Journal of Applied Psychology
1994, Vol. 79, No. 6, 845–851

Predicting Pilot and Navigator Criteria: Not Much More Than *g*

Michele Morales Olea and Malcolm James Ree

A comparison of the validity of psychometric *g* and specific ability or job knowledge, *s*, for predicting pilot and navigator criteria was conducted. Psychometric *g* and *s* were estimated from the principal components of a multiple aptitude test battery. The criteria included passing–failing training, an overall performance composite, academic performance, and work samples of pilot and navigator tasks. Regression analyses conducted to evaluate the predictive efficiency of *g* and *s* demonstrated that *g* was the best predictor of all criteria and *s* contributed little beyond g.

Charles Spearman (1904) proposed a two-factor ability model including general cognitive ability, *g*, and $s_1, s_2, s_3, \ldots s_n$, representative of test unique specific factors. Psychometric *g* typically accounts for most of the variance in a battery of cognitive tests and usually exceeds the variance accounted for by all the specific factors combined (Jensen, 1980). Tests that measure specific information, s_1 through s_n, are measures of job knowledge (Hunter, 1983). The current study investigated the predictive utility of *g* and $s_1 \ldots s_n$ in an occupational setting.

Many psychologists abandoned *g* theory and embraced multiple aptitude theory, often called the theory of differential ability or the specificity doctrine (Jensen, 1984). For example, Hull (1928) developed the concept of substitutability of specific skills for general ability, but did not provide empirical evidence. However, empirical evidence for the predictive efficacy of *g* continued to accumulate. For example, McNemar (1964) reviewed 4,096 validity coefficients and reported that multiple aptitude batteries achieved little differential validity (Brogden, 1951) compared with tests of general ability.

Recent studies have again shown the value of *g* as a predictor of occupational criteria (Carey, 1992; Hunter & Hunter, 1984; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Ree & Earles, 1991b, 1992; Ree, Earles, & Teachout, 1994; Thorndike, 1986). For example, Ree and Earles (1991b), in a sample of 78,041 airmen in 82 job specialties, regressed technical school grades on *g* and $s_1 \ldots s_n$ estimated from a multiple aptitude battery. Some portions of $s_1 \ldots s_n$ reflect job knowledge about mechanical principles, electronics information, and automotive information and are not measures of ability. For all 82 jobs, *g* was the most valid predictor with specific abilities or job knowledge measures yielding an average increase in predictiveness of about .02.

McHenry et al. (1990) found that *g* was the best predictor of Army Project A job performance measures and that specific ability or job knowledge measures incremented validity by .02 or less. Ree et al. (1994) investigating job performance, found similar results; *g* was the best predictor (average $r = .42$) and specific job knowledge or ability measures incremented predictive validity .02. Carey (1992) also conducted a study using job performance criteria and found increments above *g* of about .02.

Selection is becoming increasingly important with fewer training resources and expected increases in job complexity. The current study investigated the contribution of *g* and *s* to the prediction of pilot and navigator criteria.

Method

Subjects

The subjects were approximately 1,400 Undergraduate Navigator Training (UNT) students and 4,000 Undergraduate Pilot Training (UPT) students who tested on Form O of the Air Force Officer Qualifying Test (AFOQT; Skinner & Ree, 1987) between 1981 and 1985.

At the time of testing, a majority of the subjects possessed a high school education and more than 50 % had obtained some college education. All were tested at least 13 months prior to criterion data collection and some were tested as much as 4 years prior. All had baccalaureate degrees when they began training.

Predictors

As shown in Table 1, the AFOQT is composed of 16 tests, three of which are classified as power tests: Mechanical Comprehension, Rotated Blocks, and General Science. Electrical Maze, Instrument Comprehension, and Block Counting are primarily speeded, and the remaining tests are of a mixed power and speed model (Skinner & Ree, 1987). The tests are assembled into five composites used for officer selection and classification of pilots and navigators: Verbal (V), Quantitative (Q), Academic Aptitude (AA), Pilot (P), and Navigator–Technical (N-T). These composites are a reification of differential aptitude theory. Even though each composite measures some specific ability or job knowledge (viz. Aviation Information and Instrument Comprehension), they are all highly *g* saturated (Earles & Ree, 1991).

Ree and Earles (1991a) showed that *g* estimates based on the three commonly used methods, unrotated principal components, unrotated principal factors, and hierarchical factor analyses, were almost identical for any set of variables as long as sufficient positive manifold existed. As a consequence, the mathematically simple principal component estimates were used.

