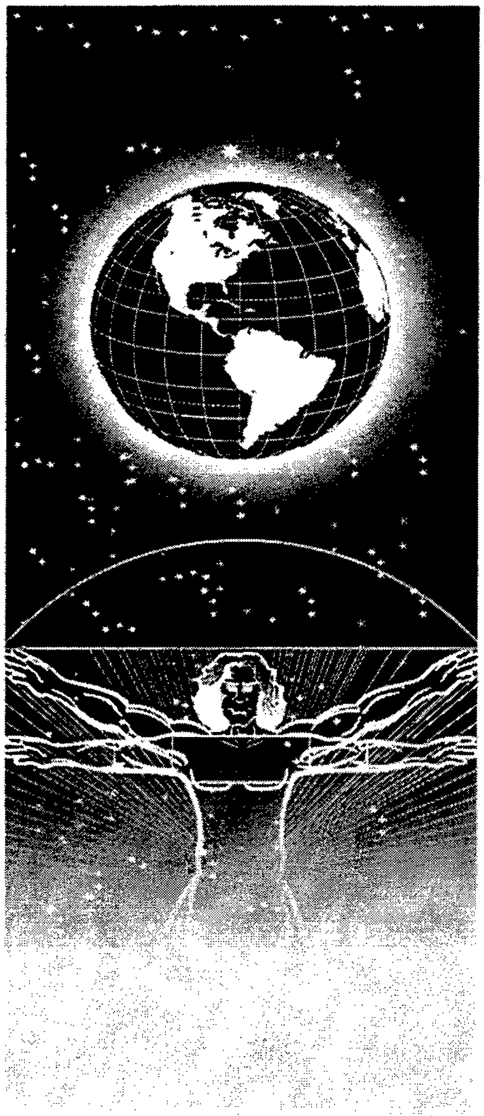
AFRL-HE-WP-TR-2001-0074



UNITED STATES AIR FORCE RESEARCH LABORATORY

PREDICTING TRAINING SUCCESS WITH THE NEO-PI-R: THE USE OF LOGISTIC REGRESSION TO DETERMINE THE ODDS OF COMPLETING A PILOT SCREENING PROGRAM

MARTIN N. ANESGART
JOSEPH D. CALLISTER

HUMAN EFFECTIVENESS DIRECTORATE
CREW SYSTEM INTERFACE DIVISION
WRIGHT-PATTERSON AFB OH 45433-7022

FEBRUARY 2001

INTERIM REPORT FOR THE PERIOD DECEMBER 1998 TO MAY 1999

Human Effectiveness Directorate
Crew System Interface Division
2255 H Street
Wright-Patterson AFB OH 45433-7022

NOTICES

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-2001-0074

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

MARIS M. VIKMANIS
Chief, Crew System Interface Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
February 2001

2. REPORT TYPE
Interim

3. DATES COVERED *(From - To)*
December 1998 - May 1999

4. TITLE AND SUBTITLE
PREDICTING TRAINING SUCCESS WITH THE NEO-PI-R: The Use of Logistic Regression to Determine the Odds of Completing a Pilot Screening Program

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

5d. PROJECT NUMBER
7184

5e. TASK NUMBER
00

5f. WORK UNIT NUMBER
00

6. AUTHOR(S)
MARTIN N. ANESGART, PH.D.
MAJ JOSEPH D. CALLISTER, PH.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory, Human Effectiveness Directorate
Crew System Interface Division
Air Force Materiel Command
Wright-Patterson AFB OH 45433

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/HEC

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-HE-WP-TR-2001-0074

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The revised NEO personality inventory (NEO-PI-R) measuring the "big five" personality domains has been used as a descriptive tool for assessing "normal" personality attributes of Air Force pilots; yet the predictive potential of NEO-PI-R has not been realized. An opportunity to transform the NEO-PI-R into a predictive instrument has been provided by the recent collection of data from the U.S. Air Force's Enhanced Flight Screening Program. This paper describes a technique used to provide a practical and intuitive method of predicting whether a student pilot with a particular profile on the NEO-PI-R will be successful in training.

