the military; those with a high combination become pilots; and those with a moderate combination become mechanics. Consequently, mechanics (represented by the empty circles) are found along a band where grit and intelligence negatively co-vary.

If what determines quality performance were no different from what determines occupational assignment, then a low-intelligence mechanic would perform about as well as a high-intelligence mechanic. This is because any low-intelligence person who nevertheless becomes a mechanic must have some compensating strength – in this case, grittiness.

However, if what determines quality performance differs from what determines occupational assignment, then quality can vary along the band. In Figure 2, grit is relatively more important for quality performance than it is for occupational assignment. The coloring captures this: lighter-colored circles indicate higher-quality mechanics, and lighter-colored lines represent higher-quality isoquants.[18] Because grit is relatively more important for quality performance than it is for selection, the isoquants slope more steeply than the band of mechanics. Moving down the band crosses over isoquants, indicating increments in quality. Thus, because the assignment mechanism under-prioritizes grit, the best mechanics have the most grit and the least intelligence.

For mechanics under these conditions, the bivariate correlation between intelligence and performance would be negative, and a careless researcher might wrongly conclude that intelligence harms performance. To support the correct conclusions, research must account for the selection mechanism. Research in the quality literature rarely does so.

---

[18] An isoquant is a curve along which output (in this case, quality performance) does not change. Isoquants represent the relative productivity of inputs and can be thought of as thresholds, the crossing of which denotes increased output.
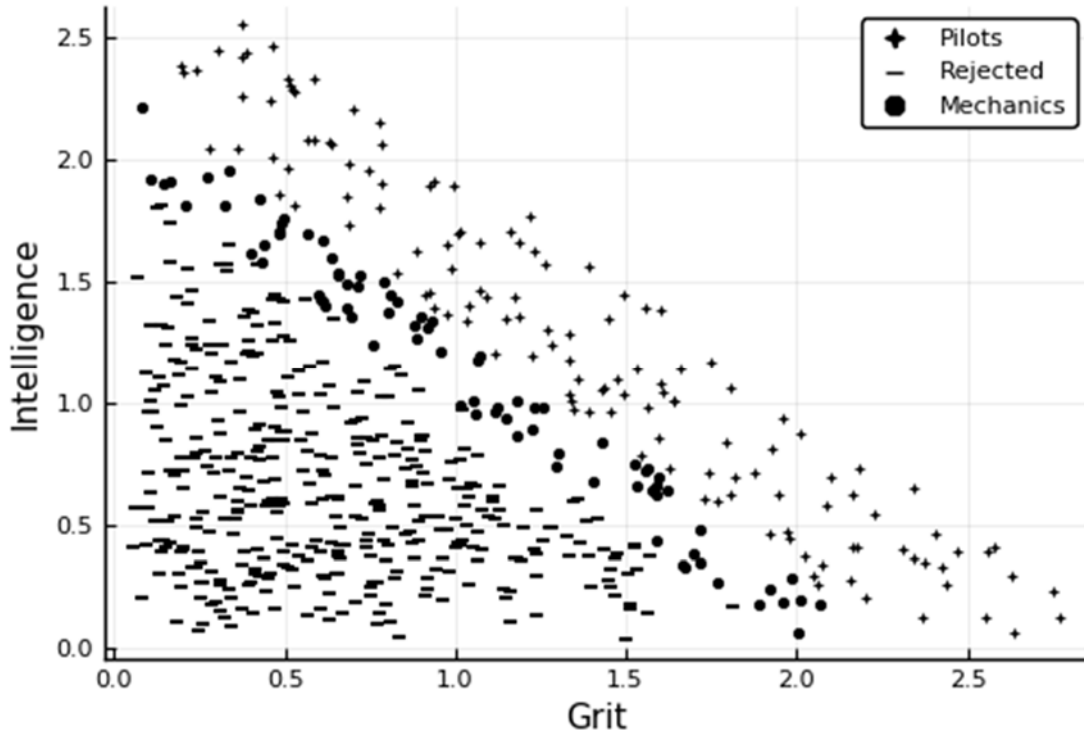
Figure 1. Selection causes occupation members to cluster in bands where traits negatively co-vary
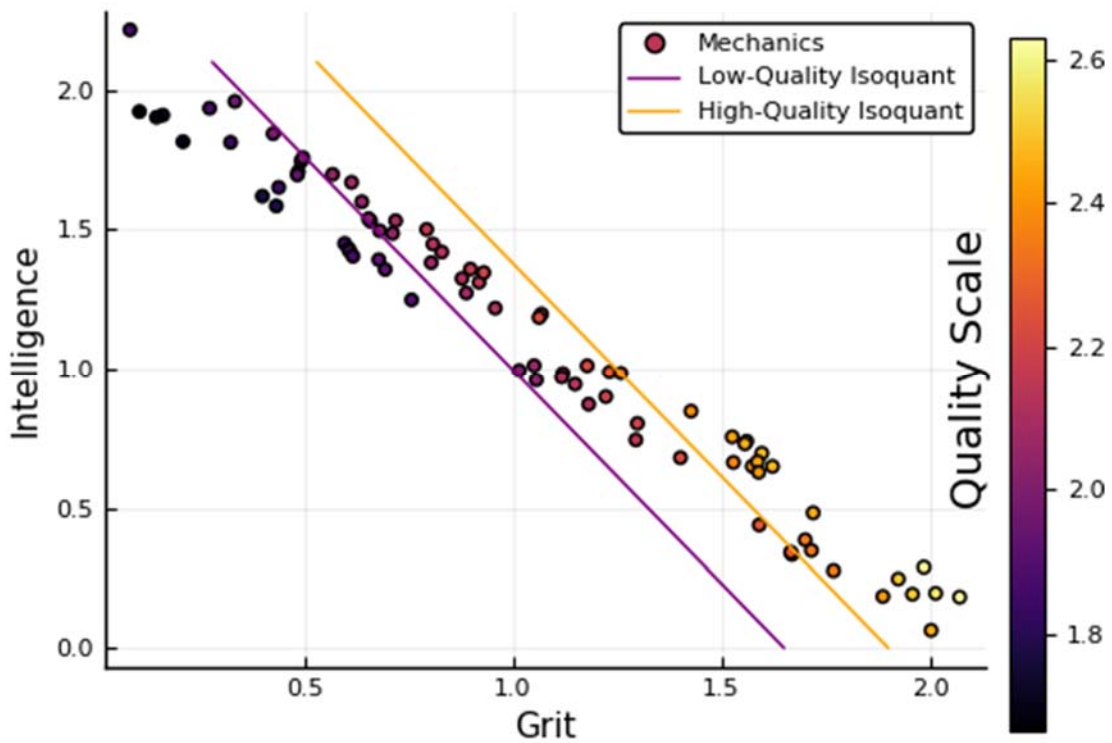

Figure 2. Quality varies along an occupational band if selection and quality functions differ

11

This page is intentionally blank.