A Logistic Regression Model for Personnel Selection

Nambury S. Raju
Illinois Institute of Technology

Stephen D. Steinhaus
Ameritech Services, Inc.

Jack E. Edwards
Navy Personnel Research and Development Center

Juyne DeLessio
Saint Joseph College of Nursing

A two-parameter logistic regression model for personnel selection is proposed. In addition to presenting a theoretical basis for the model, a unified approach is provided for studying selection, validity generalization, employee classification, selection bias, and utility-based fair selection. The new model was tested with a large database (N = 84,808). Results show the logistic regression model to be valid and also quite robust with respect to direct and indirect range restriction on the predictor. *Index terms: logistic regression, personnel selection, selection bias, utility-based fair selection, validity generalization.*

Personnel psychologists are often interested in an applicant's probability of success for a given job. To assess this probability, personnel psychologists traditionally have assumed that a certain trait (e.g., ability, aptitude, or personality trait) is functionally related to the probability of job success, and they have relied almost exclusively on the correlation coefficient to study the relationship. Unfortunately, for several reasons, a correlation coefficient does not fully express this relationship (Gulliksen, 1986; Linn, 1978). For example, a correlation coefficient does not result in a statement about the probability of success. Currently, such a statement can only be formulated indirectly with the help of Taylor-Russell tables (Taylor & Russell, 1939) if bivariate normality exists between predictor and criterion.

An alternative approach to personnel selection is proposed here that directly relates the probability of job success to trait levels. This new approach is capable of simultaneously addressing such topics as selection, validity generalization, employee classification, selection bias, and utility-based fair selection when the probability of success is of primary concern.

A Model for Personnel Selection

The criterion (job performance) score and predictor (trait) score will be denoted as y and x, respectively. The joint probability density function (PDF) of y and x may then be denoted as $H(y,x)$ and

$$H(y,x) = H(y|x)h(x) \quad , \tag{1}$$

where $H(y|x)$ is the conditional PDF of y given x, and $h(x)$ is the PDF of x. An applicant with $y \geq y_0$ is considered successful, and similarly, anyone with $x \geq x_0$ is selected. In view of these definitions,

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 15, No. 2, June 1991, pp. 139-152

0146-6216/91/020139-14$1.95

the probability of success on a job given x [$P(x)$] can be written as

$$P(x) = \int_{y_0}^{\infty} H(y|x)dy \quad . \tag{2}$$

This equation is not very practical unless the specific form of $H(y|x)$ is known. As a solution to this problem, $P(x)$ is proposed as a two-parameter logistic regression (LR) function:

$$P(x) = \frac{e^{\mathrm{D}a(x-b)}}{1 + e^{\mathrm{D}a(x-b)}} \quad , \tag{3}$$

where D is a constant that is usually set equal to 1.7 in order to make $P(x)$ correspond to a normal ogive, and a and b are the job parameters to be estimated (Lord & Novick, 1968). Unlike Equation 2, Equation 3 does not contain y_0. However, a person is considered successful if she or he scores at or above y_0. Therefore, Equation 3 also depends on y_0, but its dependence is implicit. In Figure 1, $P(x)$ represents a job characteristic curve (JCC).

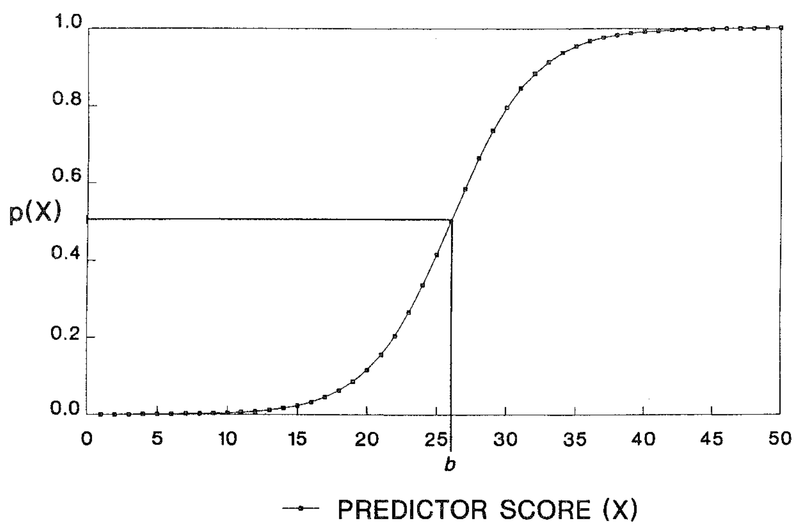
A JCC is thus comparable to an individual expectancy chart (McCormick & Ilgen, 1980). By definition (as well as in Figure 1), b is the level of the trait needed in order to have a 50% probability of job success, and a is proportional to the slope of the JCC at b. A theoretical justification for $P(x)$ (Equation 3) is given in Appendix A.