15. SUBJECT TERMS
PILOT TRAINING, SELECTION, PERSONALITY INVENTORIES, LOGISTIC REGRESSION, TABLE OF ODDS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	12	MARTIN N. ANESGART
					19b. TELEPHONE NUMBER *(Include area code)* 937-656-7214

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF TABLES

1. INTRODUCTION

Upon entering the Enhanced Flight Screening Programs at Hondo, TX and the US Air Force Academy in Colorado Springs, CO, all student pilots take a battery of psychological tests that include personality inventories (Callister & Retzlaff, 1996). Although historically the link between psychological tests and performance has been suspect (Bass & Barrett, 1981; McCormick & Iigen, 1985), two relatively recent meta-analytic reviews have established valid connections between personality measures, particularly those measures based on the "Big Five" personality dimensions, and performance (Barrick & Mount, 1991; Tett, Jackson & Rothstein, 1991). The "Big Five" factors—Neuroticism or Emotional Stability, Extraversion, Openness to Experience, Agreeableness and Conscientiousness—may provide good utility in the employee and trainee selection process. In fact, a prominent and oft-used inventory, The NEO-PI, based on the "Big Five" has been found to be significantly correlated ($p<.05$) with supervisory ratings of job performance and can increase the predictive accuracy of employee success by "6 to 24% over that expected by chance" (Piedmont & Weinstein, 1993, footnote). A revised edition of the NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) plays a critical role in the current presentation. The NEO-PI-R contains the "Big Five" factors called domains with six facets per domain (see Table 1).

Table 1. Domains and Facets

Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness
Anxiety	Warmth	Fantasy	Trust	Competence
Angry hostility	Gregariousness	Aesthetics	Straightforwardness	Order
Depression	Assertiveness	Feelings	Altruism	Dutifulness
Self-consciousness	Activity	Actions	Compliance	Achievement Striving
Impulsiveness	Excitement-seeking	Ideas	Modesty	Self-discipline
Vulnerability	Positive emotions	Values	Tender-mindedness	Deliberation

As a descriptive tool the NEO-PI-R is "highly regarded for its ability to gauge normal personality functioning" (King & Flynn, 1995, pg. 955). As in the King and Flynn study and in a paper by Callister, King, Retzlaff, and Marsh (1999) student pilot assessments are typically made to compare student pilots to experienced pilots or to the general population. While useful descriptions and inferences to the greater population of pilots have been made, predictions of the likelihood of an individual completing the screening program have not been made. Moreover, there are times when a simple and practical method is needed that would offer a means of predicting the odds that a particular individual will complete the training program. This paper describes the construction of "odds tables" of success or failure derived from the use of logistic regression. These tables can provide an efficient method for determining the probability of success of an individual with a specific set of scores.

2. METHOD

2.1 Sample

Data collection was based on NEO-PI-R domain scores from 1031 individuals accepted into the U. S. Air Force Enhanced Flight Screening program. These individuals were all graduates of Air Force Reserve Officer Training and had varying levels of previous flight hours (in civilian aircraft); including 121 people with zero flight hours and one person with 6700 hours; the mean was 255 and the median was 70. Approximately 7.5% (N=77) of the sample were women; 88% (N=907) completed the screening process; 5% (N=52) left due to flying training deficiency (FTD); 3% (N=31) left due to self-initiated termination (SIE); 4% (N=41) left due to miscellaneous reasons.

2.2 Design

Cluster analysis did not reveal any clear or distinct combinations of domains that correlated even moderately with the completion statistics. Logistic regression techniques did demonstrate a relationship between particular profiles on the NEO-PI-R and program completion versus failure to complete due to self-initiated elimination, a flying training deficiency, or other miscellaneous reasons.

