A Comparison of Predictor-Based and Criterion-Based Methods for Weighing Predictors to Reduce Adverse Impact

Keith Hattrup
San Diego State University

Joanna Rock
Aon Consulting

The present study explored the effects of differing strategies for weighing predictor and criterion measures on adverse impact and predicted performance. A Monte Carlo simulation based on meta-analytic evidence of multiple predictor and criterion relationships was used to illustrate the trade-offs involved in weighing predictors without first considering how to weigh criterion dimensions. Results illustrate that decisions to weigh predictor measures that fail to incorporate consideration of multiple predictor-criterion relationships may lead to unintended consequences for average predicted job performance in the organization.

Tests of cognitive ability have demonstrated substantial and cross-situationally consistent criterion-related validities in predicting performance in most jobs in the U.S. (e.g. Hunter & Hunter, 1984; Schmidt & Hunter, 1981; Schmitt, Gooding, Noe, & Kirsch, 1984). However, an average mean difference of about one standard deviation between African-Americans and Whites is often observed on tests of cognitive ability, resulting in a higher selection ratio for Whites than for African-Americans when cognitive ability tests are used to make personnel selection decisions (e.g. Gottfredson, 1988; Hunter & Hunter, 1984; Roth, BeVier, Bobko, Switzer, & Tyler, 2001). As a consequence, use of cognitive ability tests alone often results in *adverse impact* against African-Americans and other ethnic groups (e.g. Campbell, 1996; Gottfredson, 1988; Sackett, Schmitt, Ellingson, & Kabin, 2001).

Given the desire to increase representation of ethnic minorities in organizations, and a desire to avoid costly litigation associated with findings of adverse impact, considerable attention has recently been directed to the identification of strategies for reducing adverse impact through the use of alternative or additional predictors of job performance (e.g. Hattrup, Rock, & Scalia, 1997; Hough & Oswald, 2000; Sackett et al., 2001; Sackett & Ellingson, 1997; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). Tests of personality, in particular, have been shown to result in small to non-existent mean differences between ethnic groups (e.g. Ones, Viswesvaran, & Schmidt, 1993; Pulakos & Schmitt, 1996). Thus, findings from a number of studies have shown that adverse impact decreases when predictors with smaller mean differences between ethnic

groups, such as personality tests, are used in selecting job applicants (e.g. Schmitt et al., 1997).

Two very different conceptual approaches characterize the way in which decisions are made about how to combine tests of cognitive ability with alternative predictors when selecting job applicants (e.g. see Sackett et al., 2001). The first approach, which we call the "predictor-based" approach, weighs predictors according to the effects of the weights on (a) adverse impact, and (b) criterion-related validity in predicting "overall" job performance (e.g. Bobko, Roth, & Potosky, 1999; Ryan, Ployhart, & Friedel, 1998; Sackett & Ellingson, 1997; Schmitt et al., 1997). Studies that have emphasized this approach have shown that, although adverse impact is unlikely to be eliminated, it can be reduced substantially as the weight given to personality tests increases relative to the weight given to cognitive ability when forming a composite predictor (e.g. Sackett & Ellingson, 1997; Sackett & Wilk, 1994).

The second, "criterion-based", approach first identifies the appropriate weights to place on dimensions of job performance when forming a composite criterion (e.g. DeCorte, 1999; Hattrup et al., 1997). Predictors, such as cognitive ability and personality, are then weighed through the use of multiple regression procedures according to their relationships with the composite criterion. Research that has emphasized this second approach has demonstrated that, although adverse impact appears not to be eliminated, it can be reduced substantially when criterion dimensions that correlate with personality predictor tests are given more weight than criterion dimensions that correlate with cognitive ability measures (Hattrup et al., 1997).

The differences between the two approaches are not trivial. The criterion-based approach emphasizes choices about the importance of different job-relevant behaviors (e.g. Hattrup et al., 1997; Murphy & Shiarella, 1997). Adverse impact in the criterion-based model is a consequence of the values placed on different types of job-relevant behavior. The predictor-based approach emphasizes choices about predictors without necessarily considering the impact of such choices on the profiles of performance across multiple dimensions of job performance (Hattrup et al., 1997). Although both the predictor-based and criterion-based approaches have been recognized as alternative strategies for determining the weights to place on predictor measures when selecting job applicants, research has not yet been presented that compares the two methods.

The purpose of the present study, therefore, is to compare these two approaches in terms of their effects on adverse impact and predicted performance. In particular, we attempt to illustrate the trade-offs that are made when one approach is taken and not the other. In so doing, we offer some clarity to the kinds of costs and benefits of different approaches to weighing predictors that have, to date, not been explicitly addressed in the literature. Hence, the present study contributes both to our conceptual understanding of methods of identifying and weighing predictor measures, and to pragmatic decision making in organizations.