Implications of the Proposed Model

Validity Generalization

JCCs are useful for studying validity generalization (VG; Callender & Osburn, 1980; Pearlman, Schmidt, & Hunter, 1980; Raju & Burke, 1983; Schmidt, Gast-Rosenberg, & Hunter, 1980; Schmidt & Hunter, 1977). If the same test is used to predict success on two similar jobs in two different locations, the same a and b values should be expected at both locations. Failure to find approximately the same as and bs means that the JCCs are different for the two locations and that validity is

Figure 1
An Example of a Job Characteristic Curve



not generalizable. This procedure can be extended easily to include more than two locations and different predictors that are designed to measure the same underlying trait—provided the predictors have comparable metrics.

Employee Classification

In classification, each applicant has to be assigned to one of a number of jobs. For example, the JCCs for Jobs 1 and 2 may be denoted by $P_1(x)$ and $P_2(x)$, respectively. If $P_1(x) > P_2(x)$ for an applicant, the applicant is assigned to Job 1 because he or she has a higher probability of success on Job 1. Conversely, if $P_1(x) < P_2(x)$, the applicant is assigned to Job 2. Finally, if $P_1(x) = P_2(x)$, assignment of the applicant to either job is acceptable from the point of view of the applicant's probability of success. This simple procedure can be extended easily to include as many jobs as necessary, and is similar in some ways to univariate discriminant analysis (Tatsouka, 1988).

Selection Bias

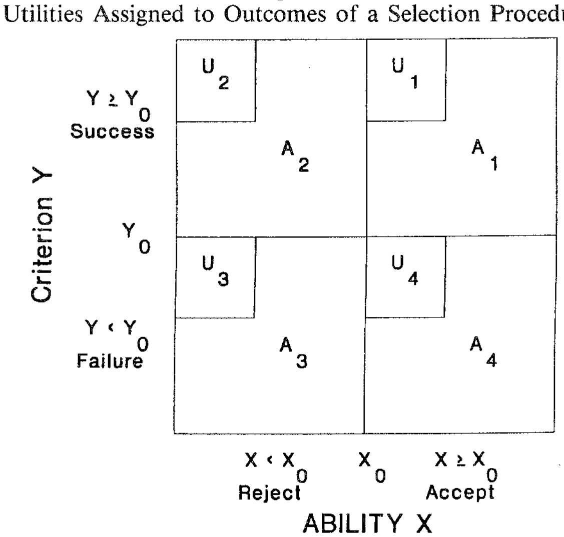
The JCCs for majority and minority applicants can be used to address the question of selection bias. Failure to obtain the same JCCs for majority and minority groups on a test that is relatively free of item bias implies that the criterion variable is measuring different variables for the two groups. The different JCCs, in turn, mean that the probability of job success is a function of x and group membership (i.e., a clear case of selection bias). Therefore, the LR model offers a simple, easily understood operational definition of selection bias.

Utility-Based Fair Selection

The use of utilities in fair selection has been considered by several investigators. The work of Gross and Su (1975) was based on the assumption that the joint distribution of job success and test (predictor) score is bivariate normal. Petersen's (1976) work emphasized the use of Bayesian principles. Using x_0 and y_0 as defined earlier, examinees with a criterion score greater than or equal to y_0 are considered successful, and those with a predictor score greater than or equal to x_0 are selected.

Figure 2 shows the four outcomes (A_i) and four utilities (U_i) that result from y_0 and x_0. According

Figure 2
Utilities Assigned to Outcomes of a Selection Procedure



to Gross and Su (1975) and Petersen (1976), the optimal selection procedure is one in which the value for x_0 is selected to maximize the expected utility. In the case of two groups, the optimal selection procedure determines cutting scores (x_0 for Group 1 and x'_0 for Group 2) that maximize the combined expected utility.

Based on Figure 2, the combined expected utility E(U) for the two-group case can be written as

$$\mathrm{E}(U) = Q\sum_{i=1}^{4} U_i P(A_i) + Q'\sum_{i=1}^{4} U'_i P(A'_i) \quad , \tag{4}$$

where Q and Q' are the proportion of applicants in Groups 1 and 2, respectively, such that $Q + Q' = 1$;
$P(A_i)$ and $P(A'_i)$ are the probabilities of outcomes A_i and A'_i, respectively;
U_i and A_i represent Group 1's utilities and outcomes; and
U'_i and A'_i represent Group 2's utilities and outcomes, respectively.

To obtain this optimal solution, both quota-free and fixed-quota selection must be considered. In quota-free selection, there is no limit on the number of applicants who will be hired. In fixed-quota selection, the proportion of applicants to be hired is fixed—that is, the selection ratio (SR) is fixed.

