Differential Prediction and the Use of Multiple Predictors: The Omitted Variables Problem

Paul R. Sackett, Roxanne M. Laczo, and Zachary P. Lippe
University of Minnesota

Moderated regression is widely used to examine differential prediction by race or gender. When using multiple predictors in a selection system, guidance as to whether differential prediction analysis should be conducted on each predictor individually, or on the set of predictors in combination, is lacking. Analyzing predictors individually creates the possibility of an omitted variable problem. Army Project A data were used to examine differential prediction by race with the use of personality measures for predictor–criterion combinations. Traditional analysis indicated predictive bias by intercept in 45 instances and by slope in 7 instances; the inclusion of an Armed Services Vocational Aptitude Battery general factor as an additional predictor changed the conclusion in 32 cases for the intercept and in 3 cases for the slope.

Differential prediction (also labeled predictive bias) by race and gender in the use of tests or other assessment instruments for personnel selection has been a long-standing concern. Differential prediction is commonly assessed by using a regression model that tests for differences in slopes, intercepts, and sometimes error variances by examining the within-group regression lines relating test scores to a job-relevant criterion (Cleary, 1968). Under this model, no predictive bias exists if the predictive relationship in the two groups being compared can be described by a common regression line. If the predictive relationship differs in terms of slopes or intercepts, it implies that bias does exist because systematic errors of prediction would be made on the basis of group membership. This is acknowledged as the accepted approach to examining predictive bias in both the Standards for Educational and Psychological Testing (American Educational Research Association/American Psychological Association/National Council on Measurement in Education, 1999) and the Principles for the Validation and Use of Personnel Selection Procedures (The Society of Industrial and Organizational Psychology, 1987).

Differential prediction by race and gender has been widely investigated in the domain of cognitive ability (e.g., Bartlett, Bobko, Mosier, & Hannan, 1978; Dunbar & Novick, 1988; Houston & Novick, 1987; Schmidt, Pearlman, & Hunter, 1981). It has been investigated much less in other domains. For example, Saad and Sackett (2002) reported finding no prior instances of examining differential prediction by race or gender in the personality domain. They examined predictive bias by gender for three personality characteristics, five criteria, and nine jobs using the Army Project A database. They reported findings that parallel those in the cognitive ability domain: Although evidence of differential prediction was found in a number of instances, it generally results in overprediction of female performance. With the growing interest in personality and in noncognitive predictors more generally as supplements to cognitive predictors, we expect that greater attention will be paid to differential prediction using such predictors.

As we began exploring differential prediction by race in using personality to supplement cognitive predictors of job performance, we came to the realization that an issue in regression analysis known as the omitted variables problem was particularly relevant to using differential prediction analysis in this domain. We elaborate below on the meaning and implications of the omitted variables problem. The purpose of this article is to explore conceptually and empirically the implications of the omitted variables problem in settings in which multiple predictors are being used in a selection system.

Consider the following scenarios. First, a researcher has assembled a trial test battery for use in a criterion-related validity study and is selecting a smaller number of predictors for inclusion in an operational battery. The researcher is concerned about predictive bias by race and gender and decides to examine each predictor separately for predictive bias. Second, consider the scenario in which an organization is considering adding a new predictor to an existing selection system. A decision rule is proposed specifying that the new predictor only be added if an analysis reveals no evidence of predictive bias by race or gender. What we show below is that each of these decisions may be misguided and may lead to incorrect conclusions about the presence or absence of differential prediction.

The Omitted Variables Problem in Differential Prediction Research

Attempts to estimate regression coefficients rely on a set of fundamental assumptions, one key assumption being that of a fully specified model. In applied selection systems, a fully specified model is a model including all predictors included in a selection system. Thus if a firm wished to use, say, cognitive ability and
Conscientiousness in a selection system, but tested each of these separately for differential prediction, there would be the potential for an omitted variables problem when examining the predictors separately. Failure to simultaneously examine both predictors can result in a misspecified model, often referred to as the omitted variables problem (e.g., Linn & Werts, 1971).

An omitted variable is only a problem under a specific set of circumstances. If a variable that is related to the criterion variable but uncorrelated with any measured predictor variable is omitted, the result is a poorer fitting model with a larger error term. The regression coefficients for the measured predictor variables, however, are not biased by the omission of such a variable. In contrast, if a variable that is related to the criterion variable and that is correlated with a measured predictor variable is omitted, the regression coefficient for the measured predictor variable can be biased.