Predictor-Based Approaches to the Identification and Weighing of Selection Tools

A large number of studies have recently explored the effects of weighing various predictor measures on adverse impact and minority hiring (e.g. Ryan et al., 1998; Sackett & Roth, 1996; Sackett & Ellingson, 1997; Schmitt et al., 1997). This research is generally driven by the recognition that predictors differ in the extent to which they demonstrate distributional differences between demographic groups (e.g. Hough & Oswald, 2000; Ones et al., 1993; Pulakos & Schmitt, 1996). Tests of cognitive ability result in the largest demonstrated mean differences between African-Americans and Whites among predictors that have been considered, with a difference between means of 1 standard deviation reported in a number of studies (e.g. Hunter & Hunter, 1984). Measures of conscientiousness, in contrast, are reported to have mean differences between African-Americans and Whites of zero, to as much as .09 standard deviations (e.g. Bobko et al., 1999). Clearly,, the use of conscientiousness alone when selecting job applicants would result in little to no adverse impact against African-Americans.

Much of the research conducted according to the predictor-based approach seeks to identify the magnitude of reduction in adverse impact when alternative predictors are used in conjunction with cognitive ability tests to select applicants. For example, Sackett and Wilk (1994) examined a simple case of a composite measure comprised of two uncorrelated predictors which differ substantially in mean subgroup differences. They showed that the reduction in adverse impact that results from the use of the composite measure is not as substantial as might be expected. Many other studies have examined the effects of differential weighing of predictors on adverse impact given more realistic assumptions about the correlations among predictors and the magnitude of subgroup differences on components of a multi-predictor composite (e.g. Bobko et al., 1999; Ryan et al., 1997). Sackett and Ellingson (1997), for example, illustrated that the reduction in adverse impact that results from combining predictors that differ in subgroup mean differences depends on, among other things, the magnitude of correlation between the predictors.

Schmitt et al. (1997) examined the effects of combining an expanded set of predictor measures, including structured interviews, biodata, personality, and cognitive ability, on adverse impact of a final composite formed from various combinations of these measures. As expected, adverse impact was reduced when predictors with smaller subgroup mean differences were included in the composite. Bobko et al. (1999) reported similar findings. Ryan et al. (1998) examined a similar set of issues using data obtained in actual criterion-related validation studies; studies reported by other authors were based on simulations of data that would likely be obtained in representative or prototypical personnel selection problems.

An additional interest of several of the studies that have adopted a predictor-based approach is the criterion-related validity of the composite measure when used to predict "overall" job performance. Ones et al. (1993), for example, showed that adverse impact is lower and overall job performance is

highest when a regression-weighted composite of cognitive ability and conscientiousness is used to select job applicants. Adverse impact was higher when cognitive ability was used alone, and overall predicted performance was lower when conscientiousness was used as the only predictor. Similar findings were reported by Schmitt et al. (1997) and Bobko et al. (1999). Ryan et al. (1998) and Sackett and Ellingson (1997) examined the effects of various predictor combination strategies on adverse impact without explicitly examining the effects of these strategies on criterion-related validity of the resulting predictor composite.

Criterion-Based Approaches to the Identification and Weighing of Selection Devices

The criterion-based approach focuses attention on validity in a multivariate model that includes multiple dimensions of job performance. Unidimensional conceptualizations of job performance, or models that rely solely on the concept of "overall" job performance, have been strongly criticized in the field (e.g. Austin & Villanova, 1992; Murphy & Shiarella, 1997; Smith, 1976). As Murphy and Shiarella (1997) noted, personnel selection is almost always a multivariate problem, in that multiple predictors are used to increase the likelihood that employees will exhibit desired behaviors in multiple criterion dimensions. A number of taxonomic models of the job performance domain have been proposed (see Austin & Villanova, 1992; Borman & Motowidlo, 1993; Campbell, Gasser, & Oswald, 1996; Campbell, McCloy, Oppler, & Sager, 1993; Motowidlo, Borman, & Schmidt, 1997; Motowidlo & Schmidt, 1999). Campbell et al. (1993; 1996), for example, describe eight dimensions of performance that are presumed to underlie performance in most jobs, including job-specific task proficiency, non-job-specific task proficiency, written and oral communication, effort, personal discipline, facilitation of team and peer performance, supervision/leadership, and management/administration.

Borman and Motowidlo (1993; see also Motowidlo & Schmidt, 1999) suggested that job performance behaviors can be categorized more parsimoniously with two broader dimensions that are hypothesized to exist in virtually all jobs. These include the dimensions of *task performance* and *contextual performance*. Task performance refers to behaviors that contribute to core transformation and maintenance processes in the organization. Behaviors such as producing products, selling merchandise, acquiring inventory, managing subordinates, or delivering services are considered examples of task performance (Motowidlo & Schmit, 1999).