Fixed-quota selection. Within the context of the proposed model, optimal utility-based fair selection can be achieved if x_0 and x'_0 are selected to satisfy

$$P(x_0)(U_2 - U_1 - U_3 + U_4) + (U_3 - U_4) = P'(x'_0)(U'_2 - U'_1 - U'_3 + U'_4) + (U'_3 - U'_4) \quad , \tag{5}$$

subject to the restriction that the total number of applicants selected from both groups is consistent with the overall proportion of applicants to be hired (i.e., overall SR). Equation 5 is similar to formulas obtained by Gross and Su (1975) and Petersen (1976), with the exception that here $P(x)$ is a two-parameter logistic regression function. Proof of Equation 5 is given in Appendix B.

When the utilities are the same for both groups, Equation 5 can be written as

$$P(x_0) = P'(x'_0) \quad . \tag{6}$$

If the JCCs are the same for the two groups ($a = a'$ and $b = b'$), Equation 6 will be satisfied if and only if $x_0 = x'_0$. Therefore, the cutting scores of Groups 1 and 2 must be identical, and the number of applicants selected must equal SR. Furthermore, the same SRs will result for each group if the distributions of x for the two groups are the same. There will be no adverse impact in such a selection strategy (this is as it should be when the utilities and the distributions of x are equal for the two groups). On the other hand, if the two distributions of x are unequal and $x_0 = x'_0$, the SRs for the two groups will be different. Under those conditions, there may be adverse impact. Such a situation can be avoided by defining the utilities differently for the two groups.

Quota-free selection. The optimal solution for quota-free selection is a special case of the optimal solution for fixed-quota selection, in the sense that cutting scores x_0 and x'_0 must satisfy only Equation 5 without any regard for how many applicants will be selected. If there is only one group, the optimal solution for x_0 can be obtained from

$$P(x_0)(U_2 - U_1 - U_3 + U_4) + (U_3 - U_4) = 0 \quad . \tag{7}$$

Equations 5 and 7 display the significant role that the two-parameter LR function plays in utility-based fair selection.

When the utilities associated with the rejected applicants are so minimal that the utilities may be treated as 0 (i.e., $U_2 = U_3 = 0$), Equation 7 can be written as

$$P(x_0) = U_4/(U_4 - U_1) \quad . \tag{8}$$

In situations in which utilities U_1 and U_4 can be considered opposite in sign but equal absolute values, Equation 8 can be rewritten as

$$P(x_0) = \frac{1}{2} \quad . \tag{9}$$

In this special case, $x_0 = b$ will satisfy Equation 9 and be the optimal cutting score. Thus, another interpretation of the b parameter is offered in this case.

Advantages of the Logistic Model

There are several theoretical and practical advantages for using logistic regression over linear regression when probability of success is a primary concern.

1. The mathematical expression for LR given in Equation 3 is significantly less complex than the probability of success function (normal ogive) given in Equation A7 for linear regression. However, as previously noted, both equations yield very similar results when $D = 1.7$ (Lord & Novick, 1968). Because of its simplicity, the LR function is more tractable when used in other contexts (e.g., utility analysis).
2. It is easier to understand the central role that probability of success plays with LR than with linear regression in such areas as selection bias, utility analysis, and VG. Within the context of linear regression, for example, correlation coefficients are used for VG analysis, and probability of success functions (e.g., Equation A7) are used for utility analysis. With LR, the same $P(x)$ is used for both purposes. Although the probability of success function can be derived from the correlation coefficient, the use of a function such as $P(x)$ in both VG and utility analysis emphasizes the commonality between these two areas and might make them easier to understand and implement in practice. Selection bias is another area in which LR offers a definite theoretical and practical advantage over linear regression. Selection bias using LR can readily be equated to item bias, and it can benefit from the extensive published research currently available on item bias.
3. The standard error of an observed correlation coefficient does not vary from one predictor score to the next, whereas the standard error of $P(x)$ in Equation 3 depends on x. Therefore, the information that LR provides about the precision of measurement is more useful.
4. Because LR is used in item response theory, $P(x)$ can be considered to be subpopulation invariant, whereas the correlation coefficient is known to be affected by range restriction (Lord & Novick, 1968).

Method

Examinees

The effect of direct and indirect restrictions of range on JCCs was empirically assessed using data from 84,808 U.S. Air Force enlistees who were tested with forms 8, 9, and 10 of the Armed Services Vocational Aptitude Battery (ASVAB). The enlistees were tested at training centers in several regions of the U.S. After initial training, the enlistees were assigned to 1 of 70 Air Force technical training schools in which each enlistee was assigned a final grade on graduation (Wilbourn, Valentine, & Ree, 1984).

Variables

ASVAB tests. The ASVAB was originally developed in 1976 around the national mobilization population (i.e., males between the ages of 17 and 24). Each of the current forms (8, 9, and 10) was constructed to be parallel in terms of number-correct scores. Each form contained 10 tests: General Science, Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, Numerical Operations, Coding Speed, Auto and Shop Information, Math Knowledge, Mechanical Comprehension, and Electronic Information.