That omitted variables may be an issue in differential prediction analysis was recognized early, that is, within 3 years of the Cleary (1968) formulation. Linn and Werts (1971) called attention to the omitted variables problem by using regression analysis for this purpose. Differential prediction analysis interprets the coefficient for the subgroup variable as indicating intercept differences by subgroup and interprets the coefficient for the interaction between subgroup membership and the test in question as indicating slope differences by subgroup. If an omitted variable correlated with the criterion is also correlated with subgroup membership, the coefficients for subgroup membership and for the subgroup–test interaction may be biased. As shown below, this bias can lead to changing conclusions about the presence and nature of differential prediction.

We note several crucial issues in considering the applicability of the omitted variables problem to differential prediction analysis. The first is that concern about omitted variables in differential prediction focuses on variables correlated with subgroup membership rather than other correlates of the test under examination. Omitting a variable correlated with the test could bias the coefficient for the test, but that coefficient is not interpreted in conducting differential prediction analysis. With a given test, the question of interest is whether its predictive relationship with the criterion of interest differs by subgroup. An omitted variable correlated with subgroup membership and with the criterion can affect both the subgroup term and the interaction term in the differential prediction model.

Demonstration of the Omitted Variables Problem

To demonstrate the omitted variables problem, we generated a data set with 1,000 cases. We generated three uncorrelated standard normal variables, A, B, and E: View A as a cognitive ability test, B as a measure of conscientiousness, and E as random error. Specifying these variables as uncorrelated corresponds to the research literature regarding the relationship between ability and Conscientiousness. Assume that the sample is made up of a Black subgroup (N = 100) and a White subgroup (N = 900). Mimicking common findings in the literature, we lowered the ability test scores by 1 SD for each member of the Black subgroup (although see Roth, Bevier, Bobko, Switzer, & Tyler, 2001, for a discussion of settings in which differences from this commonly observed value are found). Members of the Black and White subgroups did not differ on the Conscientiousness scale.

We then created a criterion variable (i.e., job performance) as A + B + 2E. In other words, we created a simple system in which performance is solely a function of A, B, and random error. 2E was used to create a situation in which half of the variance in performance was determined by the predictors of interest, and half the performance variance was error. By definition A and B are unbiased measures. They are the only two systematic determinants of the criterion, and both are measured without error.

Now imagine that a researcher is interested in using both ability and Conscientiousness in a selection system and decides to examine differential prediction by race separately for each predictor. To examine cognitive ability, the researcher would run regression models entering cognitive ability, group membership, and the interaction between the two. The results shown in Table 1 show, ability is related to performance, with no slope or intercept differences by race. This is the correct finding; ability is by definition in this simulation an unbiased predictor of performance.

Next the researcher conducts differential prediction analysis in the same manner as above to determine whether the Conscientiousness measure exhibits differential prediction by race. Here the researcher would run a regression model entering Conscientiousness, group membership, and the interaction between the two. The results are shown in Table 2. Here we see that the coefficient for race is significant, indicating intercept differences between the groups and thus that differential prediction is present. However, the simulated data were designed such that Conscientiousness is an unbiased measure. This is an example of the omitted variable problem. Ability is related to performance, but also correlated with race. With ability omitted from the equation, the shared variance between ability and race is attributed to race, biasing the race coefficient and leading to the conclusion of differential prediction for the Conscientiousness variable.

What happens when the omitted variable is included in the model? Table 3 presents the results of an alternative regression that included ability as a control variable in addition to Conscientiousness, race, and the race–Conscientiousness interaction. Table 3 shows that when the omitted variable—ability—is in the equation, the race variable is not significant. Both ability and Conscientiousness are revealed as predictive of performance, with no differential prediction.

Finally, we present a differential prediction analysis on the composite of ability and Conscientiousness. As Table 4 shows, neither race nor the Race × Composite interaction are significant, showing that the use of the composite does not result in differential prediction.