Principal components analysis (Hotelling, 1933) yields as many or-

Michele Morales Olea and Malcolm James Ree, Human Resources Directorate, Armstrong Laboratory, Brooks Air Force Base, Texas.

The views expressed are those of the authors and not necessarily those of the United States Air Force, Department of Defense, or United States Government.

Correspondence concerning this article should be addressed to Malcolm James Ree at AL/HRMAT, 7909 Lindbergh Drive, Brooks Air Force Base, Texas 78235-5352.

Table 1
Air Force Officer Qualifying Test, Form O Tests and Composites

Subtest	Reliability	Item	Time	P	N-T	AA	V	Q
				Composite				
Verbal Analogies (VA)	.76	25	8	X		X	X	
Arithmetic Reasoning (AR)	.71	25	29		X	X		X
Reading Comp (RC)	.75	25	18			X	X	
Data Interpretation (DI)	.42	25	24		X	X		X
Word Knowledge (WK)	.73	25	5			X	X	
Math Knowledge (MK)	.78	25	22		X	X		X
Mechanical Comp (MC)	.73	20	22	X	X			
Electrical Maze (EM)	.57	20	10	X	X			
Scale Reading (SR)	.65	40	15	X	X			
Instrument Comp (IC)	.55	20	6	X				
Block Counting (BC)	.51	20	3	X	X			
Table Reading (TR)	.44	40	7	X	X			
Aviation Information (AI)	.66	20	8	X				
Rotated Blocks (RB)	.65	15	13		X			
General Science (GS)	.73	20	10		X			
Hidden Figures (HF)	.45	15	8		X			
Total		350	208					

Note. Reliability based on N = 409 test–retest sample reported by Carretta and Ree (in press). P = Pilot composite, N-T = Navigator-technical composite, AA = Academic Aptitude composite, V = Verbal composite, Q = Quantitative composite, Comp = comprehension.

thogonal components as the number of tests. The first principal component from an aptitude battery typically estimates g with the remaining components representing job knowledge or specific ability measures (s_1. . .s_n). Rotation was not performed, because it redistributes first factor variance among the remaining factors and would render the first factor an inadequate measure of g.

Principal components account for all the variance in the matrix and are mathematical entities and not necessarily interpretable psychologically. Use of all 16 principal components ensures that no potentially valid sources of variance from job knowledge or specific abilities were omitted. The predictors were the scores from the 16 unrotated principal components of the AFOQT, all of which were from first-time administration to avoid practice effects.

Criteria

There were six criteria each for pilots and navigators. A dichotomous pass/fail score, four grades or work sample scores, and a unit-weighted summed composite of the four that was developed as a measure of overall performance were the criteria. A pass was reported if the grade average exceeded 70. Eighty-four percent of the UNT subjects and 79 % of the UPT subjects passed training. The criteria were established on the basis of task-based job analyses and are viewed by the trainers as minimum competency for flying aircraft. As opposed to more traditional written job knowledge tests, work samples measure job-task performance and are not usually considered training measures, even though collected during training.

The four ratings-based UNT criteria included Airmanship Grade, Basic Procedures Grade, Day Celestial Check Flight Rating, and Night Celestial Check Flight Rating. Airmanship included instruction on flight instruments and map reading. Basic Procedures included flight safety, airspace, and earth physics training. Day and Night Celestial Check Ratings were work samples of stellar observations, sun plotting, and actual flight missions. The overall performance composite was formed by unit weighting and summing the four rating-based criteria.

UPT criteria included pass–fail, Phase 2 and Phase 3 Check Flight averages, Air Education and Training Command (AETC) Phase 2 and Phase 3 Averages. Check Flight averages (work samples) were ratings of flight missions flown in jet aircraft and are measures of ability to perform the job. Phase 2 involved initial jet training and Phase 3 consisted of advanced instruction in a supersonic aircraft. Phase averages were cumulative grades covering flying performance, commanders' ratings, and written tests on various subjects such as mission planning and other aspects of airmanship. An overall performance criterion was created for UPT by summing the four unit-weighted rating-based criteria.

All work sample ratings were made by instructor pilots or instructor navigators as part of their routine duties. No reliability estimates were available for the criteria. The differing sample size for the criteria reflects the data as maintained in training records.