Contextual performance, on the other hand, represents behaviors that do not directly contribute to core transformation and maintenance processes in the organization. Instead, contextual performance contributes to the culture and climate of the organization, or more generally, the context within which transformation and maintenance processes occur (Borman & Motowidlo, 1993; Motowidlo, Borman, & Schmidt, 1997; Motowidlo & Schmit, 1999). Although contextual performance is not necessarily extra-role in nature, it is often more

affective in tone than task performance. It includes helping, prosocial, and citizenship behaviors in the organization, such as (a) volunteering for extra work, (b) persisting with enthusiasm, (c) helping and cooperating with others, (d) following rules and procedures, and (e) supporting or defending the organization both inside and outside the organization (Borman & Motowidlo, 1993).

Considerable empirical evidence has demonstrated that task and contextual performance are best predicted by different individual difference variables that might be used when selecting job applicants (e.g. Campbell, McHenry, & Wise, 1990; Day & Silverman, 1989; Hattrup, et al., 1997; Murphy & Shiarella, 1997; Motowidlo, Borman, & Schmidt, 1997; Motowidlo & VanScotter, 1994; Van Scotter & Motowidlo, 1996). In particular, task performance behaviors are more strongly correlated with individual differences in cognitive ability than with individual differences in personality constructs, such as conscientiousness or work orientation. Contextual behaviors, in contrast, are more strongly correlated with individual differences in personality traits, such as conscientiousness or work orientation.

There is a broad consensus that both task performance and contextual performance are important to organizations (e.g. Motowidlo & Schmit, 1999; Murphy & Shiarella, 1997). Supervisor ratings of overall job performance have been shown to be influenced by both types of behaviors (e.g. Borman, White, & Dorsey, 1995; Motowidlo & Van Scotter, 1994; Van Scotter & Motowidlo, 1996), and both dimensions have been related to indices of organizational effectiveness (Motowidlo & Schmit, 1999; Podsakoff & MacKenzie, 1997). Therefore, personnel selection procedures should seek to identify applicants who can exhibit high levels of both contextual and task performance. By implication, therefore, organizations would be wise to use measures of both cognitive ability and personality when selecting job applicants. Applicants who obtain high scores on measures of cognitive ability are predicted to demonstrate effective task performance behaviors, and applicants who score high on measures of personality, such as conscientiousness or work orientation, are predicted to demonstrate high levels of contextual performance.

Although empirical research has documented the relevance of both task and contextual performance to organizations, it is also important to recognize that the relative importance of these two general dimensions of behavior is likely to vary across jobs and organizations (e.g. Motowidlo & Schmit, 1999; Murphy & Shiarella, 1997). As a consequence, the relative weight given to task versus contextual performance must be considered an expression of values in the organization about the types of behaviors that are desired and rewarded (e.g. Motowidlo & Schmit, 1999; Murphy & Shiarella, 1997). Given that cognitive ability and personality have different relationships with task versus contextual performance, the values placed on these two dimensions of performance will have implications for the weights to be assigned to predictor measures when selecting job applicants (Hattrup et al., 1997; Murphy & Shiarella, 1997). As the importance of contextual performance to the organization increases relative to the importance of task performance, personality becomes more relevant and useful as a predictor relative to cognitive ability (e.g. Hattrup et al., 1997; Murphy &

Shiarella, 1997). This can be seen explicitly when a composite of value-weighted criterion dimensions is regressed on ability and personality predictors (Hattrup et al., 1997).

Thus, the criterion-based approach to weighing predictors emphasizes the importance of explicitly considering organization values about the importance of different types of behaviors in the organization, or for a particular job (Hattrup et al., 1997). These values then determine the relative weights to be assigned to task and contextual performance when forming a composite criterion. When the value-weighted composite is regressed on predictors, the regression weights assigned to predictors are optimal in predicting the profile of job-relevant behaviors that the organization values (Hattrup et al., 1997; Murphy & Shiarella, 1997). Because predictors differ in the extent to which demographic group mean differences are observed, the relative values of task and contextual performance in the organization will determine the extent to which a composite predictor will result in adverse impact (Hattrup et al., 1997). As contextual performance increases in importance in the organization, an optimally weighted predictor composite will emphasize personality to a greater extent, and thus result in lower adverse impact. A greater emphasis on task performance relative to contextual performance will result in an optimally-weighted predictor composite in which cognitive ability is weighed more, and thus, results in greater adverse impact (Hattrup et al., 1997). The criterion-based approach, therefore, leads to decisions about weighing predictors that result from an explicit consideration of organization values about the importance of different kinds of behavior in the organization. As such, adverse impact varies as a function of the relative importance of different kinds of behavior at work, in a fashion that, as Campbell (1996) noted, fits "…squarely within the domain of business necessity (p. 144)."