This finding that the Conscientiousness measure, which we know to be an unbiased measure, produces differential prediction when ability is not included in the regression equation draws attention to two very different ways in which differential prediction analyses can be interpreted. It highlights the crucial distinction between bias in a test and bias in a selection system (Draggo, 1984). If it can be shown that differential prediction is present when a test is examined alone in the traditional moderated regression framework but disappears once all other relevant variables are included in the regression equation, one can conclude that the test itself is not the source of the predictive bias. One also concludes that a selection system that includes all of the variables in the fully
specified model does not exhibit predictive bias. However, the decision to use the test alone as the selection system, without also including the other variables in the fully specified model, would result in a biased selection system. In the example being treated here, using Conscientiousness alone as a predictor would indeed result in biased prediction; however, the fact that the differential prediction disappears once ability is added to the model shows that the remedy is not to be found in modifying the Conscientiousness measure but rather in broadening the selection system to include other relevant predictors.

As a concrete example, we describe one domain in which underprediction of subgroup performance is found for a protected subgroup, namely the use of the SAT in predicting college performance. The performance of women is commonly underpredicted. Stricker, Rock, and Burton (1993) examined a large set of potential omitted variables to gain insight into this underprediction. Several key variables were found, the inclusion of which reduced the underprediction to a large degree. Key variables included number of hours studying and the percentage of readings and other assignments completed. With SAT held constant, women studied more and completed more of the readings and other assignments. Without these variables in the model, variance in academic performance because of these variables is erroneously attributed to gender, resulting in differential prediction. Adding these other variables (e.g., study habits) resulted in a more fully specified equation in which the underprediction was reduced to nearly zero. This leads to the conclusion that the SAT itself, as a test, does not exhibit substantial bias against women, but the use of the SAT alone as the basis for selection would be biased because such a selection system would indeed underpredict women’s performance. If a composite of the SAT and the relevant other variables (e.g., study habits) were used as the basis for selection, the selection system would be unbiased.

Present Investigation

In the section above we illustrated the omitted variables problem with simulated data and by describing an empirical example from the education literature. In this section we used data from the U.S. Army’s Project A for an empirical illustration of the omitted variables problem in a personnel selection setting. Note that the setting in which Project A was conducted represents the scenario used in our examples above in that the Army has long been using the Armed Services Vocational Aptitude Battery (ASVAB) in screening potential recruits, and the Project A research considered a variety of possible supplements to the ASVAB, including several personality measures. Thus we illustrate the omitted variables problem by examining the personality variables for differential prediction by race, using both a fully specified model that includes the ASVAB and using a model that includes the personality variables alone.

Personality measures are often proposed as supplements to measures of cognitive ability, with the joint goals of increasing validity and reducing subgroup differences (Sackett, Schmitt, El-Din, & Kabin, 2001). In instances in which personality and cognitive ability predictors are both part of a selection system, the analyses presented above suggest that examining the personality measures for differential prediction by race using the traditional moderated regression method is subject to a potential omitted variables problem. The omission of a predictor in the cognitive domain has the potential to influence conclusions about differential prediction because cognitive ability is a strong predictor of a number of important performance constructs and because of the racial subgroup differences documented above. Note that we would not expect cognitive ability to produce an omitted variables problem in looking at differential prediction by gender in the use of personality measures (as in the case of Saad & Sackett, 2002) because of the generally small to nonexistent relationship between cognitive ability and gender.

Again, we emphasize that an omitted variable can influence conclusions about differential prediction when it is correlated with the criterion and with subgroup membership. As a result of the
commonly observed mean differences by race on cognitive ability measures, such measures will commonly be correlated with race when examining differential prediction by race. When cognitive ability is also correlated with the criterion of interest, the conditions where omitted variables are of concern are present. In the investigation described here, we examined three different types of criterion measures, which differ in the degree to which they are linked to cognitive ability. The degree to which omitting cognitive ability affects conclusions about differential prediction by race in using personality variables to predict performance was expected to vary across these criterion measures. As detailed below, we used data from the Army’s Project A, which contains measures of task proficiency criteria, correlating on average .45 with general cognitive ability; measures of effort and leadership criteria, correlating on average .22 with general cognitive ability; and measures of personal discipline criteria, correlating on average .11 with general cognitive ability (correlations are uncorrected; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990). Thus, we expected the effects of cognitive ability on conclusions about differential prediction to be greatest for task criteria, less for effort and leadership criteria, and least for personal discipline criteria.

