Six Alternative Methods of Use of Employee Selection Test Data

Joel P. Wiesen, Ph.D. and Aaron A. Brown*
Applied Personnel Research
27 Judith Road
Newton, MA 02459
(617) 244-8859
jwiesen@appliedpersonnelresearch.com
aab405@nyu.edu

* Attending New York University as of Fall, 2008

Authors' note:
This paper and some related material are available on the web at http://apr-boston.com/mc

Six Alternative Methods of Use of Employee Selection Test Data

by Joel P. Wiesen, Ph.D.
and Aaron A. Brown

Abstract

Four innovative, alternative methods of using employee selection test data were evaluated in a Monte Carlo study that simulated the selection of Firefighters. Validity and adverse impact for women and minorities were evaluated at 3 selection ratios and for 2 proportions of minority applicants. (Women always comprised 50% of the applicant group.) The Greatest Strength Method was found to have the best combination of minimizing adverse impact against both women and minorities while maintaining validity equal to that typically seen with written tests of cognitive ability.

Problem

The pervasive adverse impact[1] of employee selection tests sorely challenges our profession. It exposes our clients to employment discrimination litigation, appears unfair to the public and the news media, works against diversity in the workforce, and results in pressure to lower standards and pressure to abandon testing. One simple and reasonably effective approach to this intractable problem, within-group norming, is proscribed by the federal *Civil Rights Act of 1991*. The federal *Uniform Guidelines on Employee Selection Procedures* (Equal Employment Opportunity Commission, the Department of Labor, the Department of Justice, and the Civil Service Commission, 1978) call for consideration of alternative selection procedures and alternative methods of using selection procedures:

> "... the user should include, as a part of the validity study, an investigation of suitable alternative selection procedures and suitable alternative methods of using the selection procedure which have as little adverse impact as possible..." (Section 3.3 B).

Our profession has produced several possible alternative selection procedures (most notably, work-style tests) but only two main alternative methods of using selection test scores: banding (including variants) and pass-fail (Cascio, Outtz, Zedeck, & Goldstein, 1991). A small body of literature has

---

[1] Adverse impact is defined in terms of the ratio of two employee selection (or hiring) rates, such as the selection rate for minority applicants divided by the selection rate for non-minority applicants. Adverse impact ratios less than .8 are especially problematic, particularly as viewed by the federal agencies that enforce Title VII of the Civil Rights Act of 1964, as amended.

suggested that innovative methods of combining test scores may decrease adverse impact while maintaining validity (Wiesen, 2006a, 2006b, 2007). Sackett and Roth (1996) found different levels of job performance and minority representation with different composite and pass-fail combinations of two tests. Pulakos and Schmitt (1996) found including non-verbal tests increased validity and decreased subgroup differences compared to using traditional paper-and-pencil tests. The present study further examines the one innovative method for using selection test data with some research support (the greatest strength method) and proposes and evaluates the adverse impact and validity of three other innovative alternative methods that use tests of both cognitive and non-cognitive abilities.

One common[2] concern (or complaint) of unions and others when using either a banding or pass-fail approach is that the hiring standards are being watered down. Countering this concern has sometimes been difficult. The written M/C cognitive ability test is assumed to be the gold standard, because of its historic near universal use for most jobs. Here we take the written M/C cognitive ability test as the standard for comparing the adverse impact and validity of other innovative, alternative methods of using test scores.

Subjects
The data generation specifications employed in this Monte Carlo study are described in the procedure section.

Procedure
A Monte Carlo simulation was employed consisting of seven steps, listed next and then described in more detail:

Step 1. Multi-normal data sets were generated to mimic typical applicant scores on an entry-level firefighter selection test battery that might be used in fire departments of various sizes, as well as subsequent employee job performance data, with intercorrelations chosen based on the professional literature.

Step 2. Mean score differences between male and female applicants and between minority and non-minority applicants were created on certain tests. Mean score differences between minority and non-minority applicants were created on the criterion.

Step 3. Any applicant with a low score on any test was failed for three of the innovative approaches only, as described below.

Step 4. Each applicant's grade was determined for each of the following methods of test data use: (1) written M/C cognitive ability test alone, (2) equally weighted composite of all tests, (3) greatest

---

[2]In my (JPW) experience.

strength method (in which an applicant's grade is determined by his or her highest score), (4) two greatest strengths method, (5) drop the lowest score method, (6) equally weighted composite of all tests other than the M/C test of cognitive ability, and (7) random selection. (Methods 1 and 2 are considered traditional, methods 3-6 are innovative, and method 7 is included for comparison purposes.)

Step 5. Three selection ratios were employed. Hiring was done based on the grade created for each method, with the applicants with the highest grades hired first.

Step 6. Levels of adverse impact in hiring and the (uncorrected) criterion-related validity were calculated for each of the methods of test score use.

Step 7. The steps above were repeated a chosen number of times, and then the observed adverse impact and validity were summarized.

Step 1 Detail

An existing program that generates multi-variate normal distributions (Multivar, Aguinis, 1994)[3] was used to generate random multivariate normal data sets of specified numbers of applicants, to model the use of four selection tests:
- Written M/C test of cognitive ability (M/C Cognitive)
- Structured oral interview (Oral)
- Physical performance test (PPT)
- Test of conscientiousness (Conscientiousness)

All variables were created to have a mean of 0 and a variance of 1. The intercorrelations specified are shown in Table 1. Applicant groups of three sizes were created: 1,000, 10,000, and 1,000,000.

Step 2 Detail

The mean score differences by group were created by subtracting a constant from scores on specified tests for each member of certain groups. The values of the constants used for these groups are shown in Table 2. The applicant groups were created to have an equal number of men and women. The applicant groups were created to have either 10% or 50% minority applicants.

Step 3 Detail

Any applicant that scored lower than the 10th percentile on any test received a failing grade for any of these innovative grading methods:
- Greatest Strength Method
- Two Greatest Strengths Method
- Drop the Lowest Score Method

---

[3] This program is available on the web at http://carbon.cudenver.edu/~haguinis/mmr/index.html

Step 4 Detail
    The applicants' grades under each approach were calculated as follows.
    (1) Written M/C cognitive ability test alone: An applicant's grade was set equal to his or her score
         on the written M/C cognitive test. The other test scores did not enter into the grading process.
    (2) Equally weighted composite of all tests: This is simply the average of the z scores for each of the
         tests.
    (3) Greatest Strength Method: An applicant's highest score on the 4 tests was determined and that
         score became the grade for the applicant.
    (4) Two Greatest Strengths Method: An applicant's two highest scores on the 4 tests were
         determined and an unweighted average of these two scores was calculated.
    (5) Drop the Lowest Score Method: An applicant's lowest score on the 4 tests was determined
         and that score was dropped. The unweighted average of the other 3 test scores was
         calculated.
    (6) Equally weighted composite of all tests other than the M/C test of cognitive ability: This is the
         average of the z scores for each of the tests other than the written test of cognitive ability.
    (7) Random Selection: Selection was done randomly, independent of the test scores.

Step 5 Detail
    Three selection ratios were employed: 1%, 20% or "all passers." All passers includes all applicants
scoring in the top 90% for any given method omitting, for three methods, those who failed (as described
in Step 3). Other than the random method, when selecting 1% or 20%, the applicants with the highest
grades were selected first.

Step 6 Detail
    Based on the selection of candidates, the adverse impact in hiring for each of the methods of test
data use was calculated. The uncorrected criterion-related validity for each of the methods of test data
use was calculated for all test takers. (Thus the validity correlations do not suffer from range
restriction.)

Step 7 detail
    The steps above were repeated 750 times for the applicant groups of size 1,000, 75 times for the
applicant groups of size 10,000, and once for the applicant groups of size 1,000,000.

Checks on the Monte Carlo Procedure
    We undertook checks on several aspects of the Monte Carlo procedure prior to running the study.

    We evaluated the data the program Multivar generates, looking for degeneracy in random number
generation over cases. We used Multivar to generate many thousands of multivariate normal samples
of size 1,000 and checked some samples at the beginning, middle and end. Scatter plots and calculated
correlations for the samples we checked were seemingly quite similar. We found no apparent change in

the effectiveness of the program.

SPSS programs were written to implement the 7 steps. These programs were verified by manually duplicating the steps in a spreadsheet program (Excel). Each operation of the SPSS program was verified in this fashion.[4]

Table 1. Specified Intercorrelations for Firefighter Selection Tests and Criterion						
	Oral	PPT	Conscientiousness	Gender	EEO Group	Job Performance
M/C Cognitive	.31*	0	.03*	0	0	.51*
Oral		0	0	0	0	.48*
PPT			0	0	0	.35**
Conscientiousness				0	0	.22*
Gender					0	0
EEO Group						0

* Value was taken from Potosky, Bobko, and Roth (2005), Table 1, p. 308.
** Estimate made for the sake of this study.

Table 2. Constants Used to Create Specified Mean Group Differences, in z-Scores					
	M/C Cognitive	Oral	PPT	Conscientiousness	Job Performance
Women	0	0	-1.25*	0	0
Minorities	-.72**	-.31**	0	.07***	-0.27****

* Estimate made for the sake of this study.
** Potosky, Bobko, Roth, 2005, Table 1, p. 308)
*** Foldes, In Press, Table 2, p. 50
**** Roth, Huffcutt, and Bobko, 2003, Table 2, p. 699; and McKay and McDaniel, 2006, Table 2, p. 544

Results

The results are presented in charts and figures as follows: (1) summaries of adverse impact and validity for 1,000,000 applicant group (with proportion minority = .5) are presented in Tables 3 and 4, (2) summaries of the means and standard deviations of the observed adverse impact and validity for the three selection ratios for all conditions are shown in Tables 5, 6, and 7, (3) summaries of the medians and interquartile ranges for adverse impact for the 1,000 applicant groups are presented in Figures 1-4, (4) summaries of the medians and interquartile ranges for adverse impact for the 10,000 applicant groups are presented in Figures 5 to 8, (5) summaries of the adverse impact for the 1,000,000 applicant groups are presented in Figures 9-12, and (6) summaries of the validity for one 1,000,000 applicant group are shown in Figure 13.