Logistic regression differs from the more typically used linear regression in that in logistic regression the dependant variable is discrete. Normally the dependent or outcome variable is dichotomous, but under special circumstances multiple categories can be used. The independent variables or predictors can be either discrete or continuous and are related to the dependent variable in an exponential function. Turning the continuous scores of the predictor NEO-PI-R domains into categories or intervals as described below enabled the computation of "odds" that an individual with a particular domain profile will complete screening or wash out as compared to a benchmark profile.

Construction of the tables of odds proceeded in the following manner through the use of the *Statistical Package for the Social Sciences* (SPSS, release 6.0, 1993): (1) for each domain, a pre-trainee's facet scores were combined into a summed domain score with a frequency distribution derived; (2) sigmas or the normal standard deviation cutoffs (+ or - 0.5 sigma, + or - 1.5 sigma), equivalent to the T scores expressed in the NEO-PI-R professional manual, were imposed on the distribution of summed domain scores, resulting in the five categories (Table 2), very low (coded 0), low (coded 1), average (coded 2), high (coded 3) and very high (coded 4) with the category labels being the same as shown on the NEO-PI-R rating forms; (3) the dependent variable was coded 1 for completion and 0 for termination under all circumstances (or 0 for each termination reason examined separately); (4) logistic regression models were used to compare domain categories simultaneously; and (5) matrix tables were constructed with all possible comparisons or odds within one domain or combination of domains.

Table 2. Category Cutoffs for Each Domain

Category	Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness
Very low (0)	<40	<102	<87	<89	<108
Low (1)	40 to 58	102 to 118	87 to 104	89 to 105	108 to 124
Average (2)	59 to 77	119 to 134	105 to 122	106 to 122	125 to 141
High (3)	78 to 97	135 to 151	123 to 140	123 to 139	142 to 157
Very high (4)	>97	>151	>140	>139	>157

3. RESULTS

Of the reasons for termination or leaving the screening program, SIE is most related (in terms of generating statistically significant odds) with the individual domains; neither FTD nor miscellaneous reasons had significant odds associated with them. Table 3 presents a cross tabulation of frequencies for each level of the Neuroticism domain in terms of completing or leaving the program due to SIE. Table 4 displays the logistic regression comparisons between the category levels of Neuroticism for the dependent variable of completion vis-a-vis SIE. Odds can be calculated from either table. For example, from the contingency table, dividing the proportion of SIE to completion in the low condition (4/217) by the proportion in the very low condition (1/63) yields 1.1613, the same value in the logistic regression table. The benefit of the latter is providing the statistics from which confidence intervals can be derived. Note that in Table 4 the Wald statistic is like the *t*-statistic in that the regression coefficient (*B*) is divided by the standard error (*S.E*). Unlike the *t*-statistic the result is squared. The last column in the table is the result of taking *B* as the exponent to the base *e*, the anti-natural log. The results in this column are the odds that an individual who scores in one category (e.g. low Neuroticism) will leave the program compared to an individual who scores in another category (e.g. very low Neuroticism). Individuals with high and very high Neuroticism have greater odds of leaving the screening program than individuals who score in the other levels. For instance, an individual with high Neuroticism is over 6 times as likely to leave the program than an individual with low Neuroticism. Note that taking the reciprocal of the odds gives the relative probability of completing the program. For example, the reciprocal of 6.2028^{-1} is .1612, so the high Neuroticism individual is 16.12% as likely to complete the screening program as the low Neuroticism individual.

Table 3. Frequencies for Each Level of Neuroticism by Completion

	Very Low	Low	Average	High	Very High
SIE	1	4	5	16	5
Completion	63	217	378	195	54

Table 4. Logistic Regressions for Each Neuroticism Comparison

Comparisons	B	S.E.	Wald	Sig.	R	Exp(B)
Low vs. Very Low	.1495	1.1272	.0176	.8945	.0000	1.1613
Average vs. Very Low	-.1823	1.1039	.0273	.8688	.0000	.8334
High vs. Very Low	-.3318	.6762	.2408	.6236	.0000	.7176
Very High vs Very Low	1.6427	1.0409	2.4906	.1145	.0424	5.1692
Average vs. Low	1.4932	.5677	6.9192	.0085	.1344	4.4513**
High vs. Low	1.8250	.5199	12.3236	.0004	.1947	6.2028**
Very High vs. Low	1.7636	1.1110	2.5197	.1124	.0437	5.8333
High vs. Average	1.6141	.6878	5.5063	.0189	.1135	5.0231*
Very High vs. Average	1.9459	.6490	8.9905	.0027	.1602	6.9997**
Very High vs. High	.1209	.5349	.0511	.8212	.0000	1.1285