Method

In this study, a moderated regression framework was used to test for the presence of differential prediction by race (White = 1, Black = 2) in an Army Project A database (N = 5,044) that contained predictor and criterion data for 13 separate jobs, known as military occupational specialties (MOS). More information about this database can be found in Campbell (1990) and Young, Houston, Harris, Hoffman, and Wise (1990).

Two inclusion criteria were used in this study. First, any job with too few individuals in one or both subgroups available for analysis was excluded. A minimum of 75 individuals per group was used as the criterion for inclusion, resulting in a total of 10 jobs to be examined. Table 5 provides sample sizes by subgroup for each MOS that was included in this study. Second, analyses were conducted only for predictor–criterion combinations that displayed significant relationships within each job (predictors and criteria used in this study are explained shortly). The logic here was that only if a test exhibits some evidence of criterion-related validity would the question arise of whether the test predicts the criterion in a similar manner for different subgroups. On the basis of these specifications, 79 predictor–criterion combinations were examined, which are listed in Tables 6, 7, and 8 for the three criterion constructs, respectively.

Measures

The omitted variable in this study was derived via principal components analysis of the 10 ASVAB subtests, considered by the Army to be a measure of aptitude. Scores on the first unrotated factor were obtained and labeled as the ASVAB general factor (Eysenck, 1988; Jensen, 1986). The Black–White d for the general factor was 1.05.

Three personality constructs obtained from the Assessment of Background and Life Experiences (ABLE; Peterson et al., 1990) were used in this study. These included adjustment (degree of emotional stability and stress tolerance), dependability (level of Conscientiousness and discipline), and surgency (degree of impact, influence, and energy). The Black–White ds for these measures were .03, .25, and .05, respectively.

Four job performance constructs were used in this study, two reflecting task proficiency (core technical proficiency and general soldiering proficiency), with these two hereafter referred to jointly as task proficiency), one reflecting effort and leadership, and one reflecting personal discipline. These constructs were originally developed on the basis of taxonomic work on the latent structure of job performance (Campbell, McHenry, & Wise, 1990). Core technical proficiency reflects the degree to which one can perform job-specific tasks, and general soldiering proficiency reflects the degree to which one can perform tasks common across all jobs in the military. Effort and leadership refers to the degree to which one leads and supports others and demonstrates effort, and personal discipline refers to the degree to which one follows rules and avoids counterproductive behaviors. Black–White ds for these performance measures were .49, .55, .19, and −.01, respectively. A fifth criterion construct from Project A, namely military bearing, was not included in this study due to a lack of a clear analog to the civilian workforce. See McHenry et al. (1990) for a complete correlation matrix among Project A measures.

Procedures

Traditional differential prediction analyses were conducted according to the Cleary (1968) model, in which the predictor (i.e., personality), subgroup (i.e., race), and the subgroup–predictor interaction were entered sequentially in the regression model. This was termed the omitted variables model because it treats the ASVAB general factor as the omitted variable. The results of these analyses were compared with what we term analyses for the full model. For these analyses, the same procedure was followed except that the ASVAB general factor was also included as a predictor at the first step of the regression to test whether its addition changed conclusions about the presence of differential prediction by slope or intercept as indicated by traditional methods.

Results

Tables 6, 7, and 8 present the results of the differential prediction analysis for the predictor–criterion combinations that were significant within job; each table presents results for a different criterion construct. In each table, the regression coefficients for the race variable and the race–predictor interaction variable are presented first for the analysis excluding the ASVAB general factor as a predictor (i.e., omitted variables model), then for the analysis that included the ASVAB general factor as a predictor (i.e., full model). The last two columns of the table indicate whether adding the ASVAB general factor to the regression equation changed conclusions about the presence of differential prediction.

