

The Relationship Between ASVAB and Training School Performance for USMC Field Radio Operators

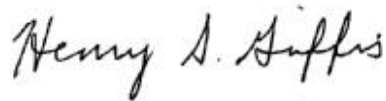
Catherine M. Hiatt



4825 Mark Center Drive • Alexandria, Virginia 22311-1850

Approved for distribution:

July 2005

A handwritten signature in black ink that reads "Henry S. Griffis". The signature is written in a cursive style with a large initial 'H' and 'G'.

Henry S. Griffis, Director
Workforce, Education and Training Team
Resource Analysis Division

CNA's annotated briefings are either condensed presentations of the results of formal CNA studies that have been further documented elsewhere or stand-alone presentations of research reviewed and endorsed by CNA. These briefings represent the best opinion of CNA at the time of issue. They do not necessarily represent the opinion of the Department of the Navy.

Cleared for Public Release: Distribution Unlimited. Specific authority: N00014-00-D-0700.
For copies of this document call: CNA Document Control and Distribution Section (703) 824-2123.

Copyright © 2005 The CNA Corporation



The Relationship Between ASVAB and Training School Performance for USMC Field Radio Operators

Catherine M Hiatt

27 July 2005

Summary



- ASVAB EL composite is a valid predictor of training performance for Field Radio Operators
- The EL composite is the best combination of ASVAB subtests for predicting training performance in this course
- There is equal prediction of training grades for racial and gender subgroups in this course

This slide summarizes the findings of our validation analyses for USMC Field Radio Operators. The Marine Corps uses the Armed Services Vocational Aptitude Battery (ASVAB) Electronics (EL) composite to select Marines for this occupation. We found that EL is still the best predictor of initial skills training performance for this occupation. We also found equal predictions of training grade in racial and gender subgroups.

Generalization



- This procedure could be completed on all training courses if:
 - All schools maintained final course grades
 - These grades were kept in a central training database

The general validation procedure presented here could be applied to every initial skills training course if the data were readily available. In general, it takes 2 to 3 years of training data to complete this type of analysis. Final course grades are necessary; dichotomous pass/fail or class standing information does not provide a rigorous enough measure of performance. If final course grades are assigned to students who fail the course they should also be retained. Course start and end dates or the total number of days required to complete the course should be recorded in the training grade database. Currently, school performance data are collected separately for each validation study, which usually means a substantial data collection period before any analysis can begin. Collecting these data on a regular basis and storing them in a USMC master file would allow for routine and less expensive validation.

Background



- The ASVAB comprises nine subtests that may be combined to create aptitude composites
- Composites are used to classify people with respect to military jobs
- As military jobs change, the classification composites may also need to change

The ASVAB is made up of nine subtests (listed on the next slide) that measure various aptitudes. Factor analyses conducted on the ASVAB subtests have shown a four-factor solution.^{1,2} The subtests can be associated with the following four content factors: verbal, math, technical, and spatial aptitude. Standard scores are computed for the subtests in the 1980 Youth Population. These standard scores have a mean of 50 and standard deviation of 10.

The standardized subtest scores are combined to create various aptitude composites. The Marine Corps uses four composites to assign Marines to the appropriate occupation.

This project will provide a basic procedure for validation of ASVAB against training school performance using the Field Radio Operators course as a reference.

1.P. H. Stoloff, *A Factor Analysis of ASVAB Form 8A in the 1980 Youth Population*, Aug 1983 (CNA Research Memorandum 83-3155).

2.M. J. Ree et al., *Armed Services Vocational Aptitude Battery: Item and Factor Analyses of Forms 8, 9, and 10*, 1982 (AFHRL, Brooks Air Force Base, TX: AFHRL-TR-81-55).

ASVAB subtests



Subtest		Content factor
General Science	GS	Verbal
Arithmetic Reasoning	AR	Math
Word Knowledge	WK	Verbal
Paragraph Comprehension	PC	Verbal
Auto Shop Information	AS	Technical
Mathematics Knowledge	MK	Math
Mechanical Comprehension	MC	Technical
Electronics Information	EI	Technical
Assembling Objects	AO	Spatial

These are the current ASVAB subtests and their associated content factors. The Assembling Objects (AO) subtest is a new experimental test in the battery. The Marine Corps does not use AO in any of its aptitude composites at present. We include AO in this analysis to explore the possibility that it may add unique information that may be helpful in predicting the training performance of field radio operators.