Procedure

A predictive study was conducted. Twelve stepwise multiple regressions were computed using the principal components as predictors with a Type I error rate of $p < .01$ for each step. Stepwise regression was chosen to select the most parsimonious set of predictors. Each pilot and navigator criterion was regressed on the 16 principal components. Analogous regressions were performed after the data were corrected for range restriction (Lawley, 1943). The variables included in these regressions were only those which were found to be significant in the regressions computed in the raw data. This was necessary because no statistical test exists for correlations corrected for range restriction.

Because range restriction corrections increase sampling error for corrected correlations, effective sample size estimates were used in the cross-validation procedures. Using the original sample size in the estimates of cross-validated correlations would bias the estimates upward. Schmidt, Hunter, and Larson (1988) noted that the increase in standard error of corrected correlations was equivalent to (effectively) using a smaller sample and solved the usual equation of the standard error of r for this effective sample size. The correction has the consequence of reducing the effective sample size proportional to the increase in r after correction for range restriction. Multiple correlations and effective sam-

Table 2
Descriptive Statistics for Undergraduate Navigator Training (UNT) and Undergraduate Pilot Training (UPT) Criteria

Criteria	*N*	Minimum	Maximum	*M*	*SD*
UNT					
Pass/Fail	1,411	0.00	1.00	0.84	0.36
Airmanship	1,341	60.00	100.00	93.59	6.00
Basic Procedures	1,176	50.90	100.00	93.23	6.54
Day Check Flight	1,224	0.00	100.00	87.80	13.33
Night Check Flight	1,182	0.00	100.00	85.60	15.40
Overall Composite	957	0.00	400.00	363.00	24.71
UPT					
Pass/Fail	3,942	0.00	1.00	0.79	0.40
Phase 2 Check Flight	2,203	6.42	98.90	84.69	15.23
Phase 3 Check Flight	1,867	21.00	100.00	90.52	8.08
Phase 2 Average	2,203	6.54	92.14	72.04	13.02
Phase 3 Average	1,867	24.34	93.73	81.59	7.40
Overall Composite	1,867	58.30	384.77	338.09	21.06

ple sizes were then used in the computation of the Stein's expectancy operator (Kennedy, 1982) to estimate cross-validation shrinkage. The procedure to correct correlations between a dichotomous variable and continuous variables (Hunter & Schmidt, 1990) was applied to the two pass-fail criteria after computation of the effective sample size and Stein's operator in order to produce better estimates of population correlations.

Results

Corrected and uncorrected correlations of the predictor tests and criteria are presented in Appendixes A and B. The UNT and UPT criteria were predicted with samples ranging from 957 to 3,942 subjects. Table 2 shows the descriptive statistics for the criteria and Table 3 shows the regression results.

The r_g column shows the bivariate correlation denoting the predictive efficiency of g, and R_{g+s} is the multiple correlation of g and $s_1 \ldots s_{15}$ with the criteria. The incremental validity ($R_{g+s} - r_g$) of job knowledge and specific abilities appears in the column labeled *Diff.*

The best predictor for all the criteria was g when the data were corrected for range restriction. The average validity coefficient of g was .332 versus the average for job knowledge and specific abilities of .068. There was no overlap in the two ranges (.209 to .523 vs .023 to .115) of correlations.

General ability entered first in all (uncorrected and corrected) but two uncorrected regression equations. The second principal component (s_1) entered first in the uncorrected regressions for UPT Phase 3 Check Flight Average and UPT ATC

Table 3
Regression Results for the 10 Criteria

Criterion	Uncorrected		Corrected		Cross validated	
	r_g	R_{g+s}	r_g	R_{g+s}	R^c_{g+s}	Diff
UNT						
Pass–Fail	.248	.311	.375	.429	.409	.034
Pass–Fail–D[a]	.350	.439	.529	.606	.577	.048
Airmanship	.372	.406	.509	.532	.515	.006
Basic Procedures	.366	.390	.523	.556	.536	.013
Day Check Flight	.136	.172	.242	.292	.290	.048
Night Check Flight	.159	.228	.254	.313	.279	.024
Overall Composite	.306	.336	.462	.490	.482	.020
UPT						
Pass–Fail	.170	.304	.284	.376	.366	.082
Pass–Fail–D[a]	.257	.460	.430	.569	.554	.104
Phase 2 Check Flight	.204	.361	.338	.445	.431	.093
Phase 3 Check Flight	.131	.210	.209	.283	.263	.053
Phase 2 Average	.211	.390	.352	.467	.455	.102
Phase 3 Average	.141	.232	.237	.312	.295	.058
Overall Composite	.175	.316	.314	.403	.398	.084

Note. R^c_{g+s} is the corrected, cross-validated correlation using the Stein Estimator with the effective sample size. Phase 2 and 3 are cumulative averages. Diff = difference.
[a] Correction for dichotomous data.