We focus here first on the results for task proficiency in Table 6, specifically the analyses for the race variable. Recall that a
The effort and leadership criterion. When excluding the ASVAB general factor, 11 of the 26 race coefficients are significant; this number reduces to 3 coefficients when ASVAB is added to the model. Thus, including ASVAB in the model changes conclusions about differential prediction in 8 of the 11 instances in which intercept differences are found using the restricted model. The results for the personal discipline criterion in Table 8 show that including the ASVAB general factor does little to change conclusions about differential prediction; 2 of 21 cases show differential prediction by race initially; this number actually increases to 3 cases when ASVAB is added. In general, the results are very different for the interaction term. Recall that a significant coefficient for the interaction term is interpreted as indicating slope differences by subgroup. When excluding the ASVAB general factor, traditional differential prediction analysis for task proficiency indicates that differential prediction is present in only 1 of 32 cases. When the ASVAB general factor is included in the analysis, conclusions change for that case because the differential prediction disappears. Thus, the exclusion of the ASVAB general factor alters the conclusion that slope differences by subgroup

---

Table 6

Differential Prediction Results: Technical Proficiency Criteria

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Predictor</th>
<th>MOS</th>
<th>Excluding ASVAB</th>
<th>Including ASVAB</th>
<th>ΔRace?*</th>
<th>ΔInteraction?*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Race B</td>
<td>Interaction B</td>
<td>Race B</td>
<td>Interaction B</td>
</tr>
<tr>
<td>CTP</td>
<td>ADJ</td>
<td>54E</td>
<td>5.71**</td>
<td>−0.22</td>
<td>1.64</td>
<td>−0.22</td>
</tr>
<tr>
<td>CTP</td>
<td>ADJ</td>
<td>63B</td>
<td>12.50*</td>
<td>0.35</td>
<td>7.14**</td>
<td>0.27</td>
</tr>
<tr>
<td>CTP</td>
<td>ADJ</td>
<td>71L</td>
<td>5.98**</td>
<td>0.07</td>
<td>0.49</td>
<td>0.02</td>
</tr>
<tr>
<td>CTP</td>
<td>ADJ</td>
<td>76W</td>
<td>7.45**</td>
<td>0.14</td>
<td>1.56</td>
<td>0.08</td>
</tr>
<tr>
<td>CTP</td>
<td>ADJ</td>
<td>91A</td>
<td>7.91**</td>
<td>−0.13</td>
<td>1.51</td>
<td>−0.14</td>
</tr>
<tr>
<td>CTP</td>
<td>DEP</td>
<td>54E</td>
<td>6.12**</td>
<td>−0.02</td>
<td>1.73</td>
<td>−0.08</td>
</tr>
<tr>
<td>CTP</td>
<td>DEP</td>
<td>71L</td>
<td>6.14**</td>
<td>0.06</td>
<td>0.69</td>
<td>0.01</td>
</tr>
<tr>
<td>CTP</td>
<td>DEP</td>
<td>76Y</td>
<td>5.21**</td>
<td>0.10</td>
<td>0.54</td>
<td>0.08</td>
</tr>
<tr>
<td>CTP</td>
<td>DEP</td>
<td>91A</td>
<td>8.33**</td>
<td>−0.19</td>
<td>1.90</td>
<td>−0.24</td>
</tr>
<tr>
<td>CTP</td>
<td>DEP</td>
<td>94B</td>
<td>5.93**</td>
<td>0.05</td>
<td>1.46</td>
<td>0.04</td>
</tr>
<tr>
<td>CTP</td>
<td>SUR</td>
<td>54E</td>
<td>5.72**</td>
<td>0.04</td>
<td>1.67</td>
<td>0.04</td>
</tr>
<tr>
<td>CTP</td>
<td>SUR</td>
<td>71L</td>
<td>6.25**</td>
<td>0.03</td>
<td>0.84</td>
<td>0.01</td>
</tr>
<tr>
<td>CTP</td>
<td>SUR</td>
<td>76W</td>
<td>7.27**</td>
<td>−0.02</td>
<td>1.50</td>
<td>−0.03</td>
</tr>
<tr>
<td>CTP</td>
<td>SUR</td>
<td>76Y</td>
<td>4.83**</td>
<td>0.04</td>
<td>0.27</td>
<td>0.01</td>
</tr>