Adverse Impact

The mean adverse impact varied greatly across the several methods of using test scores in all applicant group sizes (Tables 3, 5, 6, and 7). The levels of adverse impact of the several methods on EEO group and gender were similar for all sample sizes and both proportions of minorities. The general patterns may be seen in the summaries of the adverse impact for the 1,000,000 applicant groups (see Figs. 9-12). For example, for the 1,000,000 applicant group, with a proportion of minorities of .5, and a selection ratio of .01, for the M/C cognitive ability test the adverse impact was .13 and .97 for EEO group and gender, respectively; for the Composite the adverse impact was .32 and .21 for EEO group and gender, respectively; and for the Greatest Strength Method the adverse impact was .55 and .62 for EEO group and gender, respectively (See Table 3).

Validity

The validity varied greatly across the several methods of using test scores in all applicant group sizes. The observed validities for the several methods were similar for all conditions (see Tables 5 and 6, and Fig 13). Reported in this paragraph are values for the 1,000,000 applicant group with a proportion of minorities of .5. For the M/C cognitive ability test the validity was .52, for the Composite validity was .70, and for the Greatest Strength Method validity was .50 (see Tables 4 and 7).

	M/C Cognitive		Composite		Composite Without M/C Cognitive		Greatest Strength		Two Greatest Strengths		Drop Lowest Score		Random Selection	
S.R.	EEO	Gender	EEO	Gender	EEO	Gender	EEO	Gender	EEO	Gender	EEO	Gender	EEO	Gender
0.01	0.13	0.97	0.32	0.21	0.70	0.13	0.55	0.62	0.33	0.54	0.31	0.41	1.09	1.01
0.2	0.35	1.00	0.55	0.45	0.83	0.35	0.66	0.64	0.59	0.59	0.56	0.54	1.00	1.00
All Passers	0.88	1.00	0.92	0.90	0.98	0.88	0.86	0.83	0.86	0.83	0.86	0.83	1.00	1.00

Note: S.R. = selection ratio

* With 1,000,000 applicants and proportion of minorities = .5

Table 4. Validity in Appointments for Seven Methods of Using Test Scores*						
M/C Cognitive	Composite	Composite Without M/C Cognitive	Greatest Strength	Two Greatest Strengths	Drop Lowest Score	Random Selection
0.52	0.70	0.57	0.50	0.61	0.67	0.00

* With 1,000,000 applicants and proportion of minorities = .5

TABLE 5. Validity and Adverse Impact by EEO, Gender, Selection Ratio, Proportion of Minorities, and Grading Method

Number of Applicants = 1,000

Grading Method			M/C Cognitive Ability			Composite			Composite Without M/C			Greatest Strength			Two Greatest Strengths			Drop Lowest Score			Random Selection		
Selection Ratio	Proportion of Minorities		Validity	Adverse Impact EEO	Adverse Impact Gender	Validity	Adverse Impact EEO	Adverse Impact Gender	Validity	Adverse Impact EEO	Adverse Impact Gender	Validity	Adverse Impact EEO	Adverse Impact Gender	Validity	Adverse Impact EEO	Adverse Impact Gender	Validity	Adverse Impact EEO	Adverse Impact Gender	Validity	Adverse Impact EEO	Adverse Impact Gender
0.01	0.1	Mean	0.51	0.14	1.28	0.70	0.37	0.23	0.57	0.85	0.15	0.52	0.63	0.79	0.62	0.39	0.73	0.67	0.35	0.53	0.00	1.16	1.22
		N	750	750	747	750	750	750	750	750	750	750	750	747	750	750	749	750	750	750	750	750	749
		Std. Dev.	0.02	0.42	1.06	0.02	0.70	0.22	0.02	1.08	0.15	0.02	0.99	0.63	0.02	0.77	0.65	0.02	0.67	0.48	0.03	1.47	0.91
	0.5	Mean	0.52	0.16	1.33	0.70	0.37	0.26	0.57	0.88	0.15	0.50	0.66	0.77	0.61	0.38	0.67	0.67	0.36	0.50	0.00	1.43	1.26
		N	750	750	743	750	750	750	750	749	750	750	749	746	750	750	750	750	750	748	750	746	747
		Std. Dev.	0.02	0.21	1.23	0.02	0.33	0.33	0.02	0.74	0.18	0.02	0.54	0.71	0.02	0.35	0.71	0.02	0.31	0.62	0.03	1.24	1.06
	Total	Mean	0.52	0.15	1.30	0.70	0.37	0.24	0.57	0.87	0.15	0.51	0.64	0.78	0.61	0.38	0.70	0.67	0.36	0.51	0.00	1.30	1.24
		N	1500	1500	1490	1500	1500	1500	1500	1499	1500	1500	1499	1493	1500	1500	1499	1500	1500	1498	1500	1496	1496
		Std. Dev.	0.02	0.33	1.15	0.02	0.55	0.28	0.02	0.93	0.16	0.03	0.80	0.67	0.02	0.60	0.68	0.02	0.52	0.55	0.03	1.36	0.99
0.2	0.1	Mean	0.51	0.30	1.01	0.70	0.52	0.44	0.57	0.83	0.36	0.52	0.60	0.67	0.62	0.54	0.62	0.67	0.52	0.56	0.00	1.01	1.01
		N	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750
		Std. Dev.	0.02	0.12	0.13	0.02	0.16	0.07	0.02	0.20	0.06	0.02	0.18	0.10	0.02	0.17	0.09	0.02	0.15	0.08	0.03	0.23	0.15
	0.5	Mean	0.52	0.36	1.00	0.70	0.55	0.45	0.57	0.84	0.36	0.50	0.66	0.65	0.61	0.60	0.59	0.67	0.57	0.54	0.00	1.01	1.01
		N	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750
		Std. Dev.	0.02	0.05	0.14	0.02	0.08	0.07	0.02	0.11	0.06	0.02	0.10	0.10	0.02	0.08	0.09	0.02	0.08	0.08	0.03	0.12	0.14
	Total	Mean	0.52	0.33	1.01	0.70	0.53	0.45	0.57	0.83	0.36	0.51	0.63	0.66	0.61	0.57	0.60	0.67	0.54	0.55	0.00	1.01	1.01
		N	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500
		Std. Dev.	0.02	0.09	0.13	0.02	0.12	0.07	0.02	0.16	0.06	0.03	0.15	0.10	0.02	0.13	0.09	0.02	0.12	0.08	0.03	0.18	0.15
All Passers	0.1	Mean	0.51	0.82	1.00	0.70	0.90	0.90	0.57	0.97	0.88	0.52	0.79	0.83	0.62	0.79	0.83	0.67	0.79	0.83	0.00	1.00	1.00
		N	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750
		Std. Dev.	0.02	0.05	0.07	0.02	0.04	0.06	0.02	0.04	0.06	0.02	0.07	0.07	0.02	0.07	0.07	0.02	0.07	0.07	0.03	0.04	0.07
	0.5	Mean	0.52	0.88	1.00	0.70	0.92	0.90	0.57	0.98	0.88	0.50	0.86	0.83	0.61	0.86	0.83	0.67	0.86	0.83	0.00	1.00	1.00
		N	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750	750
		Std. Dev.	0.02	0.02	0.07	0.02	0.02	0.06	0.02	0.02	0.06	0.02	0.04	0.07	0.02	0.04	0.07	0.02	0.04	0.07	0.03	0.02	0.07
	Total	Mean	0.52	0.85	1.00	0.70	0.91	0.90	0.57	0.97	0.88	0.51	0.82	0.83	0.61	0.82	0.83	0.67	0.82	0.83	0.00	1.00	1.00
		N	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500
		Std. Dev.	0.02	0.05	0.07	0.02	0.04	0.06	0.02	0.03	0.06	0.03	0.07	0.07	0.02	0.07	0.07	0.02	0.07	0.07	0.03	0.03	0.07
Total	0.1	Mean	0.51	0.42	1.10	0.70	0.60	0.53	0.57	0.88	0.46	0.52	0.67	0.76	0.62	0.57	0.72	0.67	0.55	0.64	0.00	1.06	1.08
		N	2250	2250	2247	2250	2250	2250	2250	2250	2250	2250	2250	2247	2250	2250	2249	2250	2250	2250	2250	2250	2249
		Std. Dev.	0.02	0.38	0.63	0.02	0.47	0.31	0.02	0.64	0.32	0.02	0.59	0.38	0.02	0.49	0.39	0.02	0.44	0.31	0.03	0.86	0.54
	0.5	Mean	0.52	0.46	1.11	0.70	0.61	0.54	0.57	0.90	0.46	0.50	0.73	0.75	0.61	0.61	0.70	0.67	0.59	0.62	0.00	1.14	1.09
		N	2250	2250	2243	2250	2250	2250	2250	2249	2250	2250	2249	2246	2250	2250	2250	2250	2250	2248	2250	2246	2247
		Std. Dev.	0.02	0.33	0.73	0.02	0.30	0.33	0.02	0.44	0.33	0.02	0.33	0.42	0.02	0.29	0.43	0.02	0.28	0.39	0.03	0.74	0.63
	Total	Mean	0.52	0.44	1.10	0.70	0.60	0.53	0.57	0.89	0.46	0.51	0.70	0.76	0.61	0.59	0.71	0.67	0.57	0.63	0.00	1.10	1.08
		N	4500	4500	4490	4500	4500	4500	4500	4499	4500	4500	4499	4493	4500	4500	4499	4500	4500	4498	4500	4496	4496
		Std. Dev.	0.02	0.36	0.68	0.02	0.40	0.32	0.02	0.55	0.33	0.03	0.48	0.40	0.02	0.40	0.41	0.02	0.37	0.35	0.03	0.80	0.59

TABLE 6. Validity and Adverse Impact by EEO, Gender, Selection Ratio, Proportion of Minorities, and Grading Method

