THE EFFICACY OF BIOGRAPHICAL INVENTORY DATA IN PREDICTING EARLY ATTRITION IN NAVAL AVIATION OFFICER CANDIDATE TRAINING

D. R. Street, Jr., and D. L. Dolgin

Approved for public release; distribution unlimited.
The research reported in this paper was completed under the Naval Medical Research and Development Command Work unit number 63706N M0096.002-M00960.01.

The views expressed in this report are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.

Trade names of materials and/or products of commercial or non-government organizations are cited as needed for precision. These citations do not constitute official endorsement or approval of the use of such commercial materials and/or products.

Reproduction in whole or part is permitted for any purpose of the United States Government.
Attrition in the training of U.S. naval aviation officer candidates represents a historic problem. The early identification of those likely to attrite during training would significantly reduce overall training expenditures. In this study, we assessed the value of biographical information for predicting early attrition at the indoctrination level of naval aviation officer training. We selected a random sample of 1551 aviation officer candidates and naval aviation cadets for analysis. The subjects selected had taken the Aviation Selection Test Battery (ASTB) between 1987 and 1990 and had completed the aviation indoctrination program operated by the Naval Aviation Schools Command in Pensacola, Florida. A principal component factor analysis of Biographical Inventory items was conducted with those who passed \((N = 1176)\) and also with those who attrited \((N = 375)\) basic aviation indoctrination. The resultant factors were then forced into a discriminant function analysis to determine if the factors obtained were different for the two groups. We found that the factors were significantly different for the two groups. The results indicate that biographical data may be useful in identifying candidates who are most likely to attrite early from naval aviation training.
SUMMARY PAGE

THE PROBLEM

Attrition in naval aviation officer training represents a continuing problem. The attrition rate differs for the various accession pipelines ranging from roughly 5% for Naval Reserve Officer Training Corps and Naval Academy accessions to nearly 25% for aviation officer candidates and naval aviation cadets in the initial stages of preflight training. The early identification of those likely to attrite naval aviation training would reduce overall training costs.

FINDINGS

We found statistical differences in many Aviation Selection Test Battery (ASTB) Biographical Inventory (BI) responses for those who attrited compared to those who passed preflight training. The results indicate that the ASTB BI could be used to reduce the number of Aviation Officer Candidate (AOC) and Naval Aviation Cadet (NAVCAD) attritions by 50%. However, the cost of implementing this model is the non-selection of approximately 20% of those who would have otherwise passed as false rejections. We also found that some principal derived BI item groupings were statistically different for the two groups.

RECOMMENDATIONS

We believe that a scoring system for the ASTB BI derived by principal components technique could be useful in targeting AOC/NAVCAD candidates likely to attrite naval aviation training. A scoring key based on all scalable ASTB BI items should be considered for implementation as a method of attrition prediction in preflight AOC/NAVCAD training. The results are limited to linear composites formed by principal components. Nonlinear transformations may reveal important associations between ASTB BI and attrition.

Acknowledgments

The authors gratefully acknowledge the assistance of CDR Michael Pianka and LCDR Tommy Morrison of the Naval Aerospace Medical Institute.
INTRODUCTION

Attrition in the training of U.S. naval aviation officer candidates represents a historic problem. Attrition may occur at any point during the aviation officer training program ranging from basic aviation indoctrination through winging as naval flight officers and naval aviators. The identification of those who are likely to attrite during naval aviation training could reduce overall training expenditures. Recent figures from the Naval Aviation Schools Command demonstrate differential rates of attrition for various groups of officers and officer candidates in preflight training in Pensacola, Florida. Attrition rates for Naval Reserve Officer Training Corps (NROTC) and Naval Academy graduates have been low (roughly 5%), while rates for aviation officer candidate (AOC) and naval aviation cadet (NAVCAD) officer candidates have been comparatively high (roughly 30%). Various reasons for attrition have been indicated by students. These include failure to meet physical requirements, voluntary drop on request, not considered officer material, and failure to meet academic standards. The attrition of any candidate from the aviation training program represents a serious financial and manpower loss for the Navy at any stage of aviation training.