Data



- FY03 and FY04 course data
- Final course grades are required
- No curriculum changes
- ASVAB subtest scores

A total of 1,519 cases were collected from the Marine Corps training school for Field Radio Operators in Twentynine Palms. All students received a final course grade on completion of the course. No major curriculum changes that would affect comparing the grades across time occurred during the data collection.

The Defense Manpower Data Center (DMDC) provided the ASVAB scores. Scores could not be found for 72 cases, and an additional 73 cases did not have complete school data. Most of the records have the current configuration of the ASVAB, which includes the AO subtest and excludes the Coding Speed (CS) and Numerical Operations (NO) tests. We excluded from the study any people tested on the older configuration—before 2001. All ASVAB scores were reported before July 2004, so they have been standardized to the 1980 Youth Population.

Data quality



- Only first attempt in the course
- Course grades need to be consistent over time
- Identification of outliers

Our purpose was to predict initial skills training, so our analyses included only the first attempt made in the course. Only the first attempt through the course was retained for all Marines identified as recycles by the service school. Eight additional cases were dropped based on their service school record on the Headquarters Master File. The record showed they attended the course for Field Radio Operators on multiple occasions.

To rule out those who were making a lateral move into this military occupational specialty (MOS), we eliminated cases with more than 23 months of service before their class begin date. Marines with records indicating anything other than a first enlistment were also excluded.

We examined the final course grades by class to determine if any systematic shift in the grades occurred over time. There were no major changes over time.

Identification of outliers



- Extreme values on either course grade or aptitude information can be a problem
- Analysis can help identify outliers
 - Examine scatter plot
 - Examine standardized residuals from two separate regressions

Extreme values on either the predictor or the performance measure can influence the regression results. We examined a scatter plot of course grades by the EL composite, shown in the appendix. We also conducted regression analyses³ and examined the plots of the standardized residuals by predicted training grades to identify these outliers. Residuals are the difference between the actual training grade and the grade predicted by the regression model. These residuals are standardized to a mean of zero and standard deviation of one. In general, standardized residuals more than plus or minus three standard deviations are considered extreme. The appendix also includes the plot of the standardized residuals. The analysis identified seven extreme cases, which were eliminated from the analysis.

3. P. W. Mayberry and C. M. Hiatt. *Validation of Armed Services Aptitude Battery Against Training Performance*, Sep 1996 (CNA Research Memorandum 96-84).

Sample demographics



Subgroup	Description	Count
Race		
	White	676
	Black	160
	Hispanic	227
	Other/Unknown	97
Gender		
	Male	1,037
	Female	123
Total		1,160

These are the sample sizes after all the data editing. There are enough cases to complete the analysis for all subgroups. Only 17 cases were identified as not being high school graduates. This is not a large enough group for any reliable analysis. All statistics in this report were computed using SPSS software.

Subtest validities



Range-corrected subtest validities									
GS	AR	WK	PC	AS	MK	MC	EI	AO	VE
.57	.60	.55	.52	.49	.60	.55	.55	.38	.56

Once the data have been cleaned, the first step in the analysis is to examine the correlation between the ASVAB subtests and training school performance. This relationship is called validity. The size of the observed correlation between ASVAB scores and performance in a course-specific sample is lower than what would be expected in the entire pool of recruits available for assignment. This is because every course has minimum aptitude requirements. The minimum requirement for the Field Radio Operators Course is an ASVAB EL composite score of 90. Marines not meeting this requirement are excluded from the course and therefore not included in our data sample. The validities shown here have been corrected⁴ for this restriction of range using the 1980 Youth Population as a reference. (See the appendix for the observed correlations.) All the subtests have fairly high validities. AO is much lower than the other subtests.

4. T.L. Mifflin and S.M. Verna. *A Method to Correct Correlation Coefficients for the Effects of Multiple Curtailment*, Aug 1977 (CNA Research Contribution 336).

Stepwise regression



- Identifies which subtests are best predictors of training performance
- Considers intercorrelation among subtests
- Best predictors are combined to create composite

Next, we completed a stepwise regression to identify which combination of subtests would be the best predictor of initial skills training for the Field Radio Operator. This method allows for the interaction of subtests as they enter the regression equation. The results indicate which subtests should be used to create the best composite for predicting training performance.