Number of Applicants = 10,000

Proportion of Minorities		M/C Cognitive Ability Validity	Adverse Impact EEO	Adverse Impact Gender	Composite Validity	Adverse Impact EEO	Adverse Impact Gender	Composite Without M/C Validity	Adverse Impact EEO	Adverse Impact Gender	Greatest Strength Validity	Adverse Impact EEO	Adverse Impact Gender	Two Greatest Strengths Validity	Adverse Impact EEO	Adverse Impact Gender	Drop Lowest Score Validity	Adverse Impact EEO	Adverse Impact Gender	Random Selection Validity	Adverse Impact EEO	Adverse Impact Gender
	0.1 Mean	0.51	0.12	1.00	0.70	0.32	0.20	0.57	0.73	0.13	0.52	0.54	0.65	0.62	0.29	0.59	0.67	0.31	0.44	0.00	1.01	1.00
	N	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75
	Std. Dev.	0.01	0.12	0.19	0.00	0.19	0.05	0.01	0.30	0.04	0.01	0.26	0.14	0.01	0.18	0.12	0.00	0.18	0.11	0.01	0.38	0.19
	0.5 Mean	0.52	0.14	1.01	0.70	0.32	0.21	0.57	0.71	0.13	0.50	0.56	0.64	0.61	0.33	0.54	0.67	0.30	0.42	0.00	1.07	1.01
	N	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75
	Std. Dev.	0.01	0.04	0.23	0.00	0.08	0.05	0.01	0.15	0.04	0.01	0.12	0.14	0.01	0.07	0.10	0.01	0.07	0.08	0.01	0.22	0.17
Total	Mean	0.52	0.13	1.00	0.70	0.32	0.21	0.57	0.72	0.13	0.51	0.55	0.64	0.61	0.31	0.57	0.67	0.31	0.43	0.00	1.04	1.01
	N	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150
	Std. Dev.	0.01	0.09	0.21	0.01	0.14	0.05	0.01	0.23	0.04	0.01	0.20	0.14	0.01	0.13	0.11	0.00	0.13	0.10	0.01	0.31	0.18
	0.1 Mean	0.51	0.30	1.00	0.70	0.52	0.44	0.57	0.82	0.35	0.52	0.61	0.67	0.62	0.55	0.61	0.67	0.52	0.55	0.00	1.00	1.00
	N	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75
	Std. Dev.	0.01	0.04	0.04	0.00	0.05	0.02	0.01	0.07	0.02	0.01	0.06	0.03	0.00	0.05	0.02	0.00	0.05	0.02	0.01	0.06	0.04
	0.5 Mean	0.52	0.35	1.00	0.70	0.55	0.45	0.57	0.84	0.35	0.50	0.66	0.64	0.61	0.59	0.59	0.67	0.57	0.54	0.00	1.01	1.00
	N	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75
	Std. Dev.	0.01	0.01	0.04	0.00	0.02	0.02	0.01	0.03	0.02	0.01	0.03	0.03	0.01	0.02	0.03	0.00	0.02	0.02	0.01	0.03	0.04
Total	Mean	0.52	0.33	1.00	0.70	0.53	0.44	0.57	0.83	0.35	0.51	0.63	0.66	0.62	0.57	0.60	0.67	0.54	0.55	0.00	1.00	1.00
	N	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150
	Std. Dev.	0.01	0.04	0.04	0.00	0.04	0.02	0.01	0.05	0.02	0.01	0.05	0.03	0.01	0.05	0.03	0.00	0.04	0.02	0.01	0.05	0.04
	0.1 Mean	0.51	0.81	1.00	0.70	0.90	0.90	0.57	0.97	0.88	0.52	0.79	0.83	0.62	0.79	0.83	0.67	0.79	0.83	0.00	1.00	1.00
	N	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75
	Std. Dev.	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.02
	0.5 Mean	0.52	0.88	1.00	0.70	0.92	0.90	0.57	0.97	0.88	0.50	0.86	0.83	0.61	0.86	0.83	0.67	0.86	0.83	0.00	1.00	1.00
	N	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75	75
	Std. Dev.	0.01	0.00	0.02	0.00	0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.02
Total	Mean	0.52	0.85	1.00	0.70	0.91	0.90	0.57	0.97	0.88	0.51	0.82	0.83	0.61	0.82	0.83	0.67	0.82	0.83	0.00	1.00	1.00
	N	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150
	Std. Dev.	0.01	0.03	0.02	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.04	0.02	0.01	0.04	0.02	0.01	0.04	0.02	0.01	0.01	0.02
	0.1 Mean	0.51	0.41	1.00	0.70	0.58	0.51	0.57	0.84	0.45	0.52	0.65	0.71	0.62	0.54	0.68	0.67	0.54	0.61	0.00	1.00	1.00
	N	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225
	Std. Dev.	0.01	0.30	0.11	0.00	0.27	0.29	0.01	0.20	0.32	0.01	0.19	0.12	0.01	0.23	0.13	0.00	0.22	0.18	0.01	0.22	0.11
	0.5 Mean	0.52	0.46	1.00	0.70	0.60	0.52	0.57	0.84	0.45	0.50	0.69	0.70	0.61	0.59	0.65	0.67	0.57	0.60	0.00	1.02	1.00
	N	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225
	Std. Dev.	0.01	0.31	0.13	0.00	0.25	0.29	0.01	0.14	0.32	0.01	0.14	0.12	0.01	0.22	0.14	0.00	0.23	0.18	0.01	0.13	0.10
Total	Mean	0.52	0.43	1.00	0.70	0.59	0.52	0.57	0.84	0.45	0.51	0.67	0.71	0.61	0.57	0.67	0.67	0.56	0.60	0.00	1.01	1.00
	N	450	450	450	450	450	450	450	450	450	450	450	450	450	450	450	450	450	450	450	450	450
	Std. Dev.	0.01	0.31	0.12	0.01	0.26	0.29	0.01	0.17	0.32	0.01	0.17	0.12	0.01	0.23	0.14	0.00	0.23	0.18	0.01	0.18	0.11

TABLE 7. Validity and Adverse Impact by EEO, Gender, Selection Ratio, Proportion of Minorities, and Scoring Method

Number of Applicants = 1,000,000

| Scoring Method		Written			Composite			Composite Without Written			Greatest Strength			Two Greatest Strengths			Drop Lowest Score			Random Selection		
Selection Ratio	Proportion of Minorities	Validity	Adverse Impact EEO	Adverse Impact Gender	Validity	Adverse Impact EEO	Adverse Impact Gender	Validity	Adverse Impact EEO	Adverse Impact Gender	Validity	Adverse Impact EEO	Adverse Impact Gender	Validity	Adverse Impact EEO	Adverse Impact Gender	Validity	Adverse Impact EEO	Adverse Impact Gender	Validity	Adverse Impact EEO	Adverse Impact Gender
0.01	0.1	0.51	0.12	0.99	0.70	0.31	0.20	0.57	0.72	0.13	0.52	0.53	0.63	0.62	0.30	0.58	0.67	0.31	0.43	0.00	1.00	1.00
	0.5	0.52	0.13	0.97	0.70	0.32	0.21	0.57	0.70	0.13	0.50	0.55	0.62	0.61	0.33	0.54	0.67	0.31	0.41	0.00	1.09	1.01
	Total	0.52	0.12	0.98	0.70	0.31	0.20	0.57	0.71	0.13	0.51	0.54	0.63	0.61	0.31	0.56	0.67	0.31	0.42	0.00	1.04	1.00
0.2	0.1	0.51	0.30	1.00	0.70	0.52	0.44	0.57	0.82	0.35	0.52	0.60	0.66	0.62	0.54	0.61	0.67	0.52	0.55	0.00	1.00	1.00
	0.5	0.52	0.35	1.00	0.70	0.55	0.45	0.57	0.83	0.35	0.50	0.66	0.64	0.61	0.59	0.59	0.67	0.56	0.54	0.00	1.00	1.00
	Total	0.52	0.33	1.00	0.70	0.53	0.44	0.57	0.83	0.35	0.51	0.63	0.65	0.61	0.57	0.60	0.67	0.54	0.54	0.00	1.00	1.00
All Passers	0.1	0.51	0.81	1.00	0.70	0.90	0.90	0.57	0.97	0.88	0.52	0.79	0.83	0.62	0.79	0.83	0.67	0.79	0.83	0.00	1.00	1.00
	0.5	0.52	0.88	1.00	0.70	0.92	0.90	0.57	0.98	0.88	0.50	0.86	0.83	0.61	0.86	0.83	0.67	0.86	0.83	0.00	1.00	1.00
	Total	0.52	0.85	1.00	0.70	0.91	0.90	0.57	0.97	0.88	0.51	0.82	0.83	0.61	0.82	0.83	0.67	0.82	0.83	0.00	1.00	1.00
Total	0.1	0.51	0.41	0.99	0.70	0.58	0.51	0.57	0.84	0.45	0.52	0.64	0.71	0.62	0.54	0.67	0.67	0.54	0.60	0.00	1.00	1.00
	0.5	0.52	0.46	0.99	0.70	0.60	0.52	0.57	0.84	0.45	0.50	0.69	0.70	0.61	0.59	0.65	0.67	0.58	0.59	0.00	1.03	1.00
	Total	0.52	0.43	0.99	0.70	0.59	0.52	0.57	0.84	0.45	0.51	0.66	0.70	0.61	0.57	0.66	0.67	0.56	0.60	0.00	1.02	1.00

