

Adverse Impact Resulting From the Use of Clerical Tests: A Meta-Analysis

William J. Bradburn
Fu Associates, Ltd.

Paul Villar

The prevalence of adverse impact in clerical tests was investigated. Six tests which measured clerical aptitude were included in the study. Inclusion in the study was based on test manuals including the necessary descriptive statistics. A meta-analysis was done in order to yield an overall effect size (d) for race and gender. Our analysis revealed substantial differences in test scores between races ($d = .60$), but not between genders ($d = .02$). Given that adverse impact is to be found in selection tests, it is in an organization's best interest to carefully review these tests in order to choose the best selection device for a given situation

In recent years there have been many tests developed to aid in the process of employee selection. As use of these tests increased in popularity, attention was given to the legal implications of their use. No longer was it enough to prove that a test was valid, but there was a need to quantify the amount of adverse impact associated with these devices to both allow for fairer selection practices and to estimate potential legal risks. Research over the years has since been compiled that has examined the validity of tests (e.g. Barrett, 1990; Hunter, Schmidt, & Jackson, 1984; Kerr, 1990), attitudes of those taking various tests (Robertson & Kandola, 1982; Schmidt, Greenthal, Hunter, Berner, & Seaton, 1977), as well as on legal aspects of using tests (Barrett, 1990; Sackett, Burris, & Callahan, 1989). However, there has been little research done in a broad sense to examine multiple tests and their adverse impact based on race and gender.

Because there are many ready-made tests for purchase, especially in the clerical field, it was the aim of this study to measure the adverse impact of those tests. This was achieved using meta-analysis techniques whereby data on tests that are available for use in the selection of clerical employees were cumulated in order to derive an overall effect size (d).

METHOD

Tests used in this study were obtained by consulting the Radford Index (Busacca, Gregory, & Miller, 1991). This index contains a listing of all types of tests used in industry. For purposes of this study, selection tests for clerical employment were examined. Publishers were contacted to obtain testing manuals containing appropriate descriptive statistics. For inclusion in the meta-analysis, the manuals needed to contain test means and standard deviations for each gender and/or race. Through this process, six tests qualified for inclusion in the study. There were the PSI Basic Skills Test (PSI), Office Skills Tests (OST), Employee Aptitude Survey (EAS), General Aptitude Test Battery (GATB), Account Clerk I (ACI), and Entry Clerical (EC). Only the portions of the GATB that dealt with clerical categories were used. Also, the OST was broken down into two forms; Form A and Form B.

Once identified, the test means and standard deviations were converted into effect sizes using Schmidt and Hunter's (1982) technique. Using this meta-analysis technique, these individual effect sizes were used to calculate mean effect sizes (d's) for each test, as well as for each subgroup of tests. The overall effect size for race was based on data from 111,420 subjects and the overall effect size for gender was based on data from 41,910 subjects.

RESULTS AND DISCUSSION

As shown in Tables 1 and 2, a meta-analysis revealed a medium effect size for race ($d = .60$) but not for gender ($d = .02$). For this paper, positive effect sizes indicate that the majority group (e.g. Whites, males) scored higher on the test while negative effect sizes indicate that the minority group (e.g. Blacks, females) scored higher. The effect sizes involving race leads to the conclusion that Blacks, but not women, will be adversely impacted if clerical tests are used. From the data obtained, a few questions arise.

First, though these tests are valid and are purportedly measuring the same content domain, how can their degree of adverse impact vary so

Table 1

Effect Sizes for Racial Differences in Test Scores

Clerical Test Type	Test					Total
	PSI	OST-A	OST-B	GATBACI	EC	
Numerical Skills						.42
Numerical Ability						
Numerical Reasoning						
Computation	.74					
Numerical Aptitude				.40		
Numerical Skills			.26	.51		
Grammar						.59
Grammar		.46	1.2			
Punctuation		.43	.57			
Spelling	.40	.72				
Vocabulary	.60	.74	.70			
Reading Comprehension	.48	.58	.55			
Verbal Skills						.13
Verbal Comprehension						
Verbal Reasoning						
Word Fluency						
Language Skills	.45					
Verbal Aptitude				.10		