A means of reducing attrition, which is already available, could be the use of the Aviation Selection Test Battery (ASTB), especially the Biographical Inventory (BI), in identification of AOC and NAVCAD candidates who account for the single largest component of total naval aviation training attrition. Biographical inventories have been a component of naval pilot selection batteries since the 1940s. The Aviation Qualification Test/Flight Aptitude Rating (AQT/FAR) was the last revision, and has been used since the early 1970s. The ASTB is the latest revision of the AQT/FAR and will be operational in 1992. Various researchers have demonstrated the utility of the BI in earlier versions of the AQT/FAR (1-4), although a study by Booth and Peterson (5) revealed the lowest correlation with a pass/fail criterion for the BI compared with other naval aviation selection measures. A survey of naval aviation pilot candidates conducted by Bale and Ambler (6) found that the BI explained additional variance not accounted for by other aspects of the AQT/FAR against a pass/fail criterion. An earlier report by Berkshire (7) demonstrated that the BI had a higher correlation with a pass/fail criterion than the AQT but lower than the Mechanical Comprehension Test component of the battery then in use by the Navy. The majority of naval aviation selection research reviewed supports the continued use of the BI in pilot selection programs.

A study (8) on the validation of the BI of the ASTB used an empirical keying approach to item selection and weighting. The empirical keying method involves weighting of item responses according to the frequency of selection by a criterion group (i.e., those who passed flight training). The author (8) indicated that the ASTB BI provided improvements over the FAR/BI in predicting success in primary naval flight training. He (8) also found that a scoring key derived with selective weighting of various ASTB BI items accounted for 6.1% of the total predictive variance as compared to 4.0% for the best FAR/BI re-scoring keys. These results indicate that the ASTB/BI is more powerful than its predecessor in predicting naval flight training performance.

In other countries, biographical information has been used in the selection of NATO pilots (9). NATO countries in particular have relied heavily on biographical data in the pilot selection process since World War II (10). Biographical data are considered with other performance and personality testing data for a complete profile of a potential applicant's skills, abilities, motivation, and interest in an aviation career. The most recent trend in testing in most NATO countries has been the automation of selection testing to save time and money (10). Automated testing has also recently been adopted by the U.S. Air Force. The Air Force Basic Attributes Test (BAT) employs a computerized system where all testing is completed at a single computer terminal. The BAT is fully operational as the sole pilot selection system in the U.S. Air Force (10, 11). The selection system used by the U.S. Air Force does not currently include a biographical inventory.
A recent study (12) presented information regarding the use of the Armed Services Applicant Profile (ASAP) in recruit selection. The ASAP is a biographical measure of an applicant's probability to adapt to military life. When administered to more than 120,000 armed services applicants, the ASAP was a valid predictor of attrition for all groups without resulting in adverse impact on women or other minority groups. The ASAP also demonstrated validity in predictions of likelihood to complete first tour service. The results presented by Trent et al. (12) strongly support the validity of biographical measures in military selection and prediction of later performance.

In the civilian community, the use of background data in employee selection and job performance prediction has a strong historical basis. The problem in collecting this kind of information is in finding a format that minimizes self-report bias and reduces financial and manpower investment while retaining predictive validity (13). Of the two most common methods of collecting background data (i.e., interview and self-report testing), a self-report testing procedure that combines paper and pencil or automated testing with a multiple-choice format has proven the most cost effective in the civilian community (14). The multiple-choice format is also proven more valid than interview data, which may be subject to both interviewer and interviewee biases (15).

Although various researchers have suggested that multiple-choice measures of background data may be an economical and valid predictor of future job performance, problems in item construction remain. In their text, Magnusson and Endler (16) identified several item formats that increased self-report bias such as questions about general experiences. They proposed that questions that focus on real-life situations or actual experiences minimize self-report bias. Fleishman and Quaintance (17) later indicated that items should focus on the environment and conditions that underlie effective job performance. These suggestions can result in questions that produce a dimensional structure that is relatively stable over time (18).