Figure 1. Adverse Impact EEO by Grading Method

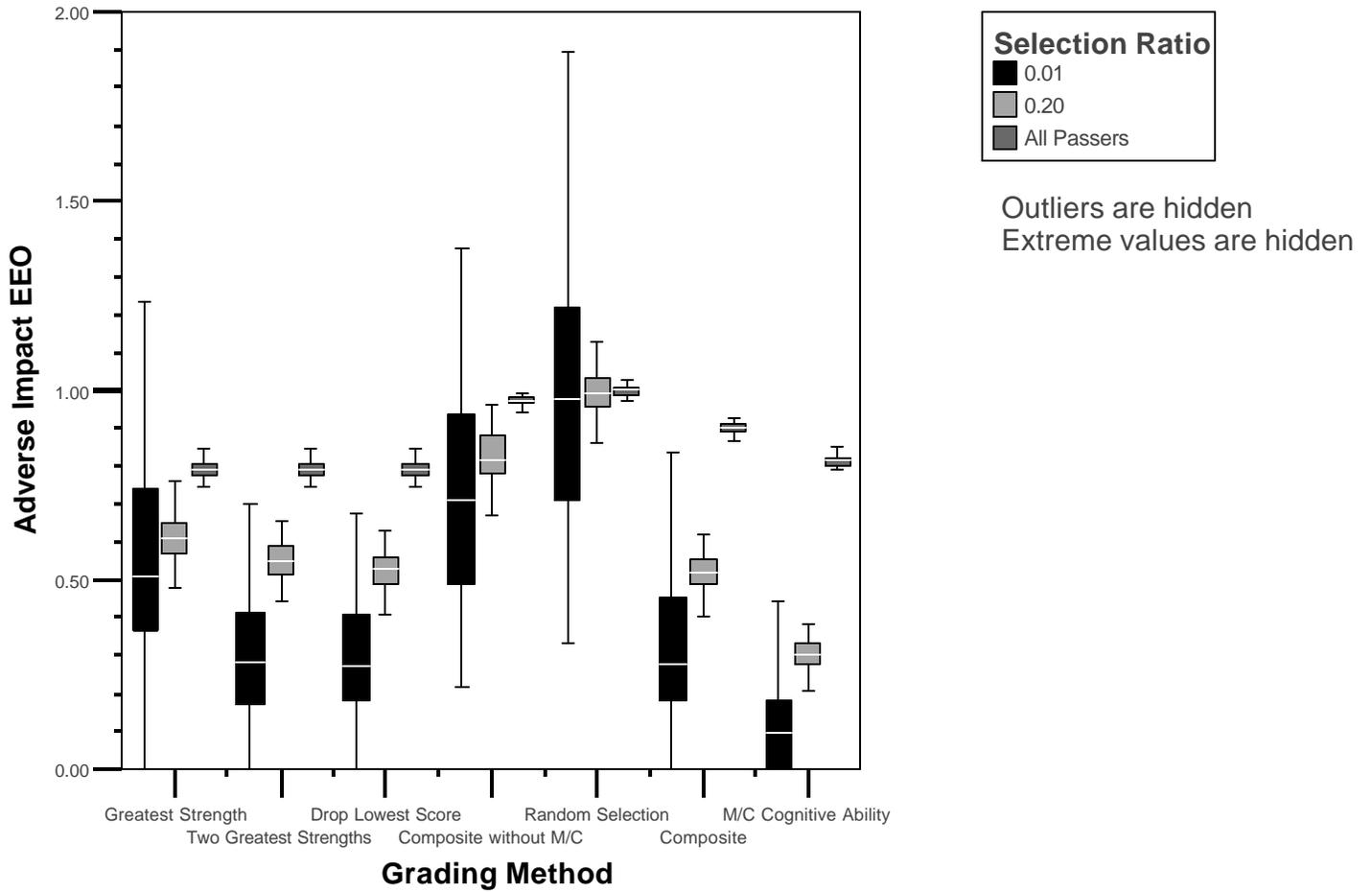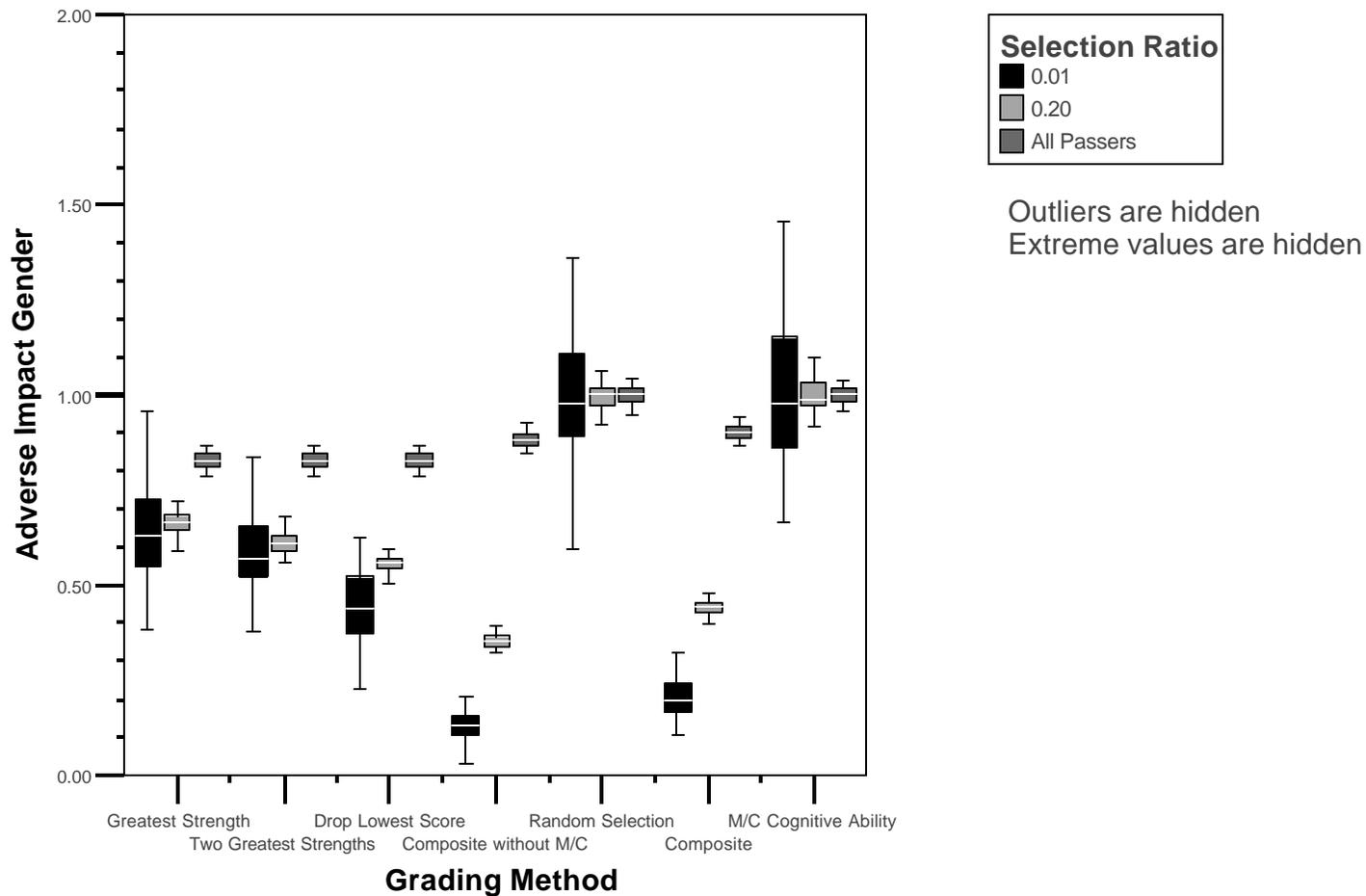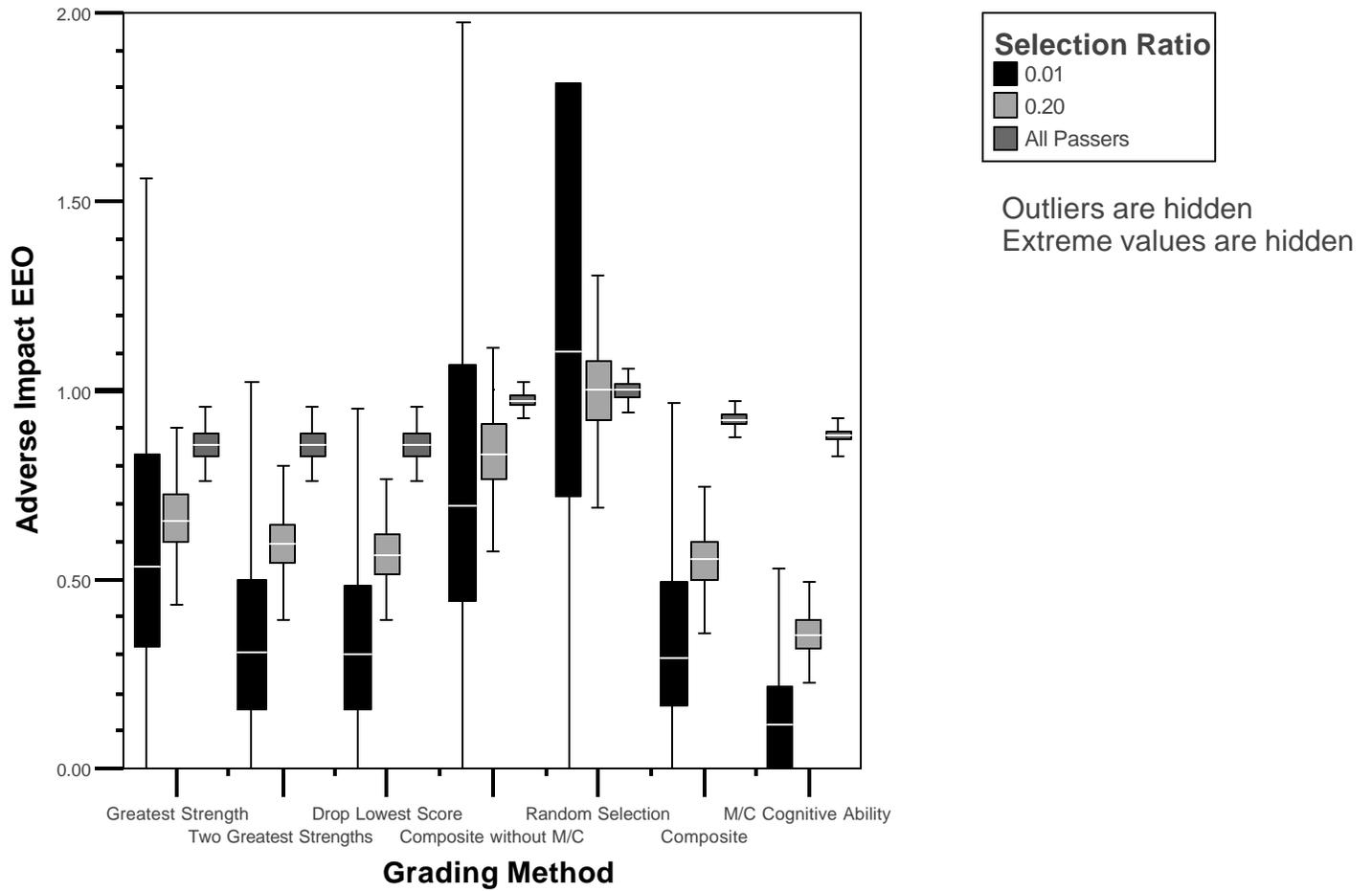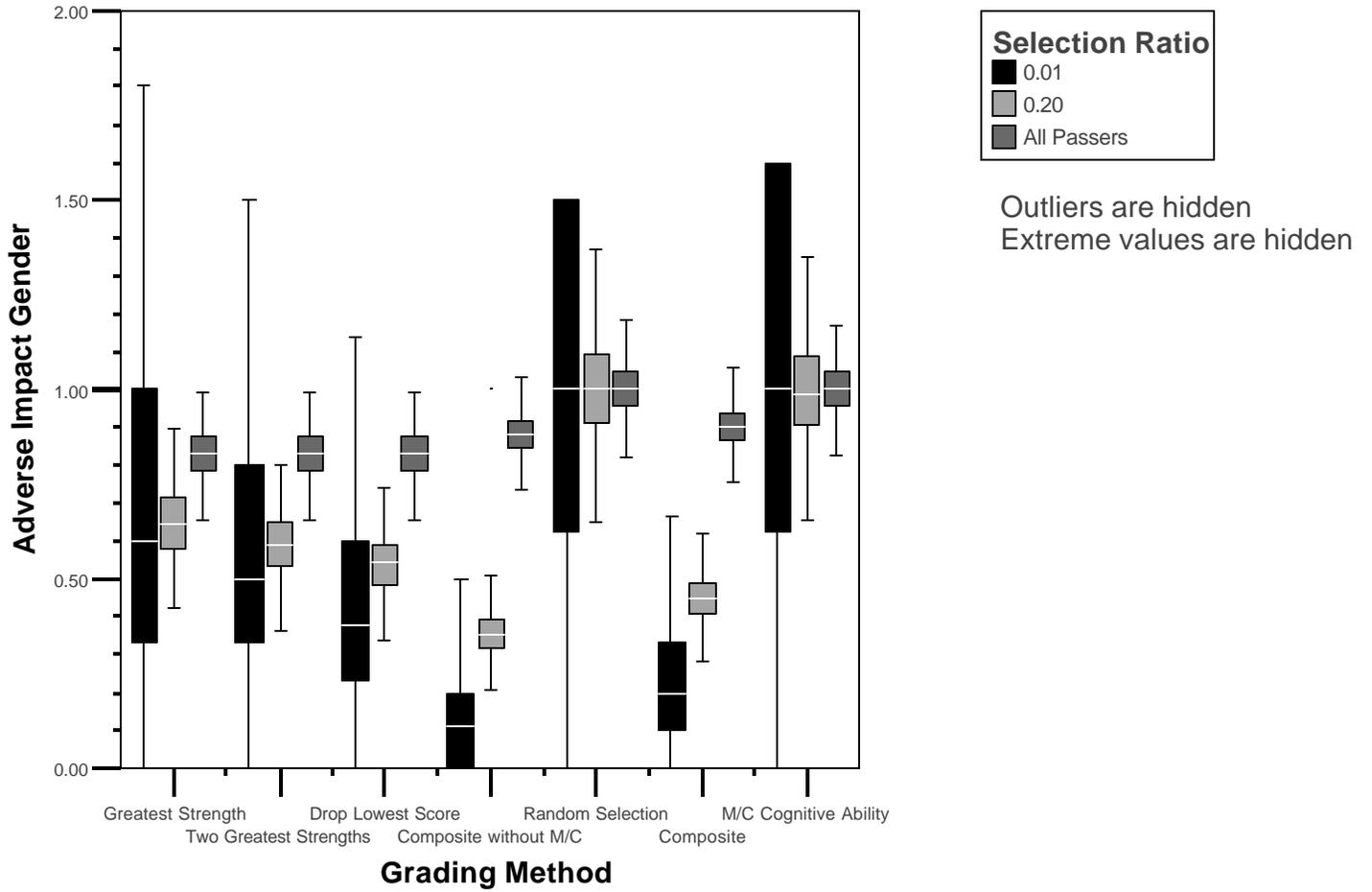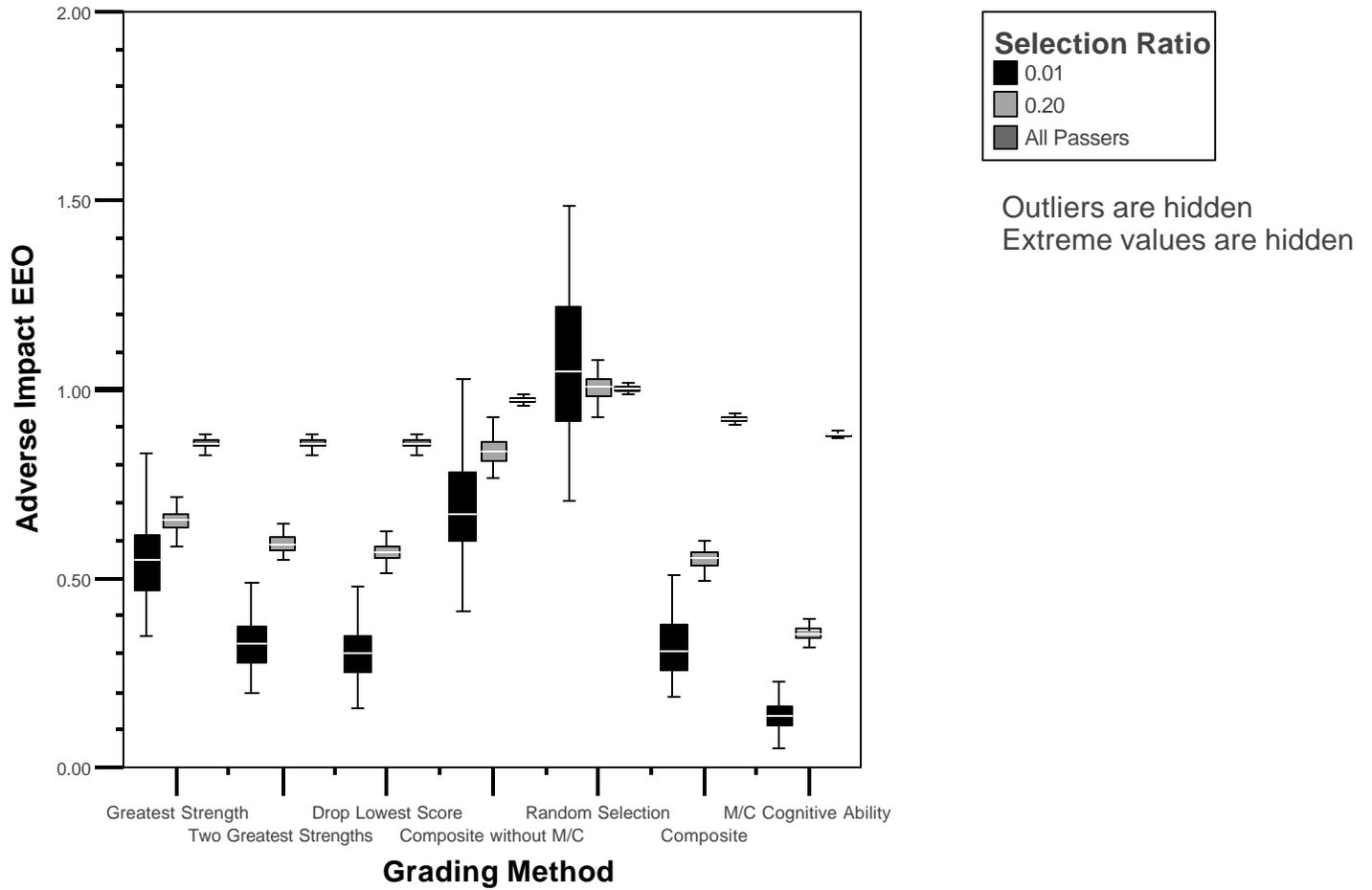