Comparing Predictor and Criterion-Based Approaches

Both the predictor-based and criterion-based approaches to weighing alternative predictors have examined the criterion-related validities of predictor batteries formed from differently-weighted components that include cognitive ability and personality. The key difference between the two approaches is the model of the criterion domain that either explicitly, or implicitly, underlies the method. To date, predictor-based strategies have examined the effects of various strategies of weighing predictor measures on criterion-related validity in predicting a unidimensional "overall" job performance construct (Bobko et al., 1999; Schmitt et al., 1997). The criterion-based approach emphasizes the multidimensional nature of job performance, and the role of organizational values in determining the relative importance of various types of job-relevant behaviors (Hattrup et al., 1997).

In a variety of instances, the differences between the two approaches would be relatively trivial in terms of the effects of the strategies on the criterion-related validity of the composite predictor measure. If task and contextual performance were highly correlated, or personality and ability were highly correlated, or the correlations of ability and personality with the two criterion

27

dimensions were equivalent, the validity of the composite predictor is unlikely to vary as predictor weights are varied (Murphy & Shiarella, 1997). However, considerable empirical evidence has accumulated that demonstrates weak relationships between task and contextual performance, weak relationships between personality and ability, and differential relationships between task and contextual performance with personality and ability (e.g. Hattrup et al., 1997; Motowidlo et al., 1997; Murphy & Shiarella, 1997). Therefore, as Murphy and Shiarella (1997) demonstrated, criterion-related validity of a composite predictor formed from ability and personality will depend on the congruence between the weights assigned to the two predictors and the relative weights given to task versus contextual performance. In particular, if personality is weighed more than ability, the criterion-related validity of the composite predictor decreases as the importance of task performance increases relative to contextual performance. Likewise, criterion-related validity of a composite predictor that weighs ability more than personality decreases as the importance of contextual performance increases relative to task performance. Criterion-related validity is maximized when the predictor weights reflect the relative importance of criterion dimensions. Given that predicted performance in the organization is a function of criterion-related validity, it follows that average predicted performance in an organization will be highest when the weights assigned to predictor measures maximize the multiple correlation with a value-weighted criterion composite.

Hence, a predictor-based strategy that involves weighing personality and ability to influence adverse impact may result in unintended consequences for the profile of performance observed across multiple criterion dimensions (Hattrup et al., 1997). Weighing personality higher than ability in an effort to reduce adverse impact will result in lower predicted performance in the organization as the relative importance of task performance behaviors increases. These unintended consequences have not been explicitly examined in the literature. Thus, this study seeks to clarify trade-offs between reducing adverse impact through predictor weighing strategies, and maximizing performance on valued criterion dimensions. It is expected that adverse impact will decrease as personality is weighed more relative to cognitive ability in selecting job applicants. Average predicted performance, however, will depend on the relative importance of task versus contextual performance. Increasing weights on personality will result in lower predicted performance as the importance of task performance increases relative to contextual performance.

Of course, these predictions follow logically and directly from the mathematics of the multivariate relationships that have been observed in the literature (Murphy & Shiarella, 1997). What is of particular interest in the present study is the magnitude of the effects of predictor weights on average predicted job performance, when the relative importance of task versus contextual performance varies. Thus, the present study helps to illustrate the size of the trade-offs that might be made between the goals of reducing adverse impact and maximizing performance across multiple criterion dimensions.

Method

A Monte Carlo simulation was performed to estimate the effects of various strategies for weighing ability and personality measures when forming a predictor composite on adverse impact and predicted performance, when the values of task and contextual performance vary. To maintain consistency with the extant literature, we rely on data obtained from the simulation reported by Hattrup et al. (1997). Thus, results reported in this study are directly comparable to the levels of adverse impact and predicted performance that were observed when criterion dimensions are first weighed according to organizational values, and predictors are then weighed through regression to optimize the prediction of the value-weighted composite criterion (Hattrup et al., 1997).

Data Set

The data used in this simulation were developed using the input correlation matrix reported by Hattrup et al., (1997; see also DeCorte, 1999). The correlations among ability, work orientation (the measure of personality), task performance, and contextual performance are reported in Table 1. These correlations were based on a synthesis of studies that reported relationships between both predictors and measures of both task and contextual performance (see Hattrup et al., 1997). Several recent studies have reported relationships that are very similar to the values presented in Table 1, although they are not included in the calculations of the mean values reported in Table 1. Given our desire to present results that are directly comparable to those presented in other related studies, we use the values reported in Table 1, which reflect the general pattern of relationships observed among these variables in the literature. As Murphy and Shiarella (1997) note, the purpose of a Monte Carlo simulation like the one employed in this study is not necessarily to present a definitive statement about the exact magnitude of effects of various weighting decisions, but rather to illustrate the implications of multidimensional models of job performance for outcomes in personnel selection decision problems.

