



1988

Selecting the Right Employee: Examining the General Validity of Employee Testing can Lead to Solid Recruitment

Dow Scott

Loyola University Chicago, dscott@luc.edu

Robert M. Madigan

Diana L. Deadrick

Follow this and additional works at: https://ecommons.luc.edu/business_facpubs



Part of the [Business Commons](#)

Recommended Citation

Scott, Dow; Madigan, Robert M.; and Deadrick, Diana L.. Selecting the Right Employee: Examining the General Validity of Employee Testing can Lead to Solid Recruitment. *Personnel Administrator*, 33, 12: 86-91, 1988. Retrieved from Loyola eCommons, School of Business: Faculty Publications and Other Works,

This Article is brought to you for free and open access by the Faculty Publications and Other Works by Department at Loyola eCommons. It has been accepted for inclusion in School of Business: Faculty Publications and Other Works by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](#).

© 2016 SHRM. All rights reserved. *Personnel Administrator* is the property of Society for Human Resource Management (SHRM) and its content may not be copied or e-mailed to multiple sites or posted to a listserv without SHRM's express written permission. However, users may print, download, or e-mail articles for individual research.

Selecting the Right Employee

*Examining the general validity of employee testing
can lead to solid recruitment*

By K. Dow Scott, Robert M. Madigan and Diana L. Deadrick

During the last 10 years a major controversy has developed concerning the total applicability of employment tests to different jobs, organizations and employee groups. The traditional approach, embodied in the *Uniform Guidelines for Employee Selection Procedures*, holds that employment tests must be validated for each applicant population and situation in which they are used.

The basic premise of this view is that tests are "situationally specific," that is, differences in the work location, job or applicant population might alter the validity of a particular application of an employment test, even if the test was previously found valid for similar jobs and circumstances. A large body of inconsistent findings from numerous validation studies seemed to support this conclusion.

However, an alternative view of the generalizability of tests was given by John E. Hunter (Michigan State University) and Frank L. Schmidt (then of the U.S. Office of Personnel Management). They reasoned that the inconsistent validation findings were primarily attributable to statistical artifacts, rather than situational differences.

Using a meta-analytic research method that adjusts for such methodological problems, Schmidt, Hunter and associates re-analyzed the findings of previous validation studies. They concluded that basic ability tests are generally more valid

than other selection procedures, and test validity is relatively stable across jobs, organizations and applicant sub-groups.¹

Based on this evidence, the U.S. Employment Service (USES) developed the Validity Generalization (VG) program, an applicant screening and referral process that relies heavily on test scores, rather than interview judgments, as the primary tool for matching applicants with job openings. The General Aptitude Test Battery (GATB) is used to assess the applicant on nine basic abilities. GATB-based composite scores are then used to predict the applicant's future job performance in virtually all jobs. This transportability (generalizability) of the validity of test scores across different jobs and different companies enables employers to use one test battery to screen all job applicants.

This article describes some of the findings from a field evaluation of the VG program.^{2,3} The program was examined from the perspective of both applicants and employers. Since applicants are not required to take the GATB, the success of the VG program depends to a con-

siderable degree on whether it is accepted by the applicant population. Therefore, a survey of the perceptions and attitudes of job applicants toward the VG program was conducted. Employers also were surveyed regarding their reactions to the VG program, as well as their perceptions of the services provided by the Job Service.

In addition, a predictive validity study was conducted in which the relationship between the VG program test score and measures of employee performance in one organization were examined. Only the findings from the applicant survey and validation study are summarized here; limitations in the employer survey data precluded meaningful generalizations.

Job applicant reactions

The extensive attention given to testing issues and abuses over the past 20 years has contributed to a general distrust of employment testing, particularly among minority groups. Thus, the focus of this facet of the study was on applicants' perceptions of the new testing procedure and whether these perceptions and beliefs varied by demographic sub-group. We were particularly interested in determining whether reactions of applicants to the VG program varied by race or gender.

Applicant perceptions were assessed at two points in time. The first survey was administered at the conclusion of the GATB testing sessions conducted by Job Service personnel

K. Dow Scott is associate professor of management at the Department of Management, Virginia Polytechnic Institute and State University in Blacksburg, Va.