* $p<.05$
**$p<.01$

Confidence intervals can be drawn around the odds. A simple simultaneous procedure taking into account the number of comparisons similar to Dunn's or the Bonferroni *t* is followed that controls for the level of alpha or Type-I error (Netter, Wasserman & Whitmore, 1988). Note that in the tables that follow a particular odds value may be significant at the $p<.05$ or $p<.01$ level (one-tailed level), but the confidence intervals may be 90% or 95%, respectively. This is due to the actual significance level not meeting the two-tailed critical alpha levels, .025 for .05 alpha and .005 for .01 alpha.

In the comparison mentioned above we can be 99% certain that the odds of a high Neuroticism leaving vis-à-vis a low Neuroticism dips to a little above even (1.1208) and rises to over 34 times as likely. Table 5 displays the odds with the confidence intervals in parentheses.

Table 5. Table Of Odds for One Domain (Neuroticism)

Category	Very low (VL)	Low (L)	Average (A)	High (H)	Very high (VH)
Very low	X	1.1613	.8334	.7176	5.1692
Low	.8611	X	4.4513**[1] (.9045,21.9054)	6.2028**[2] (1.1208,34.329)	5.8333
Average	1.1999	.2247**[1] (.0457,1.1056)	X	5.0231*[1] (.7287,34.6312)	6.9997**[2] (.827,59.2496)
High	1.3935	.1612**[2] (.0291,.8922)	.1991*[1] (.0289,1.3723)	X	1.1285
Very high	.1935	.1714	.1429**[2] (.0169,1.2092)	.8861	X

*p<.05
**p<.01
[1]the numbers in parentheses are lower and upper limits of the 95% confidence interval.
[2]the numbers in parentheses are lower and upper limits of the 99% confidence interval.

Admittedly the ranges are large and will be even greater when considering domains in combination, although being able, perhaps, to collect more data and increase the sample sizes will likely lower the standard error and decrease the size of the intervals. For a two-way analysis with two domains combined the number of comparisons increases to 300---$5^2 (5^2 - 1)/2$; 7750 comparisons for a three-way analysis; 195,000 for a four-way and 4,881,250 for a five-way. These latter two analyses will require prohibitively large N-sizes. Thus with the current sample of 1031 only tables for one-, two- and three-way comparisons were constructed. These tables are intended for a separate Air Force Research Laboratory technical report.

As domains are combined and the number of comparisons increases, the user can flag those profiles that are associated with the greatest odds of leaving or washing out of the program relative to other profiles. As an example, Tables 6 depicts what might be called a "vulnerable profile", high Neuroticism and low Extraversion, that has associated with it a repeated tendency and greater likelihood to leave the program due to SIE (as compared to several other profiles). Or the user could designate a benchmark profile, average Neuroticism and average Extraversion, for instance, to which other profiles are compared. Four such profiles are displayed in Table 7. These profiles are associated with a greater likelihood to leave the program compared to the average profile.