Table 1
Observed Correlations Between Predictors and Criteria Used in the Study

Variables	Cognitive Ability	Work Orientation	Task Performance	Contextual Performance
Predictors				
Cognitive Ability				
Work Orientation	.07			
Criterion Dimensions				
Task Performance	.41	.12		
Contextual Performance	.16	.26	.17	

Most studies of the effects of various predictor weighting strategies rely on Monte Carlo simulations that are based on observed correlations among variables of interest (e.g. Bobko et al., 1999; Murpy & Shiarella, 1997; Sackett & Ellingson, 1997; Schmitt et al., 1997). Therefore, the observed correlations reported by Hattrup et al. (1997) are used in the present analyses. Data were generated with the matrix functions in SPSS on the basis of the observed values reported in Table 1. Specifically, 7500 random cases were generated representing the *majority* group, and 2500 cases were generated to represent the *minority* group, based on an assumption of a base rate representation of minority group members in the U.S. workforce of .25 (see Hattrup et al., 1997; Jackson & Alvarez, 1992; Kutscher, 1989). Each case had a random score on the cognitive ability, work orientation, task performance, and contextual performance variables. Variables were standardized in the sample of 10,000 cases, and then for each member of the majority group a value of 1.0 was added to the score for cognitive ability, and a value of .50 was added to the task performance score. This had the effect of creating a one standard deviation difference between the majority and minority groups in cognitive ability, and a .50 standard deviation difference in task performance, which are consistent with results reported in the literature (e.g. Sackett & Wilk, 1994). The matrix functions in SPSS were then used to decompose the input correlation matrix into eigenvalues and eigenvectors, which were then used in transformations of the random data to generate a final data set that reproduced the input correlation matrix and possessed the mean group differences noted above.

Dependent Measures

We calculated adverse impact ratios by dividing the minority group selection ratio by the majority group selection ratio, at several levels of overall selection ratio for the entire sample. Adverse impact values of 1.0 indicate proportional selection ratios in the two groups, and values less than 1.0 indicate a higher selection ratio for the majority group than for the minority group. Adverse impact values below .80 violate the "four-fifths rule", and may be taken as *prima facie* evidence of unfair discrimination.

Predicted overall performance was evaluated by using the regression equation to derive the average predicted criterion performance of individuals who would be selected under the various predictor weighing strategies, depending on varying models of the relative importance of task and contextual performance. This was accomplished by first calculating a composite criterion measure that reflected the relative weights that could be applied to the task and contextual performance dimensions. The resulting composite was then re-standardized in the total sample. Predicted job performance was then determined by finding the mean performance score among those cases selected according to the different predictor weighting strategies described below.

Independent Variables

The interest in this study is the effect of predictor-based strategies for weighing cognitive ability and personality on adverse impact and predicted performance, when the relative importance of task and contextual performance varies. Hence, we compared three strategies for weighing ability and personality at each of three levels of relative importance of task and contextual performance. Specifically, we created three composite criterion measures that varied the weights on task and contextual performance. As shown in Table 2, the first composite assigned a weight of 1.0 to task performance and 0.0 to contextual performance; the second composite assigned weights of 1.0 to each dimension, and the third composite assigned a weight of 0.0 to task performance and a weight of 1.0 to contextual performance. Three predictor composites were formed by weighing cognitive ability 1.0 and work orientation 0.0, weighing both predictors 1.0, and weighing ability 0.0 and work orientation 1.0, respectively. Clearly, it is less likely that organizations would assign weights of zero and 1.0 to the predictors and criterion dimensions examined in this study than weights that are more similar across the two predictors or two criterion dimensions. The extreme weights are examined in this study to explore the range of outcomes that might occur under various strategies for weighing predictors and criteria. More realistic weights that are positive and non-zero for both predictors and/or both criteria would produce results that are between results we report for scenarios that involve weighing predictors and/or criteria the same, and scenarios that involve assigning weights of 1.0 and 0.0 for the two predictors or two criteria.

Table 2: Weighting Scenarios

	Scenarios								
	Full weight on cognitive ability			Equal weight			Full weight on work orientation		
	1	2	3	4	5	6	7	8	9
Predictor Weights									
Cognitive ability	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0
Work orientation	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
Criterion Weights									
Task performance	1.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0
Contextual performance	0.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0
R^2	0.17	0.14	0.02	0.13	0.18	0.08	0.01	0.06	0.07

Results

The effects of the predictor and criterion weighting strategies on R^2, adverse impact, and predicted performance are presented in Tables 2-4. As can be seen in Table 3, adverse impact is eliminated when work orientation is used as the only predictor in selecting job applicants. This follows, of course, from the fact that work orientation shows no mean differences between ethnic groups. In contrast, adverse impact is most significant when cognitive ability is used as the only predictor, and is most extreme at lower selection ratios. When both cognitive ability and work orientation are used to form a unit-weighted predictor composite, adverse impact is still present at all selection ratios, although it is lower when selection ratios are higher. These findings are consistent with results reported by others (e.g. Ryan et al., 1997; Sackett & Ellingson, 1998; Sackett & Wilk, 1994), and are typical in studies that adopt a predictor-based approach to the determination of relative predictor weights.

