Perspectives on Linking ASVAB's AFQT Scores to Scores from College Entrance Exams

Introduction

This paper provides an overview of linking, equating, and concordance processes and how they relate to the Armed Services Vocational Aptitude Battery (ASVAB). It details the appropriate uses and interpretations of concordance tables and explains when or whether it is appropriate to use them to compare ASVAB's AFQT scores to scores from popular college entrance exams like the Scholastic Aptitude Test **(**SAT) and the American College Testing (ACT).

Score Linking and Equating

Score linking is the process of placing scores from one test onto the same scale as another test. Numerous statistical techniques can be used to conduct score linking. There are linear techniques that ensure the two sets of scores have the same mean and standard deviation, and nonlinear techniques that transform one set of scores so that the entire distribution of scores from one test is equivalent to the distribution of scores from the other test. Regardless of the statistical method, the end goal is the placement of disparate sets of scores onto a common metric. Three excellent books provide more detailed information on the score linking process: *Linking Educational Assessments: Concepts, Issues, Methods, and Prospects* (Mislevy, 1992); *Linking and Aligning Scores and Scales* (Dorans et al., 2007); and *Test Equating, Scaling, and Linking: Methods and Practices* (Kolen & Brennan, 2014).

There are different *types* of score linking, such as *equating*, which is commonly used in educational testing. The goal of equating is to place scores from two different tests measuring the same construct on a common scale. Equated scores can be used and interpreted interchangeably. For example, imagine there are two test forms where one is slightly more difficult than the other. Without equating, examinees taking the more difficult test would be at a disadvantage. A successful equating process adjusts for the difference in difficulty between the two tests, so that it should not matter which test form an examinee takes. Once the scores are equated, it is as though the scores are from the same test form. Equating is regularly conducted for the ASVAB so that the scores for all examinees are on the same scale, regardless of the test items or test forms they take (Defense Manpower Data Center, 2008).

Making scores interchangeable is a weighty undertaking requiring strict adherence to these five primary requirements (Dorans & Holland, 2000):

1. The tests should measure the same construct. This means that the tests are built to the same specifications, in terms of content, test length, item type, etc. Essentially, the tests should be as identical as possible.

2. The tests should have the same reliability. Coupled with the first requirement, this means that not only do the tests measure the same thing, but they do so with equal precision.

3. The equating function should be symmetric. This means that converting scores from Test 1 to Test 2 should be in the inverse of converting scores from Test 2 to Test 1.

4. It should be a matter of indifference to an examinee as to which test he or she takes.

5. The equating function should be population invariant. For example, the equating function developed using an entire sample of examinees should be the same as the equating function developed using just females (or any other single demographic or subgroup).

When these requirements cannot be reasonably met, scores from two different test forms can still be statistically linked and placed on the same scale. However, the scores would not be equated and therefore could not be used interchangeably. The placement of scores from different tests onto a common scale, *without the accompanying interchangeability of scores*, is often called *concordance*.

Concordance

While linking scores via concordance has its requirements, they are less strict than the requirements for equating. *Concordance* has been defined as linking between forms/modes/tests that measure *similar* constructs at a *similar* level of reliability. Concordance also assumes similar difficulty and similar populations across the tests being linked (Holland et al., 2006; Holland, 2007). Essentially, the requirements for concordance are loosened versions of the equating requirements.

The statistical procedures used for equating are also generally used for concordance. The most commonly used procedure for establishing concordance utilizes an equipercentile function, a nonlinear method for mapping the score distribution from one test onto the score distribution of another test. Scores associated with the same percentile rank on each test are considered concordant. As an example, consider two hypothetical tests, A and B. The two tests measure similar content but have different score ranges. Scores from Test A range from 50 to 250, while scores from Test B range from 5 to 85. Using a group of examinees who have taken both tests, a researcher uses an equipercentile function to develop a concordance between the two tests, as shown in Table 1.

Table 1. Concordance Table for Test A and Test B

Test A	Test B	Percentile Rank
250	85	99th
225	70	90th
200	60	80th
175	55	70th
160	50	60th
150	45	50th
130	35	40th
110	30	30th
85	25	20th
65	15	10th

A score of 250 on Test A and a score of 85 on Test B both have a percentile rank of 99. That is, 99% of scores on Test A are at or below a score of 250, and 99% of scores on Test B are at or below a score of 85. Thus, these two scores (250 and 85) are concordant, given they share the same percentile rank.

One well-known concordance is between the SAT and ACT composite scores. It should be noted that these concordances do not mean the SAT and ACT scores can be used interchangeably. While the tests are similar in their design and purpose, they do not share the same test blueprint, nor are the populations who take each test necessarily the same. Thus, it is inappropriate to infer that a certain score on the SAT means the same thing as a corresponding score on the ACT. Instead, an appropriate interpretation of this concordance would be, "An SAT score of *x* that has a percentile rank of 90% (i.e., 90% of scores are at or below *x*) has a corresponding ACT score of *y* that also has a percentile rank of 90%." These two scores would be comparable in terms of their locations within their corresponding score distributions, but they would not be interchangeable.