A composite of four ASVAB tests (called the Armed Forces Qualifying Test) is used to select potential recruits for all four branches of the U.S. military. The Air Force then uses ASVAB-based aptitude indices to select and classify its enlistees for training and military jobs. As a result, the distributions of test scores for a given technical school may be restricted on one or both ends. Low-scoring enlistees are screened into technical schools with less rigorous standards, and the highest-scoring enlistees are often assigned to technical schools with higher admission standards and a curriculum that is relatively more difficult.

Final school grade (FSG). ASVAB Forms 8, 9, and 10 were validated using FSGs as the criterion. FSGs may vary for each school from 1 to 99, with a FSG below 70 indicating a failure. No FSG was recorded in the enlistee's file if a failure occurred. Therefore, the distribution of criterion scores was restricted, because each of the 84,808 cases used in this study had a FSG.

Procedure

For the purposes of this study, FSGs at or above the 50th percentile (FSG $\geq$ 84) for the total population were coded as successful. FSGs less than 84 were considered to be unsuccessful. Math Knowledge and Mechanical Comprehension test scores and FSGs were used. Some summary data and the intercorrelations between the three measures are presented in Table 1. The Kuder-Richardson 20 estimates of reliability for the 25-item Math Knowledge and 25-item Mechanical Comprehension tests were .87 and .85, respectively (Wilbourn et al., 1984). The Math Knowledge and Mechanical Comprehension tests were selected for this study because their validities with FSGs (.39 and .31, respectively) were similar to the validity coefficients typically reported in the personnel psychology literature.

Phase 1. Using the dichotomous FSGs as the criterion and the Math Knowledge test as the predic-

Table 1
Intercorrelations and Summary Descriptive Statistics for the Math Knowledge (MK) and Mechanical Comprehension (MC) Tests and the FSG Criterion (N = 84,808)

Statistic and Variable	MK	MC	FSG
Correlations			
MC	.41	1.00	
FSG	.39	.31	1.00
Descriptive Statistics			
Mean	15.291	16.967	83.376
Median	14.923	17.274	84.022
SD	4.954	4.213	7.894
Skewness	.125	−.317	−.420
Kurtosis	−.851	−.489	−.224

tor, the two-parameter LR model was applied to the population of 84,808 cases. The empirical odds of success were also determined for each of the 25 possible scores on the Math Knowledge test simply by dividing the number of examinees with that total test score who also had criterion scores at or above 84 by the total number of examinees at that test score. (Seven examinees with 0 scores on the Math Knowledge test were excluded from both Phases 1 and 2.)

The goodness of fit between the LR probabilities and empirical (E) probabilities was assessed with the weighted root-mean-squared error (WRMSE) and weighted average absolute deviation (WAAD) indices:

$$\text{WRMSE} = \left[\frac{\Sigma N_i (P_i - P_i')^2}{\Sigma N_i}\right]^{1/2} \tag{10}$$

and

$$\text{WAAD} = \frac{\Sigma N_i |P_i - P_i'|}{\Sigma N_i} \quad , \tag{11}$$

where P_i is the empirical probability,
P_i' is the theoretical probability,
N_i is the number of examinees at score i,
and the summation is taken over all possible scores.

Phase 2. The second phase employed a χ^2 test (Bock, 1975) to assess how well the sample-based LR-JCCs matched the sample-based E-JCCs. Because only sample-based E-JCCs are typically available in practice, such χ^2 tests are important in determining whether or not LR is a viable option for a given dataset. In view of the availability of the population-based E-JCC and LR-JCC in the current investigation, the second phase also assessed how well the population-based E-JCC and LR-JCC (determined in Phase 1) were predicted by sample-based LR-JCCs.

Random selection of 1,000 samples of 1,000 cases with replacement was performed from the 84,808 cases in the population. For each sample of 1,000 cases, three separate LR analyses were performed:

1. Each complete sample was used to develop an LR equation. The χ^2 test was used to determine how well each LR equation fit its sample-based empirical probabilities of success. Also, WRMSE and WAAD were used to compare each resulting sample-based LR-JCC to the population-based E-JCC and LR-JCC.
2. A subsample of only those examinees whose Math Knowledge number-correct scores were at or above the population median of 15 was selected from its sample of 1,000 cases. Each subsample was then used to develop a subsample-based LR-JCC. As in the preceding analysis, the χ^2 test was used to assess how well the LR model fit the sample-based empirical probabilities. Then the sample-based LR-JCC was compared to the population-based E-JCC and LR-JCC (again using the WRMSE and WAAD indices). The purpose of this analysis was to study the effect of direct range restriction on the accuracy of sample-based LR-JCCs.
3. The procedure for the second LR analysis was followed for the third LR analysis—with one exception: the subsample that was extracted from the full sample of 1,000 cases consisted of only those examinees whose number-correct scores on the Mechanical Comprehension test were at or above the population median of 17. This analysis was done to study the effect of indirect range restriction on the accuracy of sample-based LR-JCCs.