Regression results



Subtests	Multiple R ²
AR	0.37
AR + EI	0.41
AR + EI + MK	0.43
AR + EI + MK + AS	0.44
AR + EI + MK + AS + VE	0.45

The results of the stepwise regression show that performance in the course is influenced by math (AR and MK) and technical (EI and AS) abilities as measured by the indicated subtests. Adding Verbal (VE) aptitude slightly increases the multiple R².

Developing best composite



- Best composite should have
 - High validity
 - Minimal differences in validity across subgroups
 - High reliability

Composites are built as the sum of unit weighted subtest standard scores. Various composites were computed using the subtests identified by the stepwise regression as having high validity for this course. Complete results are available in the appendix.

The best aptitude composite for predicting training performance should have high validity and should predict performance for all subgroups equally. A test should also provide a consistent measure of aptitude. This consistency is shown in the test reliability.

Composite definitions



Composite definition	Validity
AR + EI	0.63
AR + EI + MK	0.66
AR + EI + MK + AS	0.66
AR + EI + MK + GS*	0.66
* This is the current EL composite	

These are the validities for the composites based on the stepwise regression results. There is very little difference in the validities for these composites and no difference between the best combination of subtests based on the stepwise regression and the EL composite currently used to select Marines for this course. The Electronics composite is already known to have high reliability.⁵ If EL can also predict performance equally across subgroups, it would be the best choice for this course.

5. P. Palmer. *Armed Services Vocational Aptitude Battery (ASVAB): Alternate Forms Reliability (Forms 8, 9, 10, and 11)*, Mar 1988, AFHRL, Brooks Air Force Base, TX:AFHRL-TP-87-48.

Subgroup analyses



- Sample size was large enough for both racial and gender subgroup analysis
- Compare regressions for each subgroup
- Use true score estimates to account for measurement error

We completed subgroup analyses for racial and gender subgroups. To determine the fairness of the selection test for this course, we computed regressions for each subgroup. Then we compared the slopes, intercepts, and standard errors of measurement.⁶ If these statistics differed across groups, the selection test was considered biased against one or more groups. Using total sample results with a biased test would result in systematic errors in classification and assignment. The appendix shows possible outcomes across groups.

A basic assumption of regression analysis is that the predictor is measured without error. Aptitude tests are not perfect measures. To account for measurement errors in the predictor, we used true score estimates instead of the observed score in the regression analysis. The appendix contains details of this computation. For completeness, the analysis was also conducted using the observed scores.

6. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association, 1985.

Results of subgroup analysis



Score type	Subgroup	Slope F value	Intercept F value
True score estimate	Race	0.20	0.01
	Gender	0.03	0.60
Observed score	Race	0.16	0.89
	Gender	0.00	1.47

These are the results of the regression analysis to compare the slopes and intercepts for both the racial and gender subgroups. Neither the slopes nor the intercepts are statistically different for any of the subgroups. This is true for both the true score estimates and the observed scores. For these samples, F values greater than 3.0 and 3.8 would be needed for the racial and gender groups, respectively, to be statistically significant at the .05 level. Since these levels are not reached, the slopes and intercepts are found to be equivalent. This implies that the EL composite will predict training performance equally for whites, blacks, and Hispanics as well as for men and women.

Conclusions



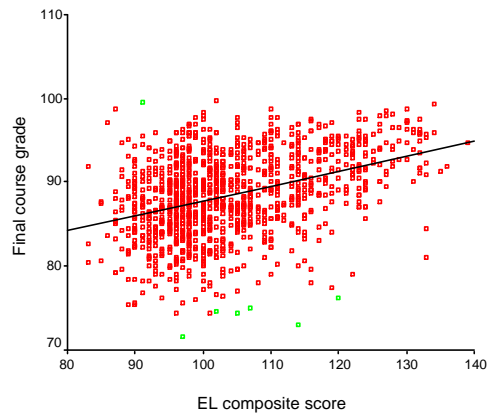
- The ASVAB is a valid predictor of initial skills training for USMC Field Radio Operators
- The EL composite remains the best composite for this course

The EL composite is a good predictor of training performance for this course. Since this composite has been shown to be a reliable predictor, and it is a fair predictor for both race and gender subgroups, it is still the best predictor composite for this course.