Phase 3 Average. Other than g, only four non-g principal components entered regression equations frequently and seven never entered the equations.

The cross-validation estimates of the multiple correlation coefficients using effective sample sizes brought an average reduction of approximately .015 across all pilot and navigator criteria.

Discussion

Consistent with past findings, g was the best predictor. The average increment to validity due to job knowledge or specific abilities across the six navigator criteria was .02 and across the six pilot criteria was .08. The smallest increment (.006) was for navigator Airmanship, a job knowledge criterion with aerodynamics, flight instrument and cockpit knowledge, and aircraft emergency procedure content. These results are consistent with the belief that g is strongly related to learning ability (Jensen, 1986). The largest increment to g (.104) was for the pilot pass–fail measure. Overall, job knowledge and specific abilities exhibited greater incremental validity for the pilot criteria than for the navigator criteria.

Those non-g portions of the battery that were predictive of navigator criteria did not overlap with those that were predictive of pilot criteria, nor were they consistent across all navigator criteria. Only g was found in every navigator prediction. For pilots, three predictors entered every equation: g, s_1, and s_3. Although the psychological nature of s_1 and s_3 cannot be assessed with any certainty, they emphasized special knowledge of aviation information and instrument comprehension and should be considered measures of job knowledge. This was consonant with Carretta and Ree (in press), who found that knowledge of aircraft instruments, controls, and terms was a good predictor of pilot training performance. Humphreys (1986) observed similar results during World War II and noted that the job of pilot was especially prone, compared with other jobs, to having the validity of g incremented by tests of job knowledge. Hunter (1983) and Schmidt, Hunter, and Outerbridge (1986) demonstrated the important causal relationship of job knowledge to job performance. The current findings demonstrated the relationship of job knowledge to training performance.

The tests used in this study did not include measures of specialized job knowledge about navigation. There were no questions about sextants, star transits, or global positioning systems. Had such tests been available, greater incremental validity for job knowledge might have been found. However, it is not clear that applicants are exposed to navigator information as frequently as to pilot and aircraft information. This would almost certainly cause the validity of these navigator special knowledge tests to be low. The use of specific job knowledge tests may pose problems for many jobs. Further studies of the incremental validity of specific knowledge should be accomplished to illuminate the issue. For example, do validities for job knowledge hold up over time or decline as people who begin without this knowledge acquire it? The policy consequences of using specific knowledge as a predictor should also be investigated, especially for women and members of minority groups who are less likely to be exposed to information about flying and navigation.

Additionally, the increment found for pilots in this study was about equal to the increment found in Carretta and Ree (in press), who used several different measures of job knowledge or specific ability. A meta-analysis could clarify the relationship of these findings.

Differences in criterion reliability and absolute level of criterion reliability affect validity correlations (Hunter, Schmidt, & Jackson, 1982). The magnitude of a correlation is dependent on reliability of the variables involved. Criterion reliabilities are likely not all the same and would therefore have increased the observed variability of both the correlations of g with the criteria and the specific abilities and job knowledge with the criteria. As no estimates of criterion reliabilities were available, no corrections could be made.

Overall, g was more predictive of navigator than pilot criteria. This difference in average correlational magnitude may be due to course content differences or to differences in reliability of the criterion measures. The cause can not be known from these data.

Additionally, the corrected correlation coefficients were also likely underestimates of the relationship between the g and $s_1 \ldots s_{15}$ and the criteria because the range restriction correction was back to a group of applicants stringently selected for college entry. This is consistent with Thorndike's (1986) explanation:

> One reason that a measure of cognitive ability sometimes does not show up so favorably in relation to other more specialized tests, or in relation to noncognitive measures, is that prior test, educational, or life hurdles have already screened out those low in g, who would have been likely to fail because of limits of cognitive ability. (p. 338)

These results extend the findings for g to new samples and confirm the value of g as a predictor of additional criteria. Again, the incremental validity of $s_1 \ldots s_n$ was small or nonexistent, especially for composite overall performance criteria. Combined with previous results, general cognitive ability continues to be the universal predictor of job and training performance. From jelly rolls (Jensen, 1980) to aileron rolls, g predicts occupational criteria.