Robert M. Madigan is associate vice president for personnel, Virginia Polytechnic Institute and State University.

Diana L. Deadrick is assistant professor of management at Owen Graduate School of Management at Vanderbilt University in Nashville, Tenn.

FIGURE 1
APPLICANT SURVEY RESPONSES

	MEAN	STRONGLY DISAGREE	DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	AGREE	STRONGLY AGREE
INITIAL SURVEY (n = 1,064)							
I was nervous about taking the test	3.2	16.5%	24.5%	10.3%	25.0%	18.7%	5.0%
I was very nervous during the test	3.1	13.6%	28.3%	11.6%	28.5%	12.6%	5.4%
The GATB was a fair test of my job abilities	3.8	9.8%	13.3%	15.8%	16.5%	36.2%	8.4%
A test is a fair way to make job referral decisions	3.9	7.9%	12.6%	16.9%	18.1%	31.9%	12.6%
The people at the JS were helpful & courteous	5.1	2.3%	2.5%	1.4%	6.0%	50.6%	37.2%
The people at the JS were respectful	5.1	2.0%	2.6%	1.7%	7.6%	52.6%	33.5%
FOLLOW-UP SURVEY (n=453)							
The GATB was a fair test of my job abilities	3.6	13.2%	15.5%	11.7%	19.6%	33.3%	6.7%
A test is a fair way to make job referral decisions	3.8	8.2%	12.0%	14.0%	26.4%	32.6%	6.8%
The people at the JS were helpful & courteous	4.8	3.8%	3.8%	3.3%	11.4%	52.8%	24.9%

*LEGEND: 1-strongly disagree; 2-disagree; 3-slightly disagree; 4-slightly agree; 5-agree; 6-strongly agree

over a four-month period. The work history and registration data of these respondents were then retrieved to generate a useable sample of 1,064 applicants (Figure 1).

A second survey was mailed to the respondents approximately 12 weeks after taking the test. The 453 applicants who completed both questionnaires, and for whom registration and work history data were available, constitute Figure 2.

The demographic composition of both samples was similar to the local area labor force in most respects, with the predictable exceptions of age and employment status. The majority of respondents in both samples were under 45

years of age (87 percent to 89 percent), white (86 percent to 89 percent) and male (54 percent to 57 percent). Blacks were the largest minority group (9 percent to 13 percent), and Hispanics, American Indians and Asians collectively comprised less than 1 percent of the samples. Approximately 40 percent of both samples were employed, over 70 percent had completed high school or GED, and 12 percent held at least a bachelor's degree.

The survey items elicited responses regarding test anxiety, test fairness and treatment by the staff of the local Job Service office. All of the items were answered on a six-point rating scale with responses ranging from "strongly disagree" to "strongly agree." The item and response scales are shown in Figure 1.

Figure 1 provides the average (mean) score on each question as well as the percentage distribution of responses for both samples. The initial responses of job applicants (top of Figure 1) indicate that reactions toward the test itself were not as positive as toward other aspects of the process. For example, approximately 61 percent agreed that the GATB was a fair test of their job abilities and that it was a fair way to make job referral decisions. But over 80 percent felt that Job Service personnel were helpful and courteous. The follow-up survey revealed a similar pattern of responses.

Some differences between sub-

groups were observed. Figure 2 reveals that a higher proportion of women felt that the GATB was a fair test of their job abilities despite the fact that test anxiety was more likely to be a problem for women than men. Test anxiety was also more likely among minority applicants, yet no differences were evident between minority and non-minority applicants regarding perceptions of the fairness of the testing process.

In addition, educational level apparently affects both anxiety and fairness perceptions. The better educated applicants were less nervous about taking the test. But, they were also less likely to agree that tests were a fair basis for determining job abilities or making job referrals.

From a practical standpoint, these findings suggest a potential problem for employers and the Job Service.

Approximately 40 percent of the applicants were somewhat skeptical about the use of employment tests.

Although the majority of respondents agreed that the GATB was fair, approximately 40 percent of the applicants were somewhat skeptical about the use of employment tests.