Appendix: Analysis details

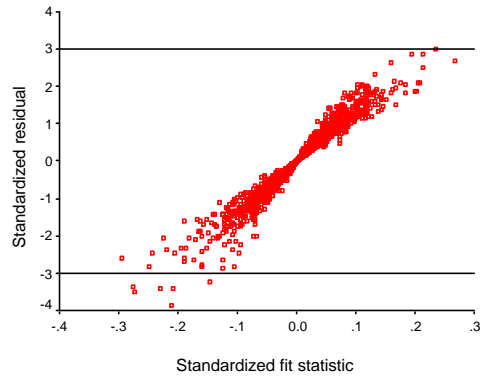


Scatter plot to identify outliers



This scatter plot shows the seven outliers identified in our analysis. The outliers are shown in green.

Plot of standardized residuals



This is the plot of the standardized residuals. Observations that exceeded plus or minus 3 for the standardized residual were considered outliers.

Uncorrected validities



Uncorrected validities								
GS	AR	WK	PC	AS	MK	MC	AO	VE
.26	.34	.24	.25	.26	.33	.26	.21	.27

These are the uncorrected subtest validities.

Experimental composites



Composite definition	Validity
AR + AS + MC	.62
AR + AS + EI	.63
AR + MC + EI	.63
MK + AS + MC	.64
MK + AS + EI	.64
MK + MC + EI	.65
AR + MK + EI + VE	.66
AR + MK + MC + GS	.66

These are other experimental composites created from the results of the stepwise regression. While they all have good validity, the last two, along with the EL composite, are better at predicting training school performance for this course.

Estimation of true scores



- $X_{\text{true}} = X_{\text{pop}} + R_{xx}(X_{\text{obs}} - X_{\text{pop}})$
 - X_{true} = estimated true score
 - X_{pop} = population subgroup mean score
 - X_{obs} = observed score
 - R_{xx} = test-retest reliability of test for subgroup

Estimated true scores are computed to account for measurement error in aptitude tests. Lord and Novick⁷ provide the equation to estimate the true scores.

7. F. M. Lord and M. R. Novick, *Statistical Theories of Mental Test Scores*. Reading, MA: Addison Wesley, 1974.

Statistics for the computation of true score estimates for the EL composite

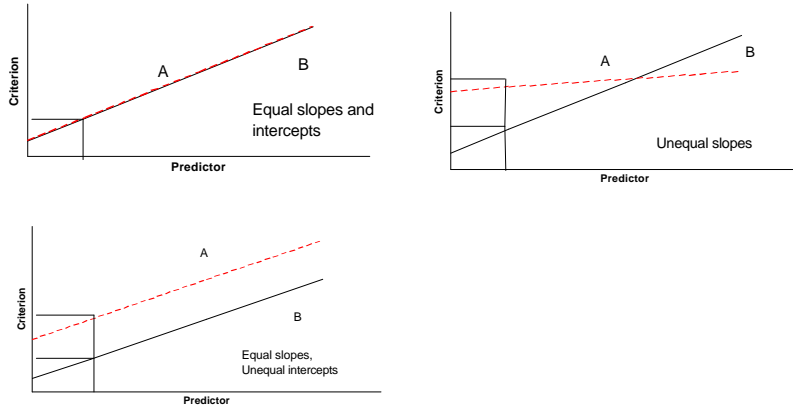


	Racial/ethnic subgroups			Gender subgroups	
	White	Black	Hisp	Male	Female
Reliability ⁸	0.92	0.88	0.87	0.93	0.90
Population Mean	104.3	81.4	85.6	104.2	95.7

These are the statistics needed to compute the estimated true score for the EL composite. All population numbers are from the 1980 Youth Population.

8. P. Palmer. *Armed Services Vocational Aptitude Battery (ASVAB): Alternate Forms Reliability (Forms 8, 9, 10, and 11)*, Mar 1988, AFHRL, Brooks Air Force Base, TX:AFHRL-TP-87-48.

Possible differential prediction outcomes



This slide shows the possible results of testing slopes and intercepts of two groups, A and B.

The first panel shows what equivalent slopes and intercepts will yield. Identical test scores will predict equal performance. This is the ideal situation.

The panel on the right illustrates the case of slopes that are determined to be different. Identical test scores do not predict the same performance level.

The lower panel shows the case of different intercepts but equal slopes. Here, the same test score will consistently predict higher performance for people in group A.