References

Brogden, H. E. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. *Educational and Psychological Measurement, 11*, 173–196.

Carey, N. (1992, August). *New predictors of mechanics' job performance: Marine Corps findings.* Paper presented at the annual meeting of the American Psychological Association, Washington, D. C.

Carretta, T. R., & Ree, M. J. (in press). Air Force Officer Qualifying Test validity for predicting pilot training performance. *Journal of Business and Psychology.*

Earles, J. A., & Ree, M. J. (1991). *Air Force Officer Qualifying Test (AFOQT): Estimating the general ability component (AL-TP-1991-0039).* Brooks Air Force Base, TX: Manpower and Personnel Division, Human Resources Directorate, Armstrong Laboratory.

Hotelling, H. H. (1933). Analysis of a complex of statistical variables with principal components. *Journal of Educational Psychology, 24,* 417–441.

Hull, C. L. (1928). *Aptitude testing.* Yonkers, NY: World Book.

Humphreys, L. G. (1986). Commentary. *Journal of Vocational Behavior, 29,* 421–427.

Hunter, J. E. (1983). A causal analysis of cognitive ability, job performance, and supervisory ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257–275). Hillsdale, NJ: Erlbaum.

Hunter, J., & Hunter, R. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin, 96,* 72–98.

Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analyses. *Journal of Applied Psychology, 75,* 334–349.

Hunter, J. E., Schmidt, F. L., & Jackson, G. (1982). *Meta analysis; Cumulating results across studies.* Beverly Hills, CA: Sage.

Jensen, A. R. (1980). *Bias in mental testing.* New York: The Free Press.

Jensen, A. R. (1984). Test validity: *g* versus the specificity doctrine. *Journal of Social and Biological Structures, 7,* 93–118.

Jensen, A. R. (1986). g: artifact or reality? *Journal of Vocational Behavior, 29,* 301–331.

Kennedy, E. (1982). Estimation of the squared cross-validation coefficient in the context of best subset regression. *Applied Psychological Measurement, 12,* 231–237.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulas. *Proceedings of the Royal Society of Edinburgh, Section A, 62, Part I,* 20–30.

McHenry, J. J., Hough, L. M, Toquam, J. L., Hanson, M., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43,* 335–354.

McNemar, Q. (1964). Lost: Our intelligence? Why? *American Psychologist, 19,* 871–882.

Ree, M. J., & Earles, J. A. (1991a). The stability of convergent estimates of g. *Intelligence, 15,* 271–278.

Ree, M. J., & Earles, J. A. (1991b). Predicting training success: Not much more than g. *Personnel Psychology, 44,* 321–332.

Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science, 1,* 86–89.

Ree, M. J., Earles, J. A., & Teachout, M. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology, 79,* 518–524.

Schmidt, F. L., Hunter, J. E., & Larson, M. (1988, August). General cognitive ability vs. general and specific aptitudes in the prediction of training performance: Some preliminary findings. In W. H. Angoff (Chair), *Promising areas for psychometric research.* Symposium conducted at the meeting of the American Psychological Association, Atlanta.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71,* 432–439.

Skinner, J., & Ree, M. J. (1987). *AFOQT Item and Factor Analysis of Form O* (AFHRL-TR-86-68). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Spearman, C. (1904). "General Intelligence" objectively determined and measured. *American Journal of Psychology, 15,* 210–293.

Thorndike, R. L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior, 29,* 332–339.

(*Appendix follows on next page*)

Appendix A

Matrix of Uncorrected (Above Diagonal) and Corrected (Below Diagonal) Correlations of the Criteria and Air Force Officer Qualifying Tests for the Navigators