Results and Discussion

Phase 1. Table 2 shows the observed frequencies of successful and unsuccessful examinees for each number-correct score on the Math Knowledge test. The LR estimates of a and b were .086 and 14.59, respectively. Table 2 also contains the population-based E and LR probabilities of success. The greatest difference between the E and LR probabilities occurred for scores from 1 through 6. For example, that difference was .16 for a score of 4. It should be noted that the number of successful and unsuccessful examinees at those six points was quite small compared to the frequencies at other scores. The differences for the highest two scores were also marked by differences greater than .05. In contrast, the greatest difference between empirical and theoretical probabilities was .04 for scores of 7 through 23. Moreover, the median difference for scores in that middle range was .01.

When the overall empirical and theoretical probability difference was examined, the WRMSE and WAAD indices were .021 and .019, respectively. The .021 WRMSE means that, on average, the E and LR probabilities differed by approximately .02; a similar interpretation applies to the WAAD index. Thus, the magnitudes of the goodness-of-fit indices imply that the two-parameter LR model provided a good fit for the empirical probabilities in the current dataset.

Phase 2. For the no-range-restriction condition, the average χ^2 between the sample-based LR-JCC and the sample-based E-JCC was 24.90 with degrees of freedom (df) ranging from a minimum of 19 to a maximum of 22 (see Table 3). Of the 1,000 χ^2 values, 2% were significant at the .01 level.

Table 2
Observed Frequencies of Successful and Unsuccessful Examinees and Empirical and Theoretical Probabilities From the Logistic Regression Analysis in Phase 1
(a = .086, b = 14.59, WRMSE = .021, and WAAD = .019)

Math Score	Successful	Unsuccessful	Empirical Probability	Theoretical Probability
1	1	1	.25	.13
2	3	13	.19	.15
3	10	36	.22	.17
4	66	123	.35	.19
5	119	369	.24	.21
6	292	716	.29	.23
7	557	1,364	.29	.26
8	935	2,010	.32	.29
9	1,437	2,709	.35	.32
10	1,722	3,297	.34	.35
11	2,259	3,563	.39	.38
12	2,508	3,692	.40	.41
13	2,716	3,587	.43	.45
14	2,783	3,160	.47	.48
15	2,759	2,791	.50	.52
16	2,899	2,531	.53	.55
17	2,925	2,102	.58	.59
18	2,868	1,871	.61	.62
19	2,760	1,581	.64	.65
20	2,688	1,360	.66	.68
21	2,720	1,103	.71	.71
22	2,610	885	.75	.74
23	2,621	722	.78	.77
24	2,501	437	.85	.79
25	1,791	226	.89	.81

Table 3
Mean and SD of WRMSE and WAAD for E-JCC and LR-JCC, and for χ^2 Between E-JCC and LR-JCC, for Three Range Restriction Conditions

	No Range Restriction		Direct Range Restriction		Indirect Range Restriction	
Statistic	E-JCC	LR-JCC	E-JCC	LR-JCC	E-JCC	LR-JCC
WRMSE						
Mean	.031	.019	.049	.049	.063	.055
SD	.006	.006	.010	.010	.017	.019
WAAD						
Mean	.024	.016	.042	.042	.058	.058
SD	.006	.009	.010	.010	.016	.018
χ^2						
Mean	24.90		9.89		25.60	
SD	7.60		4.59		7.70	

The average differences between sample-based LR-JCCs and the population-based E-JCC were .031 and .024 for WRMSE and WAAD, respectively. These latter goodness-of-fit statistics also show that the sample-based LR-JCCs were slightly closer to the population-based LR-JCC than they were to the population-based E-JCC (.031 versus .019 and .024 versus .016 for WRMSE and WAAD, respectively). The variability of the goodness-of-fit indices, however, was slightly smaller for the comparison involving the population-based E-JCC. Overall, the sample-based LR-JCCs seemed to offer a reasonable fit to sample-based E-JCCs and the population-based LR-JCC and E-JCC.

For the direct-range-restriction condition, the average χ^2 between sample-based LR and empirical probabilities was 9.89 with 9 *df.* As in the case of no range restriction, only 2% of the 1,000 χ^2s were statistically significant at the .01 level, which indicates an acceptable fit for the LR model. The WRMSE and WAAD means and standard deviations were somewhat higher for the direct-range-restriction condition than for the no-range-restriction condition. The sample-based JCCs fit the population-based E-JCC and LR-JCC exactly the same. The average differences in fit were approximately .05 for the WRMSE index and approximately .04 for the WAAD index. Thus, the direct range restriction employed in this study (i.e., rejecting examinees who scored below the population median) appeared to have a small negative effect—as compared to the no-range-restriction condition—on the accuracy of the sample-based JCCs developed with the two-parameter LR model.