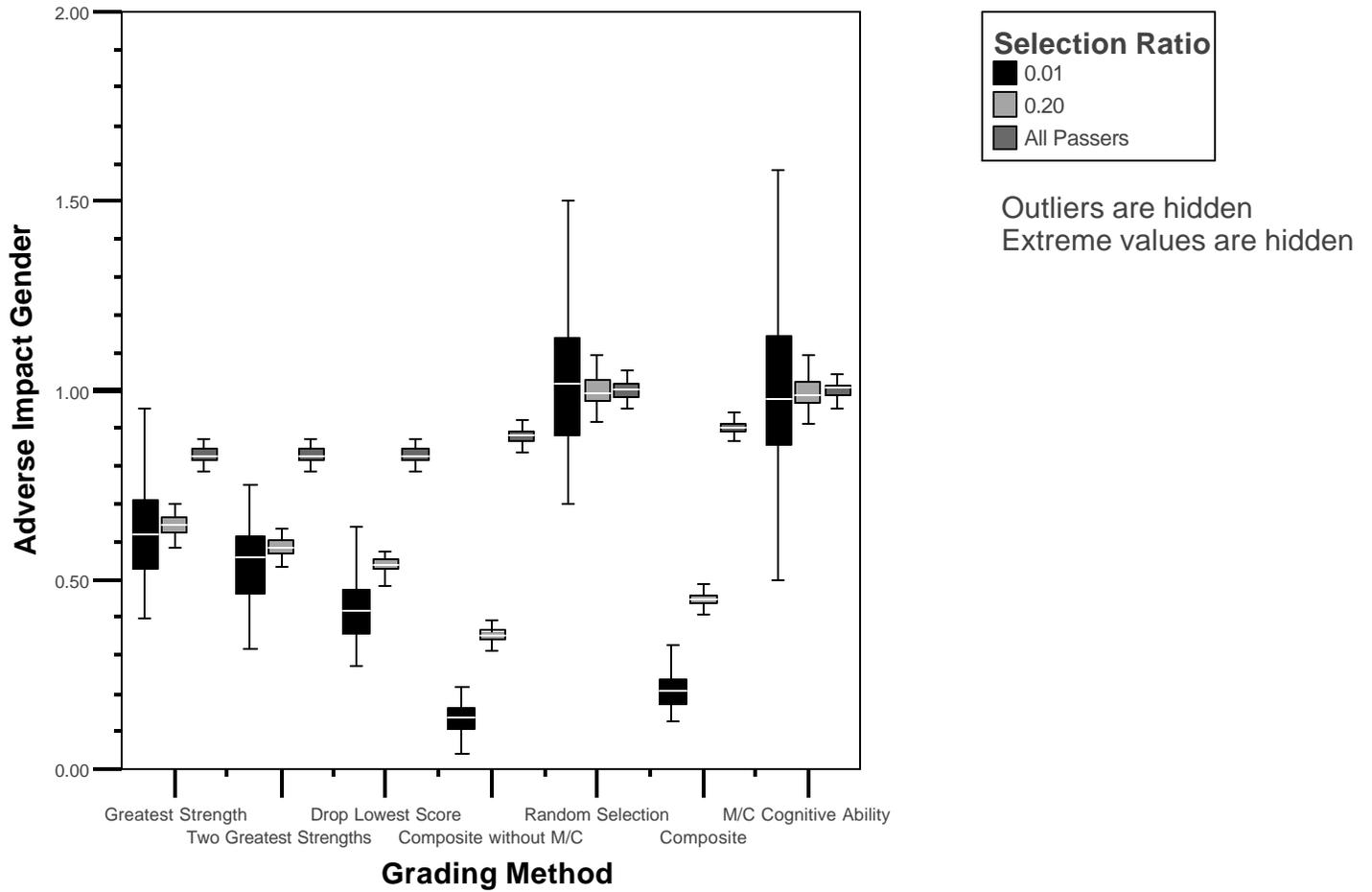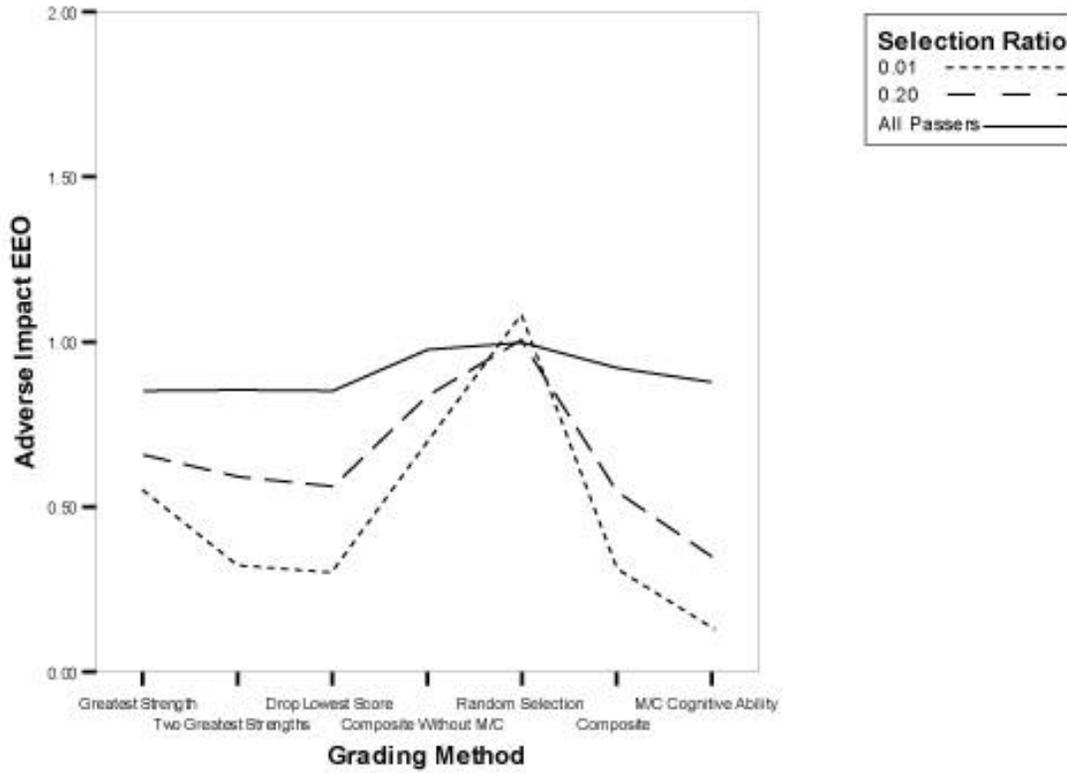*Proportion of Minorities = .1, Applicant Group Size = 1000, Number of Replications = 750*
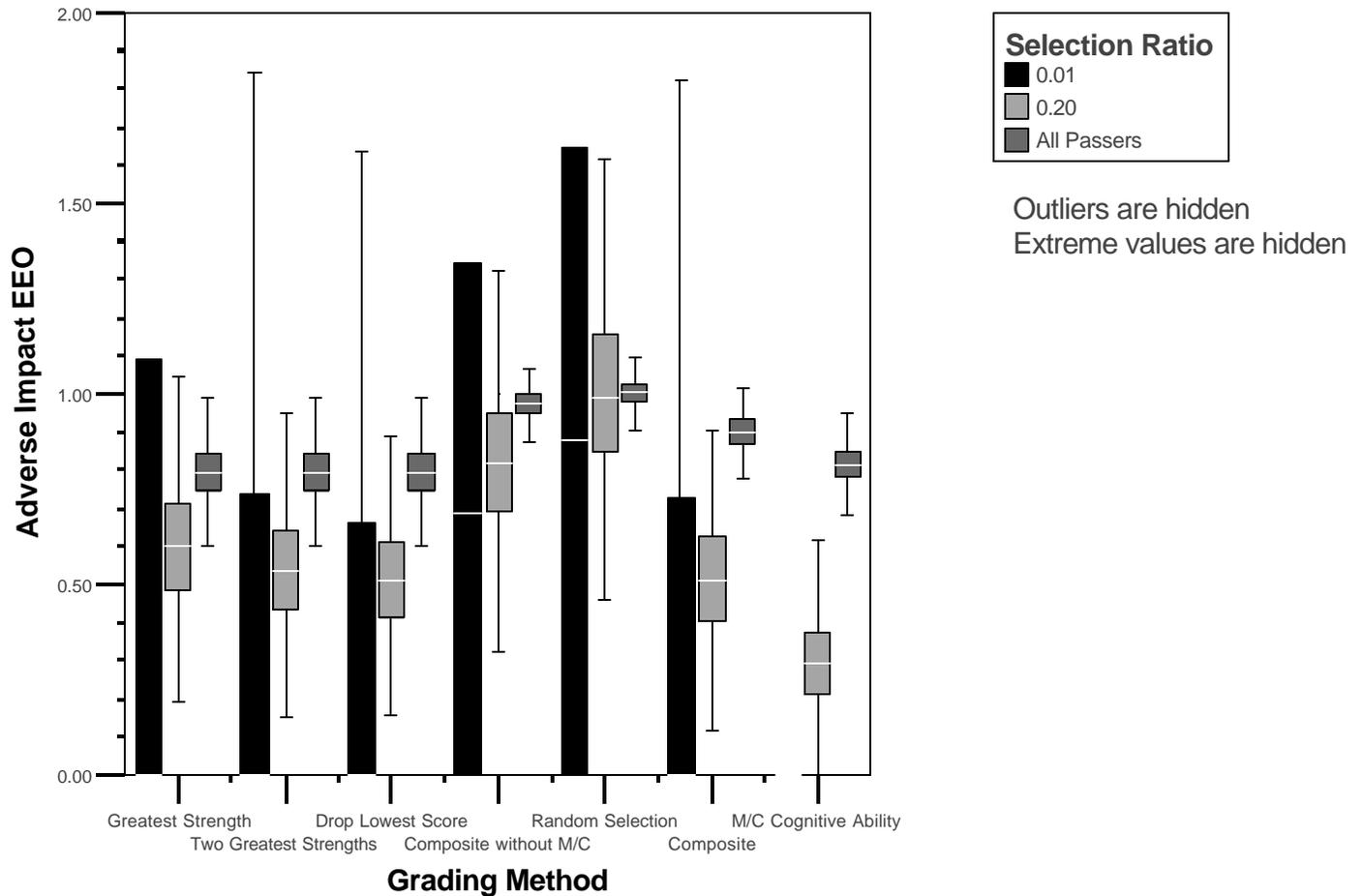
Note: Figures 1-8 plot calculated values for adverse impact. The length of the box is one interquartile range, the whiskers are 1.5 interquartile ranges from the ends of the box, and the white line in the middle of the box shows the median value.