Inappropriateness of an Armed Forces Qualification Test (AFQT) to SAT/ACT Concordance

Once concordance tables are developed, it is important to routinely update them, given any potential changes to the tests or changes in the population of examinees who take them. In the case of the ASVAB, concordance tables were developed in 2006 that linked the scores on the Armed Forces Qualification Test (AFQT), a composite of the verbal and mathematics subtests of the ASVAB, with score ranges on the SAT and ACT. However, given that both the SAT and ACT have undergone significant content, structural, and scoring changes since 2006, the existing (and outdated) concordance tables linking the AFQT to the SAT and ACT scores would provide an inaccurate "prediction," or picture, of how someone might perform on the other tests.

Even if new AFQT-SAT or AFQT-ACT concordances were to be constructed, AFQT scores could not be used interchangeably with SAT/ACT scores, nor would it be advisable to use them to predict SAT/ACT scores. One reason is that the ASVAB and the SAT/ACT were designed according to very different specifications. While there is some general content overlap (i.e., test items on verbal and math ability), the ASVAB content is not the same as, or even similar to, the SAT or ACT and likely differs in difficulty. Further, these tests were designed for very different purposes: while one could argue that both are used for making postsecondary educational decisions, ASVAB's AFQT score is used for determining qualification into the U.S. Armed Forces, while the SAT and ACT are used for determining admission to colleges and universities. Because of the differences in the difficulty, purposes, and content of these tests, it would be inappropriate to equate or concord scores on these tests. (More information about validated and approved uses of ASVAB's AFQT scores can be found in the Office of People Analytics' [2020] executive note/information paper, "Appropriate Use of ASVAB Scores.")

Another reason AFQT scores cannot be easily linked with SAT and ACT scores is due to differences in the examinee populations—both the population taking the tests as well as the population that would be used to develop the equating relationship. As noted previously, in order to develop a linking relationship, the populations of examinees must be similar; however, the population taking the ASVAB is generally different from the population taking the SAT or ACT. While a concordance between the AFQT scores and SAT/ACT scores could potentially be developed from a sample of convenience, this would require examinees who happen to take the ASVAB *and* either the SAT or ACT. The population of examinees that takes *both* the ASVAB and the SAT/ACT would likely be different from the population that takes *only* the ASVAB (or only the SAT/ACT)—the populations would likely include demographic or ability differences. Any resulting concordance would be applicable only to the convenience sample of those taking the ASVAB *and* either the SAT or ACT and would likely not generalize to the broader population of examinees who take only the ASVAB *or* the SAT/ACT.

Generally, if linked scores are used in place of actual scores to make decisions, this is a specialized type of score use that would need to be validated. While the AFQT, SAT, and ACT have been validated extensively for their particular uses specific to each test, the use of linked scores as a substitute for actual scores has never been validated (Pommerich, 2016). More information about validity and validated uses of ASVAB and AFQT scores can be found in the Office of People Analytics' (2024) executive note/information paper, "Validity Evidence Supporting the Uses of ASVAB Scores."

While concordance tables can serve as a bridge between tests that measure similar constructs, content, difficulty, and purposes, it is very easy to misinterpret or misuse these tables. Concordance tables derived from equipercentile linking, for example, simply show the corresponding scores from different tests that share the same percentile rank. They do not signify that scores from these tests can be used interchangeably, nor that one set of scores could or should be used to predict another. Doing so for the ASVAB would be an inappropriate and invalid use of these scores.

Conclusion

Although, in theory, any two sets of scores can be statistically linked, such as the scores from a geography test to scores from a dance competition, having the same corresponding percentile ranks provide little value when the test content or purpose (and subsequent scores) are unrelated. Knowing someone's level of knowledge in geography (i.e., geography score) would provide very little information about his or her dance ability (i.e., dance score). Given the significant differences between the AQFT and the SAT or ACT, as well as the risks associated with misinterpretation and misuse of concordance results, updated concordance tables between the AFQT and the SAT/ACT are no longer developed or provided. In addition, as both the SAT and ACT have undergone considerable content and structural changes since 2006, previous concordance tables should not be used as a proxy for current AFQT and SAT/ACT scores.

References

Defense Manpower Data Center. (2008). *CAT-ASVAB forms 5–9* (Technical Bulletin No. 3). Defense Manpower Data Center.

Dorans, N. J. (2004). Equating, concordance, and expectation. *Applied Psychological Measurement*, *28*, 227–246.

Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equitability of tests: Basic theory and the linear case. *Journal of Educational Measurement, 37*, 281–306.

Dorans, N. J., Pommerich, M., & Holland, P. W. (Eds.). (2007). *Linking and aligning scores and scales*. Springer-Verlag.

Holland, P. W. (2007). A framework and history for score linking. *Linking and Aligning Scores and Scales,* 5–30. Springer.

Holland, P. W., Dorans, N. J., & Petersen, N. S. (2006). Equating test scores. *Handbook of Statistics*, *26*, 169–203.

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer.

Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects.* ETS Policy Information Center.

Office of People Analytics. (2020). *Appropriate use of Armed Services Vocational Aptitude Battery (ASVAB) scores* [Executive note/information paper]. Office of People Analytics.

Office of People Analytics. (2024). *Validity evidence supporting the uses of ASVAB scores* [Executive note/information paper]. Office of People Analytics.

Pommerich, M. (2016). The fairness of comparing test scores across different tests or modes of administration. In N. J. Dorans and L. Cook (Eds.), *Fairness in Educational Assessment and Measurement*. Routledge.