The effect of indirect range restriction on the accuracy of sample-based JCCs in predicting population-based JCCs was similar to the effect observed under direct range restriction. The average χ^2 between sample-based LR-JCCs and sample-based empirical probabilities was 25.60, with *df* ranging from 18 to 21. Three percent of the 1,000 χ^2s were significant at the .01 level. All four of the WRMSE and WAAD indices indicated approximately .06 difference. The standard deviations of WRMSE and WAAD indices for indirect range restriction were also unappreciably greater than their counterparts for direct range restriction. Therefore, the accuracy under direct range restriction was only slightly better than the accuracy under indirect range restriction.

Summary and Conclusions

These data suggest two major conclusions:

1. The two-parameter LR model provided theoretical probabilities for the entire dataset of 84,808 cases that fit the empirical probabilities reasonably well.

2. When random samples were used, the fit between sample-based JCCs and population-based JCCs continued to be good, as measured by the χ^2 statistic and the WRMSE and WAAD indices. Even though the observed percentages of significant χ^2s were higher than would be expected by chance alone, they did not appear to be excessive for the three conditions: 2% for no- and direct-range-restriction conditions and 3% for indirect-range-restriction conditions. As might be expected, the WRMSE and WAAD indices reflecting the fit of the sample-based JCCs to the population-based JCCs for the directly and indirectly restricted samples were larger than those obtained for the full random sample. However, the fit indices for the nonrandom samples may be acceptable.

Overall, the two-parameter LR model appeared to offer a promising alternative to studying the question of the probability of success in selection.

The results of this study appear to support the proposed model, given three probable limitations of the dataset. First, the probability of success in all Air Force jobs/training was examined, rather than the probability of success for a given family of jobs or technical schools. It may be the case that training performance in many of the technical schools does not require Mechanical Comprehension, and that to a lesser degree, it does not require Math Knowledge. This research could have been limited to those training programs for which the Math Knowledge and Mechanical Comprehension tests were valid. However, such a study would likely result in a population size significantly smaller than the 84,808 cases used here.

Second, the standards for the 1 to 99 FSGs may have varied across schools, in addition to the more typical variability found in industry across raters. Third, the range of criterion scores was truncated even in the population. The absence of persons who were truly unsuccessful (FSG $<$ 70) may have, in part, caused some of the differences between the empirical versus the LR-derived probabilities at the lower score points. It should be noted, however, that the FSG data in Table 1 appear not to reflect a great deal of truncation, which indicates that the failure rate was probably minimal.

Given the success of the model in this study, other issues addressed by the model (i.e., validity generalization, employee classification, selection bias, and utility-based fair selection) need to be empirically investigated to determine if their theoretical bases hold up as well as the theoretical foundation for personnel selection. Furthermore, future studies using the two-parameter LR model of selection should pay careful attention to certain considerations.

Samples of 1,000 cases in Phase 2 were used for the no-range-restriction condition. The average subsample sizes across 1,000 replications for direct and indirect range restriction were 527 and 565, respectively. However, sample sizes for most validation studies are typically much smaller. Therefore, the question of what constitutes an adequate sample size for reliable estimates of LR-JCCs should be investigated. Large sample sizes were used in the present investigation to test the proposed model for personnel selection with samples of adequate size.

The proposed model assumes that there is only one predictor variable. Because the criterion of job performance is usually considered to be multidimensional in nature (Smith, 1976), the current model should be expanded to include multivariate logistic regression (Neter & Wasserman, 1974). Such an extension is straightforward, and computer packages such as SAS (1985) and SPSS-X (1988) can readily handle the data analyses required by the expanded model. An extension of the model to situations in which the criterion variable is measured on a Likert-type scale is also needed. Samejima's (1969) two-parameter model for graded response data is a viable option in this context.

Sample-based JCCs in Phase 2 were compared here to the population-based JCCs. Although such comparisons are appropriate and desirable, especially in evaluating a new model, the population-based JCCs are typically unavailable in most practical situations. When only sample data are available, it is essential that the goodness of fit between the sample-based LR-JCC and E-JCC be evaluated statistically—as was done in the current investigation—prior to accepting the sample-based LR-JCC. Just as some observed correlations or validity coefficients may not be significantly different from zero, the LR procedure may not yield an adequate fit for some datasets. The proposed LR model should not be used if the goodness of fit is inadequate.

In spite of the general favorability of the findings, some caution seems warranted. Any generalization of the conclusions from this study must take into consideration the fact that both the dichotomization of the FSG criterion and the introduction of direct and indirect range restriction were based on the median scores for the three variables investigated. The degree to which alternative methods of determining cutoffs affect the accuracy of the results obtained with the proposed model must still be investigated. The effectiveness of the proposed model should also be empirically evaluated with datasets that are more typical of those encountered in personnel psychology.

Appendix A: Theoretical Justification for *P*(*x*)

The usefulness of any mathematical model ultimately depends on how well it explains empirical data. Although a complete empirical verification of the model as given in Equation 3 will obviously take time, a theoretical justification for $P(x)$ can be readily provided.