Table 4 shows the effects of various predictor weights on predicted performance when the relative weights on task and contextual performance differ. As can be seen in the table, average predicted performance in the organization is highest when (a) both cognitive ability and work orientation are used in forming a predictor composite, *and* (b) both task and contextual performance are valued in the organization (i.e. Scenario 5 in Table 4). As hypothesized, predicted performance is lowest when work orientation is the only predictor used in selecting job applicants, and task performance is the only criterion dimension valued in the organization (i.e. Scenario 7). Similarly, predicted performance is low when cognitive ability is the only predictor used in selecting job applicants and only contextual performance is valued in the organization (i.e. Scenario 3).

Table 3: Effect of Weighting Scenarios on Adverse Impact Ratios

	Scenarios								
	Full weight on cognitive ability			Equal weight			Full weight on work orientation		
	1	2	3	4	5	6	7	8	9
Selection Ratio									
.05	0.08	0.08	0.08	0.19	0.19	0.19	1.08	1.08	1.08
.10	0.11	0.11	0.11	0.31	0.31	0.31	1.03	1.03	1.03
.20	0.17	0.17	0.17	0.37	0.37	0.37	1.00	1.00	1.00
.30	0.25	0.25	0.25	0.43	0.43	0.43	0.99	0.99	0.99
.40	0.30	0.30	0.30	0.49	0.49	0.49	0.99	0.99	0.99
.50	0.37	0.37	0.37	0.54	0.54	0.54	1.00	1.00	1.00
.60	0.45	0.45	0.45	0.61	0.61	0.61	0.99	0.99	0.99
.80	0.64	0.64	0.64	0.78	0.78	0.78	1.00	1.00	1.00

Note: Adverse impact values depart from 1.00 in Scenarios 7, 8, and 9 3 because of the small degree of sampling error in the data.

These findings are due to the fact that the R^2 in predicting the value-weighted composite criterion is highest when both cognitive ability and work orientation are used as predictors and both task and contextual performance are valued.   The R^2 is lowest when work orientation is used to predict task performance, or when cognitive ability is used to predict contextual performance. As can be seen in Table 4, the decrease in average predicted performance resulting from these scenarios compared to the scenario in which both ability and work orientation are used to predict task and contextual performance (Scenario 5) is as much as .46 to .64 standard deviations at the lowest selection ratios.

Table 4 also shows the results of scenarios that involve weighing both ability and personality, when the values of task and contextual performance vary. In particular, although using a composite of ability and work orientation reduces adverse impact relative to the use of ability alone, predicted performance is maximized only when both task and contextual performance are valued.   As expected, predicted performance is lower when both ability and personality are used in selecting applicants, but only task performance (Scenario 4), or only contextual performance (Scenario 6) is valued.   In other words, predicted performance is sacrificed when the weights used in forming a predictor composite are inconsistent with the values placed on task and contextual performance. Predicted performance is even lower when ability and work orientation are both used in selecting applicants for a job that requires either task (Scenario 4) or contextual performance (Scenario 6), than when ability alone is used to select applicants for a job that requires task performance (Scenario 1) or both task and contextual performance (Scenario 2).

Table 4: Effect of Weighting Scenarios on Mean Predicted Performance

	Scenarios								
	Full weight on cognitive ability			Equal weight			Full weight on work orientation		
	1	2	3	4	5	6	7	8	9
Selection Ratio									
.05	.81	.78	.38	.70	.84	.58	.20	.49	.56
.10	.70	.65	.30	.59	.73	.52	.17	.42	.46
.20	.56	.51	.22	.49	.58	.39	.16	.34	.36
.30	.46	.43	.19	.40	.48	.33	.13	.29	.32
.40	.40	.36	.15	.35	.41	.27	.12	.25	.26
.50	.33	.30	.13	.29	.34	.23	.11	.20	.21
.60	.27	.24	.10	.22	.27	.18	.09	.17	.17
.80	.14	.13	.05	.13	.15	.10	.04	.09	.09