Figure 2. Adverse Impact Gender by Grading Method

Proportion of Minorities = .1, Applicant Group Size = 1000, Number of Replications = 750
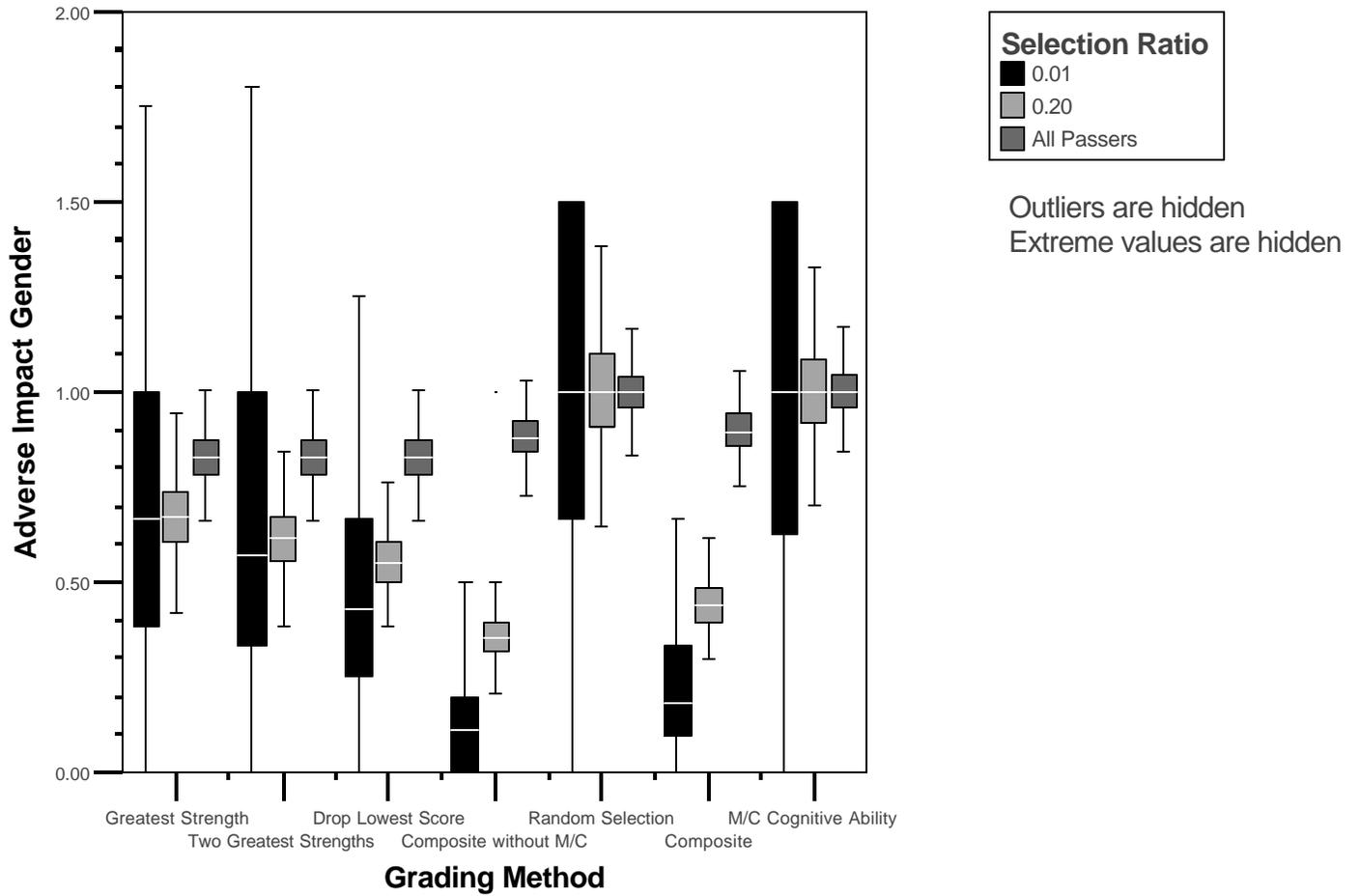
Figure 3. Adverse Impact EEO by Grading Method

Proportion of Minorities = .1, Applicant Group Size = 10000, Number of Replications = 75

Figure 4. Adverse Impact Gender by Grading Method

Proportion of Minorities = .1, Applicant Group Size = 10000, Number of Replications = 75

Figure 5. Adverse Impact EEO by Grading Method

Proportion of Minorities = .5, Applicant Group Size = 1000, Number of Replications = 750

Figure 6. Adverse Impact Gender by Grading Method

Proportion of Minorities = .5, Applicant Group Size = 1000, Number of Replications = 750

Figure 7. Adverse Impact EEO by Grading Method

Proportion of Minorities = .5, Applicant Group Size = 10000, Number of Replications = 75

Figure 8. Adverse Impact Gender by Grading Method

Proportion of Minorities = .5, Applicant Group Size = 10000, Number of Replications = 75

Figure 9. Adverse Impact EEO by Grading Method

Proportion of Minorities = .5, Applicant Group Size = 1,000,000

Note: Figures 9-13 report the results of one run of group size = 1,000,000, therefore only one value plotted for each condition (unlike in Figures 1-8).

Figure 10. Adverse Impact Gender by Grading Method

Proportion of Minorities = .5, Applicant Group Size = 1,000,000
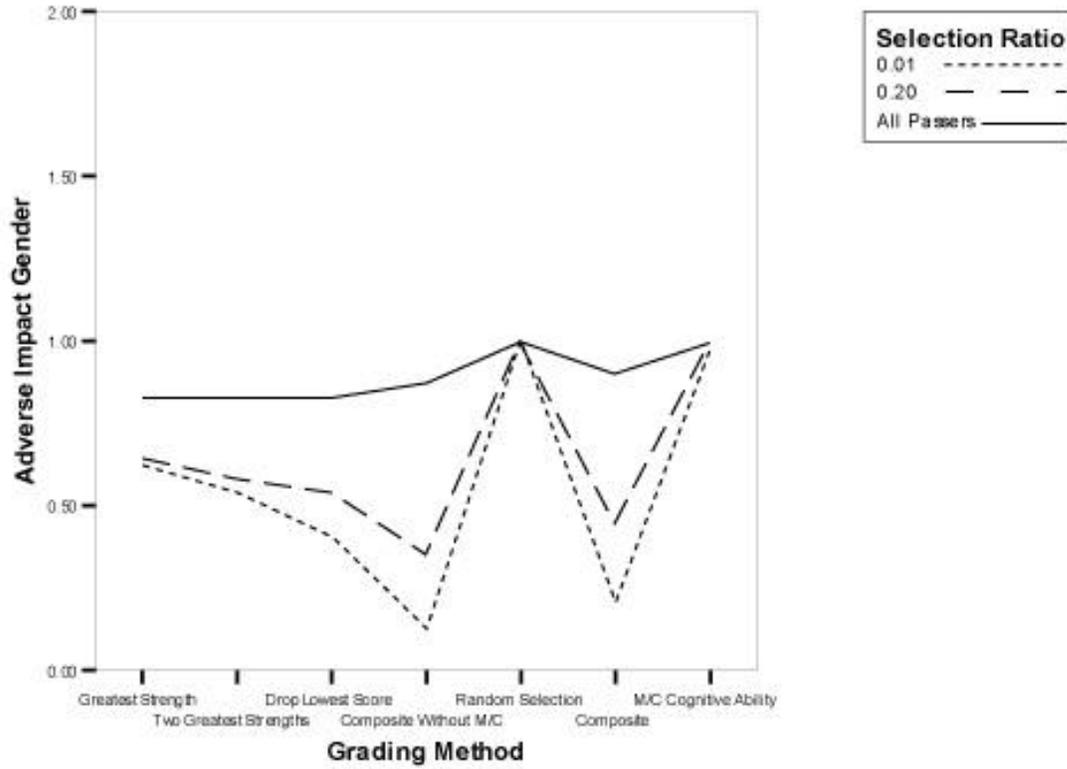
Figure 11. Adverse Impact EEO by Grading Method

Proportion of Minorities = .1, Applicant Group Size = 1,000,000
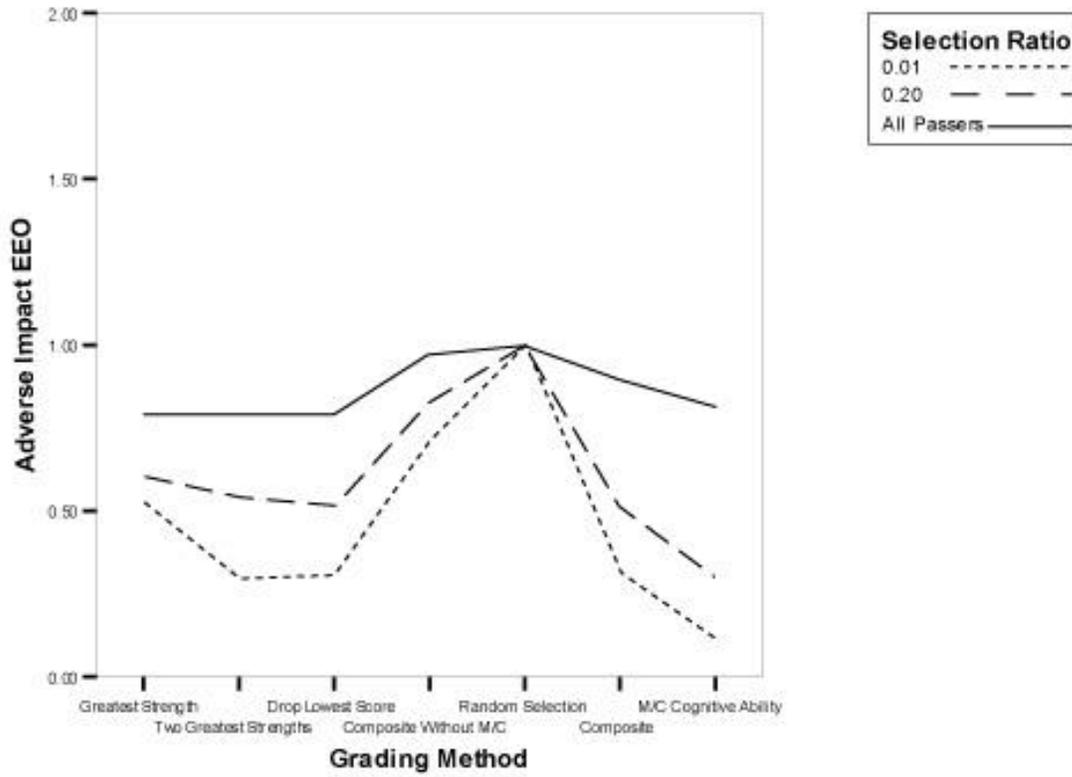
Figure 12. Adverse Impact Gender by Grading Method