The LR model is similar to some of the mathematical models that are popular in psychology. In order to demonstrate this similarity, assume that y and x have a bivariate normal distribution. Then the probability of success (i.e., $y \geq y_0$) for a given x can be written as (Hogg & Craig, 1970, p. 112)

$$\mathrm{F}(x) = [\sigma_y 2\pi(1 - \rho^2)]^{-1/2} \int_{y_0}^{\infty} \exp[-(y - s)^2/2\sigma_y^2(1 - \rho^2)dy] \quad , \tag{A1}$$

where $s = \mu_y + \rho(\sigma_y/\sigma_x)(x - \mu_x)$; and μ, σ, and ρ represent the mean, standard deviation, and correlation between y and x, respectively. If y and x are standardized so that $\mu_y = \mu_x = 0$ and $\sigma_y = \sigma_x = 1$, then $s = \rho x$ and Equation A1 can be rewritten as

$$\mathrm{F}(x) = [2\pi(1 - \rho^2)]^{-1/2} \int_{y_0}^{\infty} \exp[-(y - s)^2/2(1 - \rho^2]dy \quad . \tag{A2}$$

Furthermore, following Lord and Novick (1968),

$$t = (y - s)(1 - \rho^2)^{-1/2} \quad , \tag{A3}$$

$$-L = (y_0 - s)(1 - \rho^2)^{1/2} \quad , \tag{A4}$$

$$a = \rho(1 - \rho^2)^{-1/2} \quad , \tag{A5}$$

and

$$b = y_0\rho^{-1} \quad , \tag{A6}$$

so that $-L = a(b - x)$, because $s = \rho x$. Using the above transformations, Equation A2 can be rewritten as

$$\begin{aligned} \mathrm{F}(x) &= (2\pi)^{-1/2} \int_{-L}^{\infty} \exp(-t^2/2)dt \\ &= (2\pi)^{-1/2} \int_{-\infty}^{L} \exp(-t^2/2)dt \\ &= (2\pi)^{-1/2} \int_{-\infty}^{a(x-b)} \exp(-t^2/2)dt \quad . \end{aligned} \tag{A7}$$

Equations A2 and A7 are of special importance in the present context because (1) they are mathematically identical; (2) Equation A7 is merely the two-parameter normal ogive function (Lord & Novick, 1968); and (3) $P(x)$ is almost identical to F(x), because the normal ogive function is practically identical to the two-parameter LR function when D = 1.7 (Lord & Novick, 1968).

Even though F(x) depends on x and y_0, and $P(x)$ appears to depend only on x, the definition above of a successful employee notes that $P(x)$ also depends implicitly on y_0. This shows the extremely close relationship between this model and the normal ogive models of Alf and Dorfman (1967) and Gross and Su (1975). Furthermore, $P(x)$ can also be shown to be closely related to the probabilities given in the Taylor-Russell tables (Taylor & Russell, 1939), which can be written as the probability of success given selection—that is,

$$P(y \geq y_0 | x \geq x_0) = \frac{\int_{x_0}^{\infty} \mathrm{F}(x)h(x)dx}{\int_{x_0}^{\infty} h(x)dx} \quad . \tag{A8}$$

Equation A8 offers additional theoretical support for the proposed model, because F(x) is almost identical to $P(x)$.

Rhetorically it may be asked, "If $P(x)$ is almost identical to F(x), why consider $P(x)$?" One major reason for selecting $P(x)$ over F(x) is that F(x) assumes that the predictor and criterion have a bivariate normal distribution. This assumption is not needed for a LR function.

Appendix B: Proof of Equation 5

In the case of fixed-quota selection, optimal selection consists of maximizing Equation 4 with the restriction that

$$Q[P(A_1) + P(A_4)] + Q'[P(A_1') + P(A_4')] = \mathrm{SR} \tag{B1}$$

or

$$Q[P(x \geq x_0)] + Q'[P(x \geq x_0')] = \mathrm{SR} \quad . \tag{B2}$$

In order to find optimal cutting scores x_0 and x_0', Equation 4 must be rewritten as

$$\begin{aligned} \mathrm{E}(U) = {} & Q\Big\{U_1\int_{x_0}^{\infty} P(x)h(x)dx + U_2\int_{-\infty}^{x_0} P(x)h(x)dx + U_3\int_{-\infty}^{x_0}[1 - P(x)]h(x)dx + U_4\int_{x_0}^{\infty}[1 - P(x)]h(x)dx\Big\} \\ & + Q'\Big\{U_1'\int_{x_0'}^{\infty} P'(x)h'(x)dx + U_2'\int_{-\infty}^{x_0'} P'(x)h'(x)dx + U_3'\int_{-\infty}^{x_0'}[1 - P'(x)]h'(x)dx \\ & + U_4'\int_{x_0'}^{\infty}[1 - P'(x)]h'(x)dx\Big\} \quad . \end{aligned} \tag{B3}$$

This equation should be maximized subject to the restriction that

$$Q[P(x \geq x_0)] + Q'[P(x \geq x_0')] = Q\int_{x_0}^{\infty} h(x)dx + Q'\int_{x_0'}^{\infty} h'(x)dx = \mathrm{SR} \quad . \tag{B4}$$