In addition to economy, validity, and relative freedom from report biases, multiple-choice formats are readily quantified for statistical analysis (19). Biographical items generally present relatively low intercorrelations, which allow a few items to capture a large amount of descriptive information. The items should be scalable whenever possible so that they can be analyzed readily with sophisticated statistical procedures (13). The scalable format allows items to be scored from some lesser value to some greater value. This format for multiple-choice questions also lends itself to greater interpretability in biographical data analysis.

An important limitation in the use of predictive information derived from well-developed multiple-choice formats is the time delay between measurement and expected performance. Ghiselli (20) reviewed numerous investigations and found that as time delays increase, the validity of performance predictions decreases. Other research (21) suggests that delays less than 18 months produce the best statistical relationship with the expected criterion. Considering these limitations, the use of biographical data derived from multiple-choice measures would appear to remain a viable and cost-effective means of initial pilot candidate selection. Such measures may also make efficient use of the available pool of talent (22). Once a pool of questions has been developed and standardized, the actual scoring and statistical analysis can be completed through several procedures.

Mumford and Owens (13) identified four methods of item analysis used in biographical inventory data: 1) rational item scaling, 2) item subgrouping, 3) empirical item scaling, and 4) factorial item scaling. The rational scaling and subgrouping methods have limited statistical basis, but the empirical and factorial scaling procedures are statistically useful. Both the empirical criterion and factorial scaling methods produce similar solutions.

The empirical criterion keying method has been used extensively in the analysis of personality vocational assessment inventories such as the Minnesota Multiphasic Personality Inventory and the Strong Vocational Interest Blank (23). The empirical keying method involves item selection based on how well the
item actually discriminates a member of a selection group from a reference or criterion group. Items of the reference group that differ from the selection group are counted in the scoring key according to the relative size of the difference. Items with the largest difference would be assigned the largest weight in the key. Cronbach (23) indicated that the empirical criterion method makes no assumptions about underlying theory and can thus be used on a variety of inventories.

The factorial scaling approach employs factor analytic procedures to identify solutions that include the smallest number of items that account for the largest proportion of variance (13). The resultant factor groupings are then rotated with respect to simple structure for interpretation. A discriminant function analysis is used to derive relative item and/or item grouping weights for use in a prediction formula. The factorial scaling model should provide a solution that is more powerful than the empirical keying method, because small differences that would have been otherwise lost are retained in the final key (24).

Both the empirical criterion and factorial keying method provide similar solutions. The choice of which approach to employ rests mostly with the researcher as both the factorial scaling and empirical keying approach produce useful predictions. Factorial scaling is most appropriate when sample size is relatively large (25) and when the sample is also used as the selection pool and no separate criterion group is available. Empirical keys require a separate criterion group to avoid restricting range issues that may confound this design more than a factorial scaling approach, which can statistically control this problem (25). According to Vandeventer et. al. (21), factorial scales may not be as effective as empirical keys in initial selection, although they do display greater stability and generality.

Relatively few factorial analyses of biographical scales are noted in the research. This limitation may be related to the complexity of the operations and the availability of computer time and programs that can accommodate large numbers of cases and variables. When these resources are available, a factorial scaling approach to biographical data analysis would appear to be viable. Gorsuch (24) recommended that item analysis in any test construction procedure employ a factor analytic strategy. The preferred procedure for factor extraction is based on the kind of questions and their expected intercorrelations. Following factor extraction, an orthogonal rotation is the preferred method to arrive at simple structure and for ease of interpretation (26, 27).

We hypothesized that the factorial structure will produce interest and attitudinal patterns that are significantly different for individual subgroups. The specific groups to be studied will be the AOC/NAVCAD pass and fail groups. The factorial patterns of the two groups will be compared. A significant difference could then be used as the basis of a factorial scaling procedure in the development of an AOC/NAVCAD specific scoring key for aviation indoctrination.