Table 6. Table of Odds for Two Domains Combined:
Neuroticism (N) and Extraversion (E)

Vulnerable Profile	Referent Profile	Odds
High N + Low E	Low N + Average E	6.50* (.3932,107.4526)[1]
High N + Low E	Average N + Average E	9.3333** (.8494,102.5551)[1]
High N + Low E	Average N + High E	7.00* (.4242,115.5150)[1]

* p<.05
**p<.01
[1]the numbers in parentheses are lower and upper limits of the 95% confidence interval

Table 7. Table of Odds Compared to Benchmark Domain Profile: Average Neuroticism and Average Extraversion

Domain Profile	Odds
High Neuroticism/Low Extraversion	9.3333** (.8494,102.5551)[1]
High Neuroticism/Very High Extraversion	14.00* (.1719,1140.1818)[0]
Very High Neuroticism/Very Low Extraversion	7.00* (.2546,192.4823)[0]
Very High Neuroticism/Low Extraversion	8.00* (.2862,223.6108)[0]

* p<.05
**p<.01
[0]the numbers in parentheses are lower and upper limits of the 90% confidence interval.
[1]the numbers in parentheses are lower and upper limits of the 95% confidence interval.

Note that high Neuroticism combined with very high Extraversion shows greater odds than expected (when considered with the other three profiles). This may be due to a limitation of the overall procedure's dependency on sample sizes as the number of comparisons increases. Although the number of unsuccessful individuals that exhibited both high Neuroticism and very high Extraversion was decidedly less than the numbers for the other three categories, the number of individuals completing the program was also very much less, resulting in a higher washout likelihood ratio for the former category, 25% versus 16.67% for high Neuroticism/low Extraversion; versus 12.5% for very high Neuroticism/very low Extraversion; and versus 14.29% for very high Neuroticism/low Extraversion.

Tables 8 and 9 represent the result of adding Openness to the other two domains. Table 8 displays the vulnerable profile---very high Neuroticism combined with very low Extraversion and low Openness---which has a much greater likelihood of leaving the program compared to the referent profiles. Table 9 uses an average benchmark again displaying the greater odds associated with the other profiles.

Table 8. Table of Odds for Three Domains Combined:
Neuroticism (N), Extraversion (E) and Openness (O)

Vulnerable Profile[#]	Referent Profile	Odds
VH N + VL E + L O	L N + A E + A O	26.6667* (.2259,3148.2436)[1]
VH N + VL E + L O	A N + A E + A O	50.6667** (.4357,5892.5846)[1]
VH N + VL E + L O	A N + A E + H O	11.3333* (.1927,666.4239)[0]
VH N + VL E + L O	H N + L E + A O	16.6667* (.1404,1978.6360)[0]

[#] VL= very low; L= low; A= average; H= high; VH= very high.
* $p<.05$
**$p<.01$
[0]the numbers in parentheses are lower and upper limits of the 90% confidence interval.
[1]the numbers in parentheses are lower and upper limits of the 95% confidence interval.

Table 9. Table of Odds Compared to Benchmark Domain Profile

Domain Profile	Odds
High N + Very Low E + Low O	19.00* (.2154,1675.8068)[1]
High N + Very Low E + High O	38.00* (.1479,9760.9987)[1]
High N + Very Low E + Very High O	25.3333** (.5178,1239.3595)[1]
High N + Very High E + Average O	76.00* (.1747,33057.3761)[1]
Very High N + Very Low E + Low O	50.6667** (.4357,5892.5846)[1]
Very High N + Low E + Very High O	76.00* (.1747,33057.3761)[1]

* $p<.05$
**$p<.01$
[1]the numbers in parentheses are lower and upper limits of the 95% confidence interval.

4. DISCUSSION

The purpose of this paper is to introduce a decision making tool that can be used in the selection of future pilots. The key component of this tool is a set of odds tables developed from the logistics regression analysis of large sets of data collected from individuals in a flight screening program. These simple to use tables would allow a clinician, training or selection officer to compare the test results of a specific individual to a benchmark profile, such as average scores on all the domains or one associated, perhaps, with a highly successful performer. So, for example, the decision maker would know that candidate A, who scored high on Neuroticism and low on Extraversion, is almost 10 times more likely to self-eliminate than is the average candidate. Or, the decision maker would know that candidate B, who scored very high on Neuroticism, very low on Extraversion, and low on Openness is 50 times more likely to self-eliminate than is the average candidate.