Criterion	VA	AR	RC	DI	WK	MK	MC	EM	SR	IC	BC	TR	AI	RB	GS	HF	C1	C2	C3	C4	C5	C6
VA	—	.36	.55	.33	.57	.34	.31	.15	.24	.20	.25	.14	.12	.24	.35	.21	.09	.22	.19	.05	.06	.13
AR	.58	—	.39	.55	.35	.57	.25	.23	.53	.20	.30	.26	.07	.25	.27	.23	.24	.22	.28	.10	.15	.22
RC	.73	.58	—	.41	.69	.33	.31	.12	.23	.17	.20	.16	.14	.12	.39	.15	.07	.19	.19	.08	.04	.13
DI	.53	.67	.55	—	.37	.44	.24	.22	.45	.25	.27	.29	.12	.19	.27	.17	.17	.23	.23	.15	.13	.23
WK	.68	.46	.77	.46	—	.29	.25	.06	.19	.15	.16	.15	.17	.12	.40	.13	.05	.21	.14	.04	.05	.11
MK	.55	.71	.51	.60	.40	—	.26	.19	.40	.18	.22	.22	.04	.21	.36	.20	.20	.22	.30	.13	.13	.22
MC	.48	.51	.46	.46	.40	.48	—	.31	.20	.34	.25	.08	.39	.37	.48	.22	.12	.27	.21	.02	.03	.14
EM	.27	.37	.23	.38	.17	.40	.44	—	.26	.30	.34	.20	.15	.28	.20	.21	.10	.09	.10	.04	.09	.13
SR	.48	.66	.45	.62	.37	.60	.48	.51	—	.26	.38	.36	.06	.27	.19	.26	.23	.19	.27	.13	.15	.22
IC	.34	.41	.33	.43	.28	.39	.49	.44	.49	—	.29	.18	.42	.32	.32	.19	.12	.22	.22	.05	.05	.15
BC	.45	.53	.40	.51	.32	.49	.50	.47	.61	.49	—	.38	.07	.37	.16	.28	.17	.17	.15	.09	.09	.16
TR	.34	.44	.35	.47	.27	.44	.30	.31	.56	.34	.51	—	.00	.18	.05	.15	.17	.15	.19	.14	.13	.22
AI	.30	.31	.34	.34	.32	.25	.50	.29	.33	.56	.31	.21	—	.15	.42	.07	.04	.26	.16	.00	−.04	.05
RB	.43	.47	.35	.42	.29	.49	.54	.42	.49	.46	.55	.34	.34	—	.20	.28	.19	.18	.21	.07	.14	.21
GS	.51	.49	.55	.44	.51	.52	.57	.34	.41	.41	.37	.25	.46	.40	—	.18	.04	.25	.22	.02	.00	.11
HF	.40	.40	.36	.39	.31	.40	.39	.34	.47	.36	.45	.36	.27	.42	.31	—	.12	.16	.15	.05	.12	.17
C1	.22	.35	.19	.29	.14	.33	.26	.20	.35	.24	.30	.28	.17	.30	.17	.22	—	.40	.48	.33	.31	.34
C2	.37	.39	.34	.38	.32	.37	.40	.22	.37	.34	.34	.30	.36	.34	.36	.30	.40	—	.58	.17	.18	.50
C3	.37	.45	.36	.39	.27	.46	.37	.24	.44	.35	.35	.36	.30	.37	.35	.31	.48	.58	—	.25	.19	.52
C4	.14	.18	.16	.22	.10	.21	.11	.11	.22	.13	.18	.22	.08	.16	.11	.13	.33	.17	.25	—	.17	.65
C5	.14	.21	.11	.20	.10	.21	.12	.14	.22	.12	.17	.20	.04	.21	.09	.18	.31	.18	.19	.17	—	.77
C6	.28	.35	.27	.35	.21	.37	.29	.23	.37	.27	.31	.35	.19	.34	.24	.29	.34	.50	.52	.65	.77	—
M	15.10	13.18	18.52	13.56	15.04	16.61	12.22	9.38	24.55	14.09	13.05	30.50	14.04	9.34	10.35	10.49						
SD	3.23	3.66	4.56	3.42	5.09	5.00	3.11	4.17	5.32	4.13	3.58	5.94	4.05	2.81	3.35	2.48						

Note. C1 is Undergraduate Navigator Training Pass/Fail, C2 is Airmanship, C3 is Basic Procedures, C4 is Day Celestial Check Flight, C5 is Night Celestial Check Flight, and C6 is the Composite of C1 through C5. $N = 957$. See Table 1 for criteria names.