In addition, almost half of the respondents indicated they experienced some degree of test anxiety, especially women, blacks and the less educated. This suggests that employers who incorporate the VG

program into their selection process might reduce their applicant flow. Furthermore, because testing is less acceptable to the better-educated applicants, the VG program could reduce the number of such people that the job service can attract for referral.

Validation study

The question most frequently posed by employers contacted was whether the VG program will actually help them select the best-qualified applicants for their jobs. Although many of the employers were aware of the Job Service's claims about validity generalization, they still harbored some reservations. Would the VG program really work in their company, and would it really apply to their specific job

FIGURE 2
SURVEY RESPONSES BY GENDER AND RACE
(Mean Responses)

GENDER	SURVEY 1		SURVEY 2	
	MALE (N = 604)	FEMALE (N = 458)	MALE (N = 246)	FEMALE (N = 207)
I was nervous about taking the test	3.0	3.4	—	—
I was very nervous during the test	3.0	3.3	—	—
The GATB was a fair test of my job abilities	3.7	4.0	3.6	3.7
A test is a fair way to make job referral decisions	3.9	4.0	3.9	3.8
The people at the JS were helpful & courteous	5.1	5.2	4.8	4.8
The people at the JS were respectful	5.1	5.1	4.9	4.9
RACE	WHITE (N = 917)	BLACK (N = 145)	WHITE (N = 403)	BLACK (N = 50)
I was nervous about taking the test	3.2	3.5	—	—
I was very nervous during the test	3.1	3.3	—	—
The GATB was a fair test of my job abilities	3.8	4.0	3.6	4.2
A test is a fair way to make job referral decisions	3.9	3.9	3.8	4.1
The people at the JS were helpful & courteous	5.1	5.1	4.8	4.9
The people at the JS were respectful	5.1	5.1	4.8	5.1

openings? Hence a case study of the VG testing program in one company was conducted to evaluate both the validity and the potential economic benefit to the employer.

The research strategy used in this evaluation was criterion-related validation using a predictive design. Applicants took the GATB before they were hired, and the relationship between their test scores and subsequent objective and subjective measures of job performance was determined.

The study site was a large (8,000 employees) multi-plant manufacturer of casual-wear garments. In the hope of increasing productivity through selection of more qualified personnel, the personnel department adopted a policy of referring all applicants for production positions to the Job Service for testing. However, because the testing program was not generally understood or accepted in the plants, the VG scores were not used in the referral and selection decisions. Thus, the test scores of the new hires approximate a normal distribution. This fact, coupled with the size of the sample and an apparent lack of selectivity in the selection procedure, served to allay any serious concern about sampling error.

The job of sewing machine operator (SMO) was selected for this study due to the large number of such operators and the availability of both subjective and objective performance data. SMO's are organized into production lines, with each line responsible for the completion of a specific garment.

For example, a shirt line might consist of six sewing operations (i.e., seam sleeves, attach collar, etc.) to produce the finished garment. However, because of in-process inventory, each operator's performance was independent of the work pace of the other operators.

The job performance of the

FIGURE 3
SAMPLE
CHARACTERISTICS

Number Hired:	932
Number Tested:	751
Average Test Score:	51.3
Test Score Range:	1-99
Standard Deviation:	25.8
Race:	
White	551
Black	198
Age:	
Average	26
Range	1-59
Experience:	
Average (months)	11
No prior experience (No.)	492
Less than 1 Year (No.)	113
Turnover:	
No.	383

SMO's was measured using both supervisor ratings and actual production data. Supervisors rated each of their operators on six dimensions of performance: quality of output, quantity of output, flexibility, dependability, receptiveness to training and an overall rating. These ratings were obtained in group meetings with supervisors conducted by the researchers. Confidentiality of the ratings was assured; only the researcher had access to the data. Piece-rate (production) earnings of SMO's provided the objective measures of performance.

Unit rates were based on standard industrial engineering procedures so that two employees working at the same rate of output, but on different operations, would receive the same pay. The earnings averages used in this study were "clean" pro-

ductivity measures, that is, the various allowances and guarantees typical of piece-rate systems were not reflected in these measures.

The test score used in this study was the Job Family score for sewing machine operator jobs. The nine basic ability scores of the GATB are used to form five job family composites, each reflecting a different weighted combination of basic abilities.