The data set used in this study suggests that these particular odds tables are most predictive when used with inexperienced student pilots, in the present case with 22 previous flying hours or less (bottom third of flying hour distribution). With 110 hours or greater (top third) just 2 candidates of 336 fail to complete the program---both exhibit low scores on Extraversion but are disparate on the other domains. This compares to 73 out of 324 individuals with 22 hours or less, and 22 out of 322 individuals with greater than 22 but less than 110 hours. (Note: Previous flying experience data was missing for 49 students). In terms of odds, experienced flyers are 2% as likely to wash out than those with little or no experience. With sufficient N-sizes, other demographics or characteristics of the sampled populations can be investigated as moderators of the dependent/predictor(s) relationship.

A limitation of this approach is that as the number of comparisons increases the sample sizes required would be in the tens or even hundreds of thousands. Without these large samples, many cells would be empty and overall effort for return would be poor. However, when these large data sets do exist, as they do in the U. S. Air Force, a reasonable number of profiles can be compared.

5. CONCLUSION

In sum, predictive odds tables, easy to use and understand, as derived from logistic regression and the domains of the NEO-PI-R may have great utility (savings in costs and man-hours) as part of a battery of tools for screening potential pilots. Decision makers could consult the odds tables to identify profiles that are most associated with failing or completing the program. Important relationships between characteristics such as emotional stability and self-elimination from training will be reflected in these tables, enabling decision makers to invest in those individuals with the greatest potential for success. Finally, additional research to validate the tables of odds will likely require a cross-validation study as well as longitudinal studies. It will be essential to follow a wide range of individuals through flight training and into the operational environment to determine the profiles of the most and least successful pilots. Once validated, the methodology could be applied in other professional contexts with different types of selection or vocational instruments.

6. REFERENCES

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.

Bass, B. M., & Barrett, G. V. (1981). *People, Work, and Organizations (2nd Ed)*. Boston, MA: Allyn & Bacon.

Callister, J. D., King, R. E., Retzlaff, P. D. & Marsh, R. W. (1999). Revised NEO Personality Inventory profiles of male and female U.S. Air Force Pilots. *Military Medicine,* 164, 885-890.

Callister, J. D., & Retzlaff, P. D. (1996). The USAF's Enhanced Flight Screening Program: Psychological assessment of undergraduate pilot training candidates. *AGARD Conference Proceedings 588: Selection and Training Advances in Aviation* (pp. 14-1 to 14-5). Neuilly-Sur-Siene, France: NATO Advisory Group for Aerospace Research and Development.

Costa, P. T., & McCrae, R. R. (1992). *Professional manual Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.

King, R. E., & Flynn, C. F. (1995). Defining and measuring the "Right Stuff": Neuropsychiatrically Enhanced Flight Screening (N-EFS). *Aviation, Space, and Environmental Medicine*, 66, 951-956.

McCormick, E. J., & Iigen, D. (1985). *Industrial and Organizational Psychology (8th Ed)*. Englewood Cliffs, NJ: Prentice-Hall.

Netter, J., Wasserman, W., & Whitmore, G. A. (1988). *Applied Statistics (3rd Ed.)*. Boston, MA: Allyn & Bacon, Inc.

Piedmont, R. L., & Weinstein, H. P. (1993). Predicting supervisor ratings of job performance using the NEO personality inventory. *The Journal of Psychology*, 128, 255-265.

SPSS for Windows (1993). Chicago, IL: SPSS.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). *Personnel Psychology*, 44, 703-742.

Author Note

Martin N. Anesgart, Crew System Interface Division, Air Force Research Laboratory; Joseph D. Callister, Aerospace Engineering, Applied Mechanics, and Aviation, The Ohio State University (now at Crew System Interface Division, Air Force Research Laboratory). The authors would like to acknowledge the technical assistance provided by Pauline Schlichting and William Weaver. Correspondence concerning this article should be addressed to Martin N. Anesgart, AFRL/HECI, 2210 8th Street, Bldg. 146, Rm. 122, Wright-Patterson AFB, OH 45433.