Discussion

In this study, we sought to clarify the trade-offs that are made when predictor-based strategies for reducing adverse impact by differentially weighing predictor tests are implemented without consideration of the implications of these strategies for predicting multidimensional criteria. We illustrated the effects of different strategies of weighing cognitive ability and personality using simulated data that were based on the pattern of correlations observed between alternative predictors and alternative criterion dimensions. Results from this simulation illustrate clearly that the effects of various strategies of weighing ability and personality predictors on average performance in the organization depend on the relative values placed on task and contextual performance in the organization. Specifically, strategies that involve weighing personality in forming a predictor composite will indeed reduce adverse impact. However, such strategies may or may not result in decreases in average predicted performance, depending on the relative importance of task and contextual performance in the organization. When personality is emphasized, but job performance depends more on task than on contextual performance, average predicted performance in the organization is lower than when the relative importance of contextual performance is higher. Weighing personality to reduce adverse impact in hiring will involve a greater trade-off of increasing minority representation against average predicted performance as the importance of task performance increases relative to contextual performance. Strategies for reducing adverse impact without sacrificing job performance are more likely to succeed when the weights on the predictors match the relative values placed on task and contextual performance in the organization.

A key implication of the present study is the need to carefully consider the role of organizational values when deciding how to weigh predictor measures. Simply varying the weight on personality and ability in an effort to reduce adverse impact will have unintended consequences for average performance in the organization, unless the multidimensional nature of performance, including its implications for test validity, are explicitly considered. Unfortunately, this has not been carefully addressed in the literature. Hence, the findings of the present study should remind researchers and practitioners that predictor-based strategies for reducing adverse impact present a potentially incomplete picture of the consequences of weighing personality to reduce adverse impact.

Indeed, it can be argued that decisions about the relative values of various kinds of outcomes in organizations, including profiles of performance across multiple criterion dimensions and adverse impact, need to be explicitly considered before decisions about predictor weights are made. It is common in the literature to argue that overall organizational effectiveness is improved when staffing, performance evaluation, and compensation are clearly and explicitly linked to the goals, values, strategy, and culture of the firm (e.g. Bowen, Ledford, & Nathan, 1991; Schmidt & Kaplan, 1971; Schuler & Jackson, 1987; Wright & McMahan, 1992). In addition, organizations' value statements often include contextual types of behaviors such as respect, integrity, and orientation towards betterment of the

community. Organizations often hold employees accountable for these values in performance reviews, but often fail to consider these values when selecting employees. Therefore, staffing strategies that fail to incorporate what is known about the organization's values for various kinds of behaviors will be less effective than strategies that are based on a careful linking of desired criteria to methods for selecting applicants. The results of the present study illustrate this clearly and explicitly.

Of course, task and contextual performance are not the only criteria that may be valued in an organization. Increasing representation of groups that have been traditionally underrepresented in organizations is an important goal. Indeed, some authors have noted that ethnic and gender diversity is a strategic imperative in many modern organizations (e.g. Jackson & Alvarez, 1992). However, completely satisfactory solutions for weighing diversity *as a criterion* have not been developed. As Hattrup et al. (1997) demonstrate, increasing the importance of contextual performance relative to task performance will increase diversity in the organization compared to strategies that place greater importance on task performance. Additional criterion dimensions might be identified that reflect organizational goals that emphasize both increasing diversity and increasing predicted performance in the organization. For example, criterion behaviors related to tolerance of individual differences, sensitivity to culture, awareness and knowledge of the "minority experience", or being a role model for members of underrepresented groups might be identified and valued as important criteria. If such criteria correlate with group membership in a direction opposite to that of the correlation between task performance and group membership, valuing such criteria may result in increased diversity in the organization, for reasons that are entirely congruent with the desire to improve overall organizational effectiveness.

Although the results of the present study provide clear empirical evidence of the trade-offs involved in weighing predictors without considering multidimensional criteria, several limitations of the research should be noted. First, we relied on previous cumulated evidence of the relationships between task and contextual performance, and cognitive ability and work orientation (Hattrup et al., 1997). This was done to provide results for adverse impact and predicted performance that are directly comparable to the previous published work on criterion-based strategies for weighing predictors. Researchers should be encouraged to replicate these findings in local criterion-validation studies. Results from such local studies would be more relevant to decisions made in a particular setting than the general findings reported here. Our decision to rely on previous meta-analytic evidence and Monte Carlo simulation was driven by a desire to illustrate general patterns of results that are likely to be observed in many settings. Given that cognitive ability and personality have consistently shown differential relationships with task and contextual performance, we expect the pattern of results observed here to generalize across settings.

Another limitation relates to the limited range of predictor and criterion weighing scenarios that were explored in this study. We compared scenarios that represented extreme endpoints of continua that involved maximum differences between weights applied to cognitive ability and personality, and task and

contextual performance. This was done to illustrate the range of outcomes that are possible given differing predictor weights and differing criterion weights. In practice, the relative weights assigned to predictors and the relative values placed on criteria are unlikely to be as disparate as the values explored in this study. As noted above, outcomes under such less extreme weighting scenarios might be found by interpolating between the extreme values reported in Tables 3 and 4. The results reported in Tables 3 and 4are illustrative of the range of outcomes that might be observed given what we know about the differential relationships between ability and personality and task and contextual performance, and the magnitude of ethnic group differences on these variables. As is illustrated in Table 2, there may be substantial effects on average predicted job performance when the relative importance of task performance and contextual performance vary.