METHOD

SAMPLE IDENTIFICATION

The subjects used in the present study were a subgroup of pilot candidates entering naval aviation training who had participated in the standardization of the ASTB BI between 1987 and 1990. We included only AOC and NAVCAD candidates in the subject data pool who had data relating to graduation or attrition in initial NASC training. The AOC program requires a degree from a civilian college and no previous military indoctrination. The typical NAVCAD has been selected from the enlisted Navy ranks and has had at least 2 years of college coursework. From this pool of 1837 cases, 286 had missing discriminating data and were excluded from analysis. The remaining pool consisted of 1551 complete cases. The sample selected was made up of AOCs/NAVCADs who graduated ($N = 1176$) and who attrited ($N = 375$) pre-flight training. Subjects ranged in age from 20 to 27 years and were tested on the ASTB BI at various locations...
around the country. Both men \((N = 1493)\) and women \((N = 58)\) were included in the AOC/NAVCAD selection sample, although sex differences were not analyzed.

**DATA ANALYSIS**

The Statistical Package for the Social Sciences (SPSS, Release 10.3) was the source of statistical analysis programs used in this study. The FREQUENCIES program was utilized to obtain means and standard deviations of the dependent variable with regard to pass/attrite status, sex, and group membership (i.e., AOC/NAVCAD). The FACTOR program for principal components was used in item analysis. The principal components model of exploratory factor analysis produces factors that account for all of the variance in each variable, including that variance shared with other variables in the set, and that variance specific to the variable itself. Of the components generated, 10 were retained on the basis of eigenvalues greater than 1. A VARIMAX rotation of the retained components was used to obtain a meaningful pattern of variables and simplified factor structure. These procedures follow recommendations of Gorsuch (1983) for item analysis using a factorial scaling procedure.

Once a simplified factor structure was obtained, items with at least a low correlation \((.30)\) on one of the principal components were retained. These items were summed to form a simple linear composite score for each of the 10 principal components. The resulting factor composite scores were then treated as single variables and subjected to a canonical discriminant function analysis (SPSS DISCRIMINANT). This program entered all variables simultaneously according to a preselected minimum tolerance level \((.01)\) such that a variable was removed from the analysis when it accounted for roughly the same variance accounted for by another variable. The SPSS program prior probabilities for pass or attrition were adjusted to obtain a 50% reduction in attrition during training. The differences in BI composite scores as constructed by a principal component solution for the students who passed and the students who attrited were analyzed using Student’s \(t\)-test.

**RESULTS**

A principal component analysis was performed to assess the pattern of item intercorrelations. Of the initial pool of 240 items on the ASTB BI, 176 items could be scored on an ordinal scale. The remaining items were purely demographic and did not lend themselves to principal component analysis. The principal component analysis and orthogonal rotation yielded a 10-component solution, which accounted for 25% of the total item variance. We assigned labels to the principal components obtained based on a review of the general content of the items that made up the component.

The first component (AOC1) was defined by items reflecting personal and emotional independence. It accounted for 5.7% of the total item variance and is made up of 18 items. The second component (AOC2) accounted for 2.8% of the total variance with 13 items reflecting an interest in math, science, and engineering. The third component (AOC3) accounted for 2.6% of the total variance and included 19 items related to risk taking. Sixteen items comprised the fourth component (AOC4), which reflected economic self-sufficiency and accounted for 2.5% of total variance. The AOC5 component included 11 military orientation items and accounted for 2.4% of total variance. The 10 items found in the sixth component (AOC6) accounted for 2.2% of total variance and described willingness to take the initiative. The seventh component (AOC7) included 12 items reflecting practical self-sufficiency such as taking care of a car. They accounted for 1.9% of total variance. The eighth component (AOC8) had 10 items and accounted for 1.8% of total variance; they reflected personal and emotional maturity. The ninth (AOC9) and tenth (AOC10) components account for 1.8% and 1.7% of the total variance, respectively. The 12 items in AOC9 reflected participation and interest in sports, while the 5 items in AOC10 represented school leadership experience.
Table 1 presents means and standard deviations for the composite AOC/NAVCAD pass and attrite component scores. As shown in Table 1, the pass and attrite group means for the AOC1 and AOC9 composites were significantly different at the $p < .001$ level. The AOC4 composite group means were different at the $p < .01$ level, while the AOC5 and AOC6 composite factor means were significantly different at the $p < .05$ level. A canonical discriminant function analysis was performed to assess the prediction of membership in the attrite or pass AOC/NAVCAD groups. There was statistically significant discrimination between the two groups on the basis of five of the 10 composite factors.