Appendix B

Matrix of Uncorrected (Above Diagonal) and Corrected (Below Diagonal) Correlations of the Criteria and Air Force Officer Qualifying Tests for the Pilots

Criterion	VA	AR	RC	DI	WK	MK	MC	EM	SR	IC	BC	TR	AI	RB	GS	HF	C1	C2	C3	C4	C5	C6
VA	—	.39	.60	.36	.59	.40	.33	.12	.24	.13	.20	.11	.05	.22	.40	.20	.00	.01	.00	.00	.00	.00
AR	.58	—	.42	.55	.33	.58	.31	.22	.52	.19	.32	.29	.05	.27	.31	.20	.11	.13	.07	.14	.08	.12
RC	.73	.58	—	.41	.69	.39	.32	.12	.26	.13	.18	.16	.09	.15	.44	.15	.03	.04	.01	.04	.02	.10
DI	.53	.67	.55	—	.34	.46	.29	.23	.48	.25	.32	.31	.11	.24	.28	.20	.12	.15	.07	.16	.08	.12
WK	.68	.46	.77	.46	—	.33	.28	.04	.20	.11	.13	.09	.12	.14	.43	.14	−.02	−.02	−.02	−.01	−.02	−.02
MK	.55	.71	.51	.60	.40	—	.31	.24	.42	.15	.30	.28	−.06	.291	.38	.24	.04	.07	.03	.06	.03	.04
MC	.48	.51	.46	.46	.40	.48	—	.31	.24	.29	.27	.08	.26	.38	.46	.23	.10	.15	.08	.16	.09	.11
EM	.27	.37	.23	.38	.17	.40	.44	—	.27	.26	.34	.20	.05	.30	.19	.23	.08	.07	.08	.07	.09	.09
SR	.48	.66	.45	.62	.37	.60	.48	.51	—	.28	.40	.39	.08	.29	.20	.26	.17	.20	.12	.20	.13	.18
IC	.34	.41	.33	.43	.28	.39	.49	.44	.49	—	.27	.20	.40	.29	.23	.20	.23	.26	.15	.28	.17	.23
BC	.45	.53	.40	.51	.32	.49	.50	.47	.61	.49	—	.37	.00	.35	.17	.27	.10	.09	.09	.09	.11	.10
TR	.34	.44	.35	.47	.27	.44	.30	.31	.56	.34	.51	—	.00	.15	.07	.18	.15	.12	.13	.12	.13	.15
AI	.30	.31	.34	.34	.32	.25	.50	.29	.33	.56	.31	.21	—	.09	.24	.04	.18	.24	.10	.27	.12	.21
RB	.43	.47	.35	.42	.29	.49	.54	.42	.49	.46	.55	.34	.34	—	.26	.29	.10	.10	.05	.10	.06	.08
GS	.51	.49	.55	.44	.51	.52	.57	.34	.41	.41	.37	.25	.46	.40	—	.18	.04	.04	.01	.04	.02	.04
HF	.40	.40	.36	.39	.31	.40	.39	.34	.47	.36	.45	.36	.27	.42	.31	—	.06	.04	.07	.04	.07	.08
C1	.22	.35	.19	.29	.14	.33	.26	.20	.35	.24	.30	.28	.17	.30	.17	.15	—	.67	.74	.68	.76	.69
C2	.37	.39	.34	.38	.32	.37	.40	.22	.37	.34	.34	.30	.36	.34	.36	.15	.40	—	.39	.98	.43	.71
C3	.37	.45	.36	.39	.27	.46	.37	.24	.44	.35	.35	.36	.30	.37	.35	.14	.48	.42	—	.43	.95	.90
C4	.14	.18	.16	.22	.10	.21	.11	.11	.22	.13	.18	.22	.08	.16	.11	.16	.33	.98	.45	—	.49	.75
C5	.14	.21	.11	.20	.10	.21	.12	.14	.22	.12	.17	.20	.04	.21	.09	.15	.31	.45	.45	.51	—	.92
C6	.12	.22	.14	.22	.07	.17	.23	.19	.28	.31	.23	.24	.28	.19	.15	.18	.70	.73	.90	.76	.92	—
M	15.32	13.27	18.84	13.18	16.08	16.57	11.61	8.85	23.50	11.26	12.71	29.80	10.75	8.93	9.98	10.43						
SD	3.22	3.76	4.56	3.53	5.14	4..73	3.10	4.11	5.30	4.40	3.74	6.00	4.00	2.85	3.38	2.50						

Note. C1 is Undergraduate Pilot Training Pass/Fail, C2 is Phase 2 Checkride, C3 is Phase 3 Checkride, C4 is Phase 2 Average, C5 is Phase 3 Average, and C6 is the Composite of C1 through C5. N = 1,867. See Table 1 for criteria names.

Received January 4, 1993
Revision received March 4, 1994
Accepted March 7, 1994 ■