Finally, this study is not intended to present a definitive solution to the problem of increasing ethnic diversity in organizations while also maximizing individual performance in the organization. Instead, the present study illustrates the importance of considering the relative values placed on different kinds of behaviors in the organization *before* deciding how to weigh predictors to reduce adverse impact. We urge researchers to continue to explore alternative methods for maximizing both ethnic and gender diversity and predicted performance in organizations. The results of the present study clearly illustrate that predictors should not be differentially weighed in an effort to reduce adverse impact without also carefully considering the consequences for the predicted profile of performance across multiple job performance dimensions.

References

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology, 77*, 836-874.

Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology, 52*, 561-590.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco: Jossey-Bass.

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance, 10*, 99-110.

Borman, W. C., White, L. A., & Dorsey, D. W. (1995). Effects of ratee task performance and interpersonal factors on supervisor and peer performance ratings. *Journal of Applied Psychology, 80*, 168-177.

Bowen, D. E., Ledford, G. E., & Nathan, B. R. (1991). Hiring for the organization, not the job. *Academy of Management Executive, 5*, 35-51.

Campbell, J. P. (1996). Group differences and personnel decisions: Validity, fairness, and affirmative action. *Journal of Vocational Behavior, 49,* 122-158.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt and W. C. Borman (Eds.), *Personnel selection in organizations*. San Francisco, CA: Jossey-Bass.

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology, 43*, 313-334.

Day, D. V., & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel Psychology, 42*, 25-36.

De Corte, W. (1999). A note on the success ratio and the utility of fixed hiring rate personnel selection decisions. *Journal of Applied Psychology, 84*, 952-958.

Gottfredson, L. S. (1988). Reconsidering fairness: A matter of social and ethical priorities. *Journal of Vocational Behavior, 33*, 293-319.

Hattrup, K., O'Connell, M. S., & Wingate, P. H. (1998). Prediction of multidimensional criteria: Distinguishing task and contextual performance. *Human Performance, 11*, 305-319.

Hattrup, K., Rock, J., & Scalia, C. (1997). The effects of varying conceptualizations of job performance on adverse impact, minority hiring, and predicted performance. *Journal of Applied Psychology, 82*, 656-664.

Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future – remembering the past. *Annual Review of Psychology, 51*, 631-664.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.

Jackson, S. E., & Alvarez, E. B. (1992). Working through diversity as a strategic imperative. In S. E. Jackson & Associates (Eds.), *Diversity in the workplace: Human resources initiatives* (pp. 13-35). New York: Guilford Press.

Kutscher, R. E. (1989). Projections summary and emerging issues. *Monthly Labor Review, 112*(11), 66-74.

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology, 79*, 475-480.

Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance, 10,* 71-83.

Motowidlo, S. J., & Schmit, M. J. (1999). Performance assessment in unique jobs. In D. R. Ilgen & E. D. Pulakos (Eds.), *The changing nature of performance* (pp. 56-86). San Francisco: Jossey-Bass.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*, 579-703.

Murphy, K. R., & Shiarella, A. H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: Multivariate frameworks for studying test validity. *Personnel Psychology, 50*, 823-854.

Podsakoff, P. M., & MacKenzie, S. B. (1997). Impact of organizational citizenship behavior on organizational performance: A review and suggestions for future research. *Human Performance, 10*, 133-151.

Pulakos, E. D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance, 9,* 241-258.

Roth, P. L., BeVier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54*, 297-330.

Ryan, A. M., Ployhart, R. E., & Friedel, L. A. (1998). Using personality testing to reduce adverse impact: A cautionary note. *Journal of Applied Psychology, 83,* 298-307.

Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology, 50,* 707-721.

Sackett, P. R., & Roth, L. (1996). Multi-stage selection strategies: A Monte Carlo investigation of effects on performance and minority hiring. *Personnel Psychology, 49*, 549-572.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative action world. *American Psychologist, 56*, 302-318.

Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist, 49*, 929-954.

Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist, 36*, 1128-1137.

Schmidt, F. L., & Kaplan, L. B. (1971). Composite vs. multiple criteria: A review and resolution of the controversy. *Personnel Psychology, 24*, 419-434.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. P. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology, 37*, 407-422.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology, 82*, 719-730.

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

Van Scotter, J. R., & Motowidlo, S. J. (1996). Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of Applied Psychology, 81*, 525-531.

Address correspondence to:

Keith Hattrup
Department of Psychology
San Diego State University
5500 Campanile Dr.
San Diego, CA 92182-4611
khattrup@sunstroke.sdsu.edu