Table 1. AOC/NAVCAD Comparison of Composite Factor Means and Standard Deviations (SD) for Pass and Attrite

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pass</th>
<th>SD</th>
<th>Attrite</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOC1</td>
<td>44.26</td>
<td>5.29</td>
<td>42.77</td>
<td>5.47</td>
</tr>
<tr>
<td>AOC2</td>
<td>38.62</td>
<td>8.21</td>
<td>38.28</td>
<td>8.96</td>
</tr>
<tr>
<td>AOC3</td>
<td>46.68</td>
<td>8.14</td>
<td>46.93</td>
<td>8.13</td>
</tr>
<tr>
<td>AOC4</td>
<td>48.23</td>
<td>8.23</td>
<td>46.98</td>
<td>8.50</td>
</tr>
<tr>
<td>AOC5</td>
<td>37.72</td>
<td>6.09</td>
<td>36.93</td>
<td>6.47</td>
</tr>
<tr>
<td>AOC6</td>
<td>20.04</td>
<td>3.52</td>
<td>20.46</td>
<td>3.65</td>
</tr>
<tr>
<td>AOC7</td>
<td>35.87</td>
<td>4.69</td>
<td>35.47</td>
<td>4.52</td>
</tr>
<tr>
<td>AOC8</td>
<td>27.56</td>
<td>4.83</td>
<td>28.04</td>
<td>4.55</td>
</tr>
<tr>
<td>AOC9</td>
<td>25.02</td>
<td>4.29</td>
<td>24.09</td>
<td>4.61</td>
</tr>
<tr>
<td>AOC10</td>
<td>5.92</td>
<td>1.79</td>
<td>5.83</td>
<td>1.72</td>
</tr>
</tbody>
</table>

$^*$ $p < .05$

$^**$ $p < .01$

$^***$ $p < .001$

A significant discriminant function (DF) was calculated, with a combined $\chi^2 (10) = 61.78$, $p < .001$. This indicates that there was a statistical difference in the pass ($M_{DF} = 0.2$) and attrite ($M_{DF} = -0.3$) means for the distribution of DF scores calculated for the two groups. The Pearson correlation coefficient was $r = .19$. The amount of variance explained by the DF was $r^2 = .0361$ or 3.6%. Adjusting the pass/attrite probabilities to obtain a 50% reduction in attritions resulted in a 64.6% correct classification rate. The 50% reduction in attritions would result in a 30.6% false rejection cost. The accuracy of success predictions is approximately 81% for those selected in this model. A classification matrix based on the principal component solution is presented in Table 3.
Table 3. Principal Component Classification Matrix *

<table>
<thead>
<tr>
<th>Predicted Group Membership</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Group</td>
<td></td>
</tr>
<tr>
<td>Pass</td>
<td>816</td>
</tr>
<tr>
<td></td>
<td>69.4%</td>
</tr>
<tr>
<td>Attrite</td>
<td>139</td>
</tr>
<tr>
<td></td>
<td>50.4%</td>
</tr>
</tbody>
</table>

* Percent of grouped cases correctly classified: 64.6%

We included all 176 scalable BI items in the final DF analysis but eliminated 66 items from the analysis when they failed to meet the minimum tolerance for redundancy of .001. The results of the analysis revealed a statistically significant difference in pass (M DF = 0.1) and attrite (M DF = -0.7) means for the distribution of DF scores calculated for the two groups with the retained 110 items ($\chi^2 (110) = 214.22, p < .001$). The Pearson correlation coefficient was .36. The non principal component model explained 12.96% of the total variance and correctly classified 73.2% of the grouped cases when the criterion was adjusted to obtain a 50% reduction in attritions. The related cost in terms of false rejections is 19.1%. The new hit rate for those selected in the non principal component model is approximately 83%. The individual item classification matrix is presented in Table 4.