Weights are determined by the nature of the mental and physical requirement of jobs in that job family. Sewing machine operators fall into Job Family 5 (semi-skilled jobs), which reflects test performance on psychomotor (weighted 56 percent) and cognitive (weighted 44 percent) ability. The final step in computing the Job Family score is to convert the composite into a percentile (norm-referenced) score using separate norms for minority and non-minority candidates (within-group percentiles).

As shown in Figure 3, the total sample consisted of 932 newly hired SMO's, but only 751 were tested (primarily due to time pressure to fill openings). Of these, 26 percent were minority (black), all were women, 64 percent had no previous experience as SMO's, and the average age was 26 years old.

The turnover rate of these new employees (51 percent) was high, and the average tenure of those who left was brief (12 weeks). The average VG percentile score (Job Family 5) was 51.3 with a standard deviation of approximately 26 and a range from 1 to 99. This distribution of test scores supports the previous assertion that tests were not actually used in the employee selection process.

Figure 4 presents job performance averages broken down by

FIGURE 4
QUARTILE ANALYSIS SAMPLE CHARACTERISTICS & PERFORMANCE AVERAGES

	QUARTILE 1	QUARTILE 2	QUARTILE 3	QUARTILE 4
Sample Characteristics				
Number of Employees Tested:	137	230	225	159
Average Test Score:	13.6	38.1	62.9	86.5
Minority Percentage:	17%	16%	29%	46%
Average Age:	29	26	26	25
Experience: None	65%	61%	66%	71%
Turnover Percentage	53%	40%	46%	37%
Performance Measures (Sample)				
Month 1 Average Earnings	\$2.06 (n=107)	\$2.28 (n=203)	\$2.39 (n=190)	\$2.57 (n=141)
Month 2 Average Earnings	2.84 (94)	3.11 (187)	3.17 (174)	3.36 (128)
Month 3 Average Earnings	3.21 (78)	3.59 (158)	3.68 (149)	3.83 (116)
Month 4 Average Earnings	3.54 (63)	3.84 (133)	4.03 (128)	4.11 (105)
Month 5 Average Earnings	3.60 (55)	4.06 (119)	4.27 (116)	4.31 (88)
Month 6 Average Earnings	3.83 (47)	4.15 (101)	4.48 (97)	4.59 (75)
20-Week Average	3.16 (55)	3.54 (119)	3.69 (116)	3.80 (88)
Supervisor Ratings				
Quantity	2.0	2.3	2.3	2.7
Quality	2.7	3.0	2.9	3.2
Flexibility	3.1	3.4	3.2	3.6
Receptiveness	3.4	3.6	3.3	3.7
Dependability	3.6	3.8	3.7	3.9
Overall	3.0	3.1	3.1	3.5
Number Rated	(82)	(160)	(154)	(116)

test score quartile. Quartile 1 refers to those applicants whose Job Family 5 score was between 1 percent and 25 percent. Quartile 4 refers to the highest scores, those between 76 percent and 99 percent. The sample characteristics of each quartile describe the demographic composition of each group. These data reveal that Quartile 4 (highest scorers) had a higher proportion of black and inexperienced employees, was slightly younger and had a lower turnover rate than the other quartiles.

A comparison of performance measures by quartile supports the conclusion that Job Family 5 scores are positively related to performance on the job. First, the monthly production averages consistently

increase with the test score quartiles, not only during the early weeks of training, but through the first six months on the job. In statistical terms, the correlation coefficients between test scores and average earnings ranged from .15 to .24, with the highest values observed for months 5 and 6.

This finding suggests that the benefit of testing persists over time. Second, a similar pattern is reflected in the supervisors' ratings. These ratings reflect not only the quantity of production but other facets of sewing machine operator performance as well. Finally, the turnover data also indicate a favorable rela-

tionship with VG scores, albeit weak. This finding contradicted the prevailing belief in the company that the high GATB scorers were more likely to become bored or dissatisfied and thus have higher turnover.