Clerical Test Type	PSI	OST-A	OST-B	GATB ACI	EC	Total
Following Directions						.33
Oral Directions	.30	.43	.43			
Written Directions	.56					
Typing						.41
Speed		.36	.47			
Accuracy		.43	.38			
Clerical Skills						.42
Visual Speed and Accuracy	.14					
Classifying	.39					
Clerical Perception				.44		
Checking	.45	.66	.16			
Coding	.41	.43	.43			
Filing		.44	.64			
Names	.32					
Numbers	.48					
Forms		.42	.60			
Reasoning	.36					
Problem Solving	.75					
Decision Making	.58					
Memory	.68					
Generic Battery	.68					
Overall	.39	.46	.58	.30	1.0	.93
						.60

Table 2**Effect Sizes for Gender Differences in Test Scores**

	Test				Total
	PSI	EAS	ACI	BC	
Numerical Skills					.36
Numerical Ability		.58			
Numerical Reasoning		.19			
Computation	.28				
Grammar					-.09
Vocabulary	-.05				
Reading Comprehension	-.12				
Verbal Skills					-.33
Verbal Comprehension		-.08			
Verbal Reasoning		.04			
Word Fluency		.00			
Language Skills	-.86				
Following Directions					-.24
Oral Directions	-.26				
Written Directions	-.22				
Clerical Skills					.42
Visual Speed and Accuracy	-.29	-.08			
Checking	-.20				
Coding	-.38				
Filing					-.36
Names	-.85				
Numbers	-.28				

	Test				
	PSI	EAS	ACI	EC	Total
Reasoning	-.06	.24			
Problem Solving	.34				
Decision Making	.00				
Spatial Visualization		.57			
Memory	-.14				
Generic Battery	-.31				
Overall	-.25	.19	.05	-.08	-.02

greatly? Also, the subtests contained within these tests differ in the degree to which adverse impact is found. One test in particular, the OST, had large differences between two forms of the same test.

For professionals concerned with the fair selection of employees, this information is important for making selection decisions. With the great number of selection tests available on the market, would it not be nice to know, in addition to validity, the adverse impact associated with these tests before purchasing them for organizational use?

In the present study there were a few limitations that may have affected the results. One major problem involved obtaining test manuals for inclusion in our research. Once obtained, several manuals were found to lack the descriptive statistics necessary to be included in the meta-analysis and thus were excluded. Also, not all tests used in the study contained the same subtests, with some reporting no subtest information at all. During our study, we grouped subtests that were similar in nature under similar

categorical names. Also, when viewing results found in the adverse impact of gender, it should be noted that the sample size for males that the meta-analysis was based on was substantially smaller than the sample size for females.

REFERENCES

- Barrett, G. V. (1990). Personnel selection after Watson, Hopkins, Antonio, and Martin. *Forensic Reports*, 3, 179-203.
- Busacca, T., Gregory, R., & Miller, J. (1991). *The Radford Index*. Unpublished manuscript, Radford University, Radford, Virginia.
- Cohen, J. (1977). *Statistical Power Analysis for Behavioral Sciences*. New York: Academic Press.
- Cooperative Personnel Services (1990). *Entry Clerk*. Sacramento, CA: author.
- Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.
- Kerr, M. R. (1990). The impact of Canadian employment selection issues on validation research. *Canadian Psychology*, 31, 334-344.
- Robertson, I.T., & Kandola, R. S. (1982). Work sample tests: Validity, adverse impact, and applicant reaction. *Journal of Occupational Psychology*, 55, 171-183.
- Ruch, F. L., & Ruch, W. W. (1980). *Employee Aptitude Survey*. Glendale, CA: Psychological Services, Inc.
- Ruch, W. W., Weiner, J. A., McKillip, R. H., & Dye, D. A. (1985). *PSI Basic Skills Tests*. Glendale, CA: Psychological Services, Inc.
- Sackett, P. R., Burriss, L. R., & Callahan, C. (1989). Integrity testing for personnel selection: An update. *Personnel Psychology*, 42, 491-529.
- Schmidt, F. L., Greenthal, A. L., Hunter, J. E., Berner, J. G., & Seaton, F. W. (1977). Job sample vs. paper-and-pencil trades and technical tests: Adverse impact and examinee attitudes. *Personnel Psychology*, 30, 187-197.
- Science Research Associates (1984). *Office Skills Test*. Park Ridge, IL: MacMillan/McGraw-Hill.
- Synk, D. J., & Swarthout, D. (1987). *Comparison of black and nonminority validities for the General Aptitude Test Battery* (Report No. 51). Washington, D.C.: U.S. Department of Labor.