Table 4. Individual Item Classification Matrix *

<table>
<thead>
<tr>
<th>Predicted Group Membership</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Group</td>
<td></td>
</tr>
<tr>
<td>Pass</td>
<td>951</td>
</tr>
<tr>
<td></td>
<td>80.9%</td>
</tr>
<tr>
<td>Attrite</td>
<td>190</td>
</tr>
<tr>
<td></td>
<td>50.7%</td>
</tr>
</tbody>
</table>

* Percent of grouped cases correctly classified: 73.2%

DISCUSSION

The results of this study demonstrate statistical differences in attitudinal and interest patterns in samples of AOC/NAVCAD officer candidates who attrite and who graduate initial NASC training. A scoring key derived with 110 ASTB BI items was more accurate than the factorial item groupings in predicting pre-flight attrition. The current results are limited to linear approaches. Nonlinear transformations may produce solutions that are different from those obtained in this study.
The ASTB BI scoring keys derived through linear transformations were statistically different for AOC/NAVCAD candidates. The presence of statistical differences in the scores from the two groups must be viewed in terms of the overall value of the differences in actual group discrimination. The results indicate that a scoring key derived from discriminant function techniques can decrease attrition rates for AOC/NAVCAD candidates in preflight training.

Our findings have several implications. First, this study presented an uncommon factor analytic technique for biographical item analysis in military selection research. Earlier studies have focused on test item development through expert judgement and empirical item keying and did not assess item patterns in post hoc analysis (12). Therefore, actual comparisons with other studies were limited. In this study, factor analytic procedures were used to gain insight into the attitudes and interests of two groups with a prepackaged pool of items. The resulting analysis yielded a 10-component solution similar to the underlying psychological structure suggested in the construction of the ASTB/BI.

In addition, previous research suggested that factor analytic solutions (28) would provide a more powerful DF model than methods such as empirical keying. This procedure had not been previously applied to the AQT/FAR or the ASTB/BI. Our principal component model failed to produce a solution that could improve group discrimination beyond current pass/attrition rates. However, the application of linear DF techniques to individual scalable items indicated that ASTB BI item response differences may be sufficiently different early in the selection process to improve predictions for preflight training. Our non principal component DF analysis indicated that a 50% reduction in attritions could be achieved based on the adjustment of the DF criterion. The price for an increase in attrition prediction accuracy is a dramatic increase in the number of predicted attritions who would have otherwise passed training (i.e., a 19.1% false rejection rate using an all item DF solution). The benefit of a non principal component item solution to the ASTB BI is an 83% accuracy rate in pass predictions. This accuracy rate exceeds the current 75% hit rate in AOC/NAVCAD accessions. The impact of such a decision would require an increase in the number of accessions to keep pace with Navy demands for successful pilot graduates. The actual cost savings, if any, need to be calculated before any decision to implement this system.

Attrition prediction has historically been the primary emphasis of the BI. The current prediction model employs the FAR BI in prediction of attrition for all naval aviation candidates. Additional research should be conducted to measure the amount of unique variance explained by the ASTB BI in the naval pilot attrition prediction model. We recommend that hierarchical multiple regression procedures be used to address this concern. Hierarchical procedures allow for the statistical control of the effects of certain variables and are less subjective to error than iterative procedures such as stepwise multiple regression. The statistical control of the effects of intelligence and certain demographic variables helps to equalize the measures and more accurately judge their unique contribution. We believe that a solution obtained through hierarchical multiple regression would produce the most useful estimate of the unique contribution of the ASTB BI to attrition prediction.

Finally, further research is needed into the use of the ASTB BI in predicting attrition in other naval pilot training subgroups. Particular interest lies in the use of the BI with groups with substantially different background and interest patterns. Such groups might be women and minorities who may have experiences and interests different from the typical white male candidate. Linear discriminant function techniques may prove useful as the size of these groups increases.

RECOMMENDATIONS

We believe that a scoring system for the ASTB BI derived by principal components technique could be useful in targeting AOC/NAVCAD candidates likely to attrite naval aviation training. A scoring key based on all scalable ASTB BI items should be considered for implementation as a method of attrition prediction in preflight AOC/NAVCAD training. The results are limited to linear composites formed by
principal components. Nonlinear transformations may reveal important associations between ASTB BI and attrition.
REFERENCES