The final aspect of the employer analysis dealt with the question of the potential economic benefit — or utility — to the company from the use of the VG program. Techniques for utility analysis have been available for years and incorporate calculations of both the costs and benefits of using a valid selection procedure. The benefit of improved selection decisions can take numerous forms, such as improvement in worker productivity, reduc-

tions in training time and cost, and administrative cost savings. The estimates presented here address only the incremental improvement in sewing machine operator productivity that could have been realized by the company had it used the Job Family 5 scores as a selection tool.

The potential gain in productivity from implementing a new (valid) selection procedure depends upon a number of factors, such as:

- Validity of the existing and new procedure.
- Number of people hired.
- Extent of variability in employee performance.
- Number of applicants.
- Cost of the new procedure.

In this particular case all conditions were favorable.

The VG program entails no cost to the employer. There were typically about three applicants for each position, allowing the company to capitalize on the validity of the selection procedure by skimming the top portion of the applicant pool. Productivity differences between high and low performers were large, thus the potential payoff from accurate differentiation among applicants was substantial. Finally, a large number of sewing machine operators were hired annually, thus improving the probability that the benefits of testing would be realized and magnifying the effect.

Application of a general utility model to the data gathered indicated that the average annual gain in productivity as reflected in production earnings would be \$519 per hiree if the Job Family 5 scores were used for selection rather than the current, unsystematic selection procedure. If the company hired 800 operators and average tenure of these hires was only six months, the benefit of that one year of testing would be approximately \$200,000.

The payoff increases, of course, if average tenure is longer and if the cumulative effects are factored into the estimate.

Our point in presenting these estimates here is not to suggest that they are precise (although we believe they are quite conservative). Rather, it is to point out that the

The cost/benefit ratio would undoubtedly be very favorable if the company properly used the VG program.

cost/benefit ratio in this instance would undoubtedly be very favorable if the company properly used the VG program.

Conclusions

The Validity Generalization program is one of the most important, and perhaps riskiest, initiatives in the history of the U.S. Job Service. It represents a marked departure from the past philosophy and practices of the agency, and it contradicts the prevailing governmental attitudes toward employment testing, at least at the federal level. Whether the program will ultimately be successful will depend upon the experiences and perceptions of both employers and job seekers. The findings presented here suggest the outlook for the program is favorable.

A single study is hardly conclusive evidence of the validity of a testing program that could be used

across the occupational spectrum. However, the fact that the VG scores proved valid for a randomly chosen job is impressive. There is no reason to believe that this study is an anomaly. On the contrary, for reasons that are beyond the scope of this discussion, we would expect a higher level of validity for most other jobs. This is not to suggest that additional studies are not necessary. A large body of evidence such as that reported here will be required to establish the credibility of the program, because neither applicants nor employers are likely to accept the existing analyses of past studies as adequate proof of the validity of VG scores. ■

Notes:

1. See F. Schmidt and J. Hunter, "Old Theories and New Research Findings," *American Psychologist* 36, 1981, pp. 1128-1137; and F. Schmidt, K. Pearlman, J. Hunter and H. Hirsh, "Forty Questions about Validity Generalization and Meta-Analysis," *Personnel Psychology* 38, (1985), 697-798.
2. K.D. Scott, J. Stoddard, R. Madigan and D. Deadrick, *An Examination of Job Applicant Attitudes Toward the VG Screening Procedure*, 1987, produced for the Virginia Employment Commission, Richmond, Va. K.D. Scott, R. Madigan and D. Deadrick, *An Examination of Employer Attitudes Toward the VG Screening Procedure*, 1987, produced for the Virginia Employment Commission, Richmond, Va.
3. R. Madigan, D. Deadrick and K.D. Scott, *A Validation Study of the VG Testing Procedure for Selection of Sewing Machine Operators*, 1987, produced for the Virginia Employment Commission, Richmond, Va.
3. This study was funded by the Virginia Employment Commission, Richmond, Va. Special thanks to Ben Berryman, John Kuhn, Bob BeCraft and Jim Roberts of the Virginia Employment Commission for their support and interest.

© 2016 SHRM. All rights reserved. ***Personnel Administrator*** is the property of Society for Human Resource Management (SHRM) and its content may not be copied or emailed to multiple sites or posted to a listserv without SHRM's express written permission. However, users may print, download, or email articles for individual research.