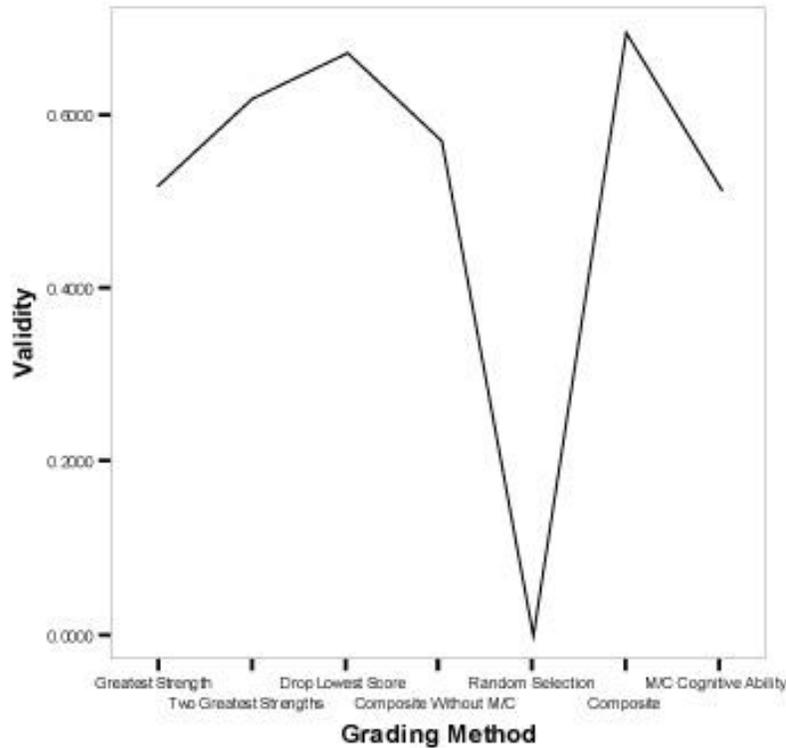Proportion of Minorities = .1, Applicant Group Size = 1,000,000
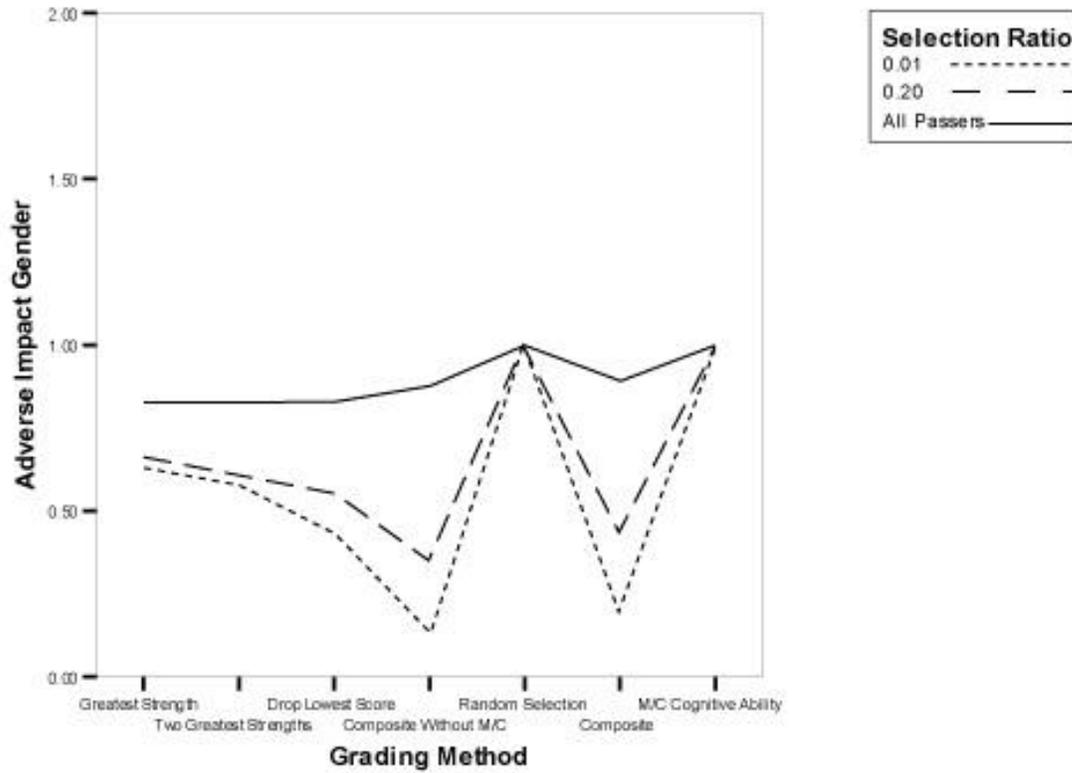
Figure 13. Validity by Grading Method

Proportion of Minorities = .1, Applicant Group Size = 1,000,000

Discussion

The two most salient findings are that the alternative methods of using the selection test scores:

(1) have validities comparable to or higher than the validity of the written M/C test of cognitive ability alone (.70, .57, .50, .61, .67 and .52 for the composite, composite without M/C, greatest strength, two greatest strengths, drop the lowest score methods, and the M/C cognitive ability test, respectively, for applicant group size 1,000,000 and proportion of minorities of .5), and

(2) have lower levels of adverse impact than the written M/C test of cognitive ability alone (.32, .70, .55, .33, .31, and .13 for the composite, composite without M/C, greatest strength, two greatest strengths, and drop the lowest score methods, and the M/C cognitive ability test only, respectively, for applicant group size 1,000,000 and proportion of minorities of .5).

The adverse impact for EEO of the Composite was substantially lower than the adverse impact for

Page 26

the M/C cognitive ability test, unlike the findings of Potosky, Bobko, and Roth (2005). This is likely due to the choice of tests simulated.

Another salient finding is that the validity of one of the innovative methods (drop the lowest score, .67) of using selection test scores is almost as high as that of the traditional composite (.70).

Finally, adverse impact varies by method of test score use and selection ratio both with respect to central tendency and variability. For some conditions the interquartile range was very large. For example, the 25th percentile of adverse impact values for all the methods (including the random selection method!) was zero when the proportion of minority applicants was .1 and the selection ratio was .01 (see Fig. 1). The practical import of the relatively high variability is that with very small selection ratios and very low proportions of minority applicants the adverse impact for any one applicant group may be expected to vary considerably from the mean.

Limitations of This Research

A limited number of values were investigated. In real life, the number of applicants and the selection ratio may be different from those studied, and may be beyond the control of the organization.

Only certain types of tests were considered. Other tests (e.g, situational judgement, memory, perceptual speed) may yield different results.

The values used for mean group differences and for intercorrelations between tests and between tests and the criterion are meant to be typical. In practical application, the actual values may differ from these for various reasons. For example, a particular written M/C cognitive ability test may be better or worse than typical and therefore may be more or less reliable and valid, and may have more or less adverse impact.

Conclusions

All of the innovative grading methods showed validity greater than or almost equal to that of the written M/C test of cognitive ability alone, and in some cases markedly higher validity was found. The Greatest Strength grading method had the least simultaneous adverse impact for EEO and gender (both had relatively little impact at the same time). While the Greatest Strength Method did not improve on the validity of the M/C cognitive ability test, the improvement in adverse impact is substantial.

This Monte Carlo study suggests that some of the innovative test use methods studied here may be expected to reduce adverse impact while maintaining traditional (if not maximal) levels of validity. There is potential for greater reductions in adverse impact if more tests are used (such as work style or personality tests).

References

Aguinis, H. (1994). A Quickbasic Program for Generating Correlated Multivariate Random Normal Scores. *Educational and Psychological Measurement*, Vol. 54 No. 3, 687-689.

Cascio, W.F., Outtz, J., Zedeck, S. & Goldstein, I.L. (1991). Statistical implications of six methods of test score use in personnel selection. *Human Performance*, 4, 233-264.

Civil Rights Act of 1964, 42 U.S.C. Section 2000e-2.k.1.A.i (2002)

Civil Rights Act of 1991, Public Law No. 102-166, 105 Stat. 1071 (1991). Codified as amended at 42 U.S.C., Section 1981, 2000e et seq.

Equal Employment Opportunity Commission, Department of Labor, Department of Justice, and Civil Service Commission (1978). *Uniform Guidelines on Employee Selection Procedures*. 43 Federal Register 38295 (August 25, 1978).

Foldes, H.J. (In Press). Group Differences in Personality: Meta-Analyses Comparing Five U.S. Racial Groups. *Personnel Psychology*, in press.

McKay, P.F., McDaniel, M.A. (2006). A Reexamination of Black–White Mean Differences in Work Performance: More Data, More Moderators. *Journal of Applied Psychology*, Vol. 91 No. 3, 538-554.

Potosky, D., Bobko, P., and Roth, P.L. (2005). Forming Composites of Cognitive Ability and Alternative Measures to Predict Job Performance and Reduce Adverse Impact: Corrected Estimates and Realistic Expectations. *International Journal of Selection and Assessment,* Vol. 13 No. 4, 304-315.

Pulakos, E.D., Schmitt, N. (1996). An Evaluation of Two Strategies for Reducing Adverse Impact and Their Effects on Criterion-Related Validity. *Human Performance*, Vol. 9, 241-258.

Roth, P.L., Huffcutt, A.I., Bobko, P. (2003). Ethnic Group Differences in Measures of Job Performance: A New Meta-Analysis. *Journal of Applied Psychology*, Vol. 88 No. 4, 694-706.

Sackett, P.R., Roth, L. (1996). Multi-Stage Selection Strategies: A Monte Carlo Investigation of Effects on Performance and Minority Hiring. *Personnel Psychology*, Vol. 49, 549-572.

Wiesen, J.P. (2006a, June) Thinking Outside the Box in Merit Selection. Paper presented at the annual International Personnel Management Association Assessment Council Conference, Las Vegas, NV.

Wiesen, J.P. (2006b, November) Thinking Outside the Box in Merit Selection. Paper presented at the MAPAC Fall Conference. Princeton, NJ.

Wiesen, J.P. (2007, June) Possible New Approaches Reduce Adverse Impact. Paper presented at the annual International Personnel Management Association Assessment Council Conference, St. Louis, MO.