The equation to be maximized can now be written as

$$W = \mathrm{E}(U) - \lambda[\mathrm{SR} - Q\int_{x_0}^{\infty} h(x)dx - Q'\int_{x_0'}^{\infty} h'(x)dx] \quad , \tag{B5}$$

where λ is the Lagrangian multiplier. By differentiating Equation B5 with respect to x_0, the resulting derivative can be set equal to 0 and then divided by Q to produce

$$\partial w/\partial x = -U_1P(x_0)h(x_0) + U_2P(x_0)h(x_0) + U_3[1 - P(x_0)]h(x_0) - U_4[1 - P(x_0)]h(x_0) - \lambda h(x_0) \quad . \tag{B6}$$

A simplification of Equation B6 is

$$P(x_0)(U_2 - U_1 - U_3 + U_4) + (U_3 - U_4) = \lambda \quad . \tag{B7}$$

Similarly, differentiating W with respect to x_0' and setting the resulting equation to 0 produces

$$P'(x_0')(U_2' - U_1' - U_3' + U_4') + (U_3' - U_4') = \lambda \quad . \tag{B8}$$

Eliminating λ from Equations B7 and B8 results in

$$P(x_0)(U_2 - U_1 - U_3 + U_4) + (U_3 - U_4) = P'(x_0')(U_2' - U_1' - U_3' + U_4') + (U_3' - U_4') \quad , \tag{B9}$$

which is identical to Equation 5. The optimal fixed-quota selection procedure is therefore one in which the cutting scores are selected to satisfy Equation B9 subject to the restriction that

$$Q\int_{x_0}^{\infty} h(x)dx + Q'\int_{x_0'}^{\infty} h'(x)dx = \mathrm{SR} \quad . \tag{B10}$$

References

Alf, E. F., Jr., & Dorfman, D. D. (1967). The classification of individuals into two criterion groups on the basis of a discontinuous payoff function. *Psychometrika, 32,* 115–123.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research.* New York: McGraw-Hill.

Callender, J. C., & Osburn, H. G. (1980). Development and test of a new model for validity generalization. *Journal of Applied Psychology, 65,* 543–558.

Gross, A. L., & Su, W. (1975). Defining a fair or unbiased selection model: A question of utilities. *Journal of Applied Psychology, 60,* 345–351.

Gulliksen, H. (1986). The increasing importance of mathematics in psychological research (Part 3). *The Score, 9,* 1–5.

Hogg, R. V., & Craig, A. T. (1970). *Introduction to mathematical statistics.* New York: MacMillan.

Linn, R. L. (1978). Single-group validity, differential validity, and differential prediction. *Journal of Applied Psychology, 63,* 507–512.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Wesley.

McCormick, E. J., & Ilgen, D. (1980). *Industrial psychology* (7th ed.). Englewood Cliffs NJ: Prentice-Hall.

Neter J., & Wasserman, W. (1974). *Applied linear statistical models.* Homewood IL: Irwin.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology, 65,* 373–406.

Petersen, N. S. (1976). An expected utility model for "optimal selection." *Journal of Educational Statistics, 2,* 333–358.

Raju, N. S., & Burke, M. J. (1983). Two new procedures for studying validity generalization. *Journal of Applied Psychology, 68,* 382–395.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement,* No. 17.

SAS. (1985). *User's guide: Statistics.* Cary NC: SAS Institute, Inc.

Schmidt, F. L., Gast-Rosenberg, I. E., & Hunter, J. E. (1980). Test of a new model of validity generalization: Results for computer programmers. *Journal of Applied Psychology , 65,* 643–661.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62,* 529–540.

Smith, P. C. (1976). Behaviors, results, and organizational effectiveness: The problem of criteria. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 745–775). Chicago IL: Rand-McNally.

SPSS-X. (1988). *User's guide.* Chicago IL: SPSS, Inc.

Tatsouka, M. M. (1988). *Multivariate statistics* (2nd ed.). New York: Macmillan.

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection. *Journal of Applied Psychology, 23,* 565–578.

Wilbourn, J. M., Valentine, L. D., & Ree, M. J. (1984). *Relationship of the armed services vocational aptitude battery (ASVAB) forms 8, 9, and 10 to the Air Force technical final school grades* (AFHRL-TP-84-8, AD-A144 213). Brooks Air Force Base TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Acknowledgments

The authors thank Malcolm Ree and Major Welsh from the Human Resources Laboratory at Brooks Air Force Base, TX, for providing the data used in the present study. The authors also thank the Editor and two anonymous reviewers for their many valuable comments. The opinions expressed in this article are those of the authors and do not necessarily reflect the views of the Navy Department.

Author's Address

Send requests for reprints or further information to Nambury S. Raju, Department of Psychology, Illinois Institute of Technology, Chicago IL 60616, U.S.A.