<tr>
<td>CTP</td>
<td>SUR</td>
<td>94B</td>
<td>5.68**</td>
<td>0.04</td>
<td>1.23</td>
<td>0.03</td>
</tr>
<tr>
<td>GSP</td>
<td>ADJ</td>
<td>31C</td>
<td>9.81**</td>
<td>0.07</td>
<td>3.77</td>
<td>0.09</td>
</tr>
<tr>
<td>GSP</td>
<td>ADJ</td>
<td>54E</td>
<td>6.77**</td>
<td>−0.07</td>
<td>2.71*</td>
<td>−0.08</td>
</tr>
<tr>
<td>GSP</td>
<td>ADJ</td>
<td>55B</td>
<td>3.58**</td>
<td>0.07</td>
<td>1.33</td>
<td>−0.01</td>
</tr>
<tr>
<td>GSP</td>
<td>ADJ</td>
<td>63B</td>
<td>6.52**</td>
<td>0.14</td>
<td>0.59</td>
<td>0.09</td>
</tr>
<tr>
<td>GSP</td>
<td>ADJ</td>
<td>71L</td>
<td>8.16**</td>
<td>0.30</td>
<td>2.46</td>
<td>0.24</td>
</tr>
<tr>
<td>GSP</td>
<td>ADJ</td>
<td>76W</td>
<td>7.46**</td>
<td>0.19*</td>
<td>3.31**</td>
<td>0.15</td>
</tr>
<tr>
<td>GSP</td>
<td>ADJ</td>
<td>76Y</td>
<td>6.73**</td>
<td>−0.09</td>
<td>2.74**</td>
<td>−0.05</td>
</tr>
<tr>
<td>GSP</td>
<td>ADJ</td>
<td>91A</td>
<td>9.82**</td>
<td>−0.25</td>
<td>3.42</td>
<td>−0.26</td>
</tr>
<tr>
<td>GSP</td>
<td>DEP</td>
<td>63B</td>
<td>7.56**</td>
<td>−0.19</td>
<td>1.32</td>
<td>−0.17</td>
</tr>
<tr>
<td>GSP</td>
<td>DEP</td>
<td>71L</td>
<td>8.20**</td>
<td>0.13</td>
<td>2.40</td>
<td>0.07</td>
</tr>
<tr>
<td>GSP</td>
<td>DEP</td>
<td>76Y</td>
<td>6.89*</td>
<td>−0.03</td>
<td>2.78**</td>
<td>−0.05</td>
</tr>
<tr>
<td>GSP</td>
<td>DEP</td>
<td>91A</td>
<td>10.15**</td>
<td>−0.05</td>
<td>3.50</td>
<td>−0.10</td>
</tr>
<tr>
<td>GSP</td>
<td>SUR</td>
<td>54E</td>
<td>6.77**</td>
<td>0.01</td>
<td>2.68*</td>
<td>0.09</td>
</tr>
<tr>
<td>GSP</td>
<td>SUR</td>
<td>71L</td>
<td>8.29**</td>
<td>0.09</td>
<td>2.58</td>
<td>0.06</td>
</tr>
<tr>
<td>GSP</td>
<td>SUR</td>
<td>76W</td>
<td>7.27**</td>
<td>−0.02</td>
<td>3.42**</td>
<td>−0.02</td>
</tr>
<tr>
<td>GSP</td>
<td>SUR</td>
<td>76Y</td>
<td>6.43**</td>
<td>−0.03</td>
<td>2.57*</td>
<td>−0.05</td>
</tr>
<tr>
<td>GSP</td>
<td>SUR</td>
<td>91A</td>
<td>9.71**</td>
<td>−0.03</td>
<td>3.19</td>
<td>−0.41</td>
</tr>
</tbody>
</table>

Note. White = 1, Black = 2. MOS = military occupational specialties; ASVAB = Armed Services Vocational Aptitude Battery; CTP = core technical proficiency; ADJ = adjustment; DEP = dependability; SUR = surgery; GSP = general soldiering proficiency.

* Yes = coefficient changed from significant to nonsignificant when ASVAB was added; No = coefficient remained significant when ASVAB was added. Dashes indicate that coefficient remained nonsignificant when ASVAB was added.

*p < .05. **p < .01.

---

significant race coefficient is interpreted as indicating that intercept differences by subgroup are present. For the omitted variables model excluding the ASVAB general factor, every race coefficient in Table 6 is significant, indicating the presence of differential prediction in all 32 cases. Coefficients are positive, indicating that the performance of the minority group is overpredicted. However, when the ASVAB general factor is included in the differential prediction analysis, the conclusions change drastically. Namely, after adding the ASVAB general factor, differential prediction by intercept is present in only 8 of 32 cases. Even for the 8 cases in which the race term remains significant, the size of the race term is markedly reduced when ASVAB is added to the model. Thus, the exclusion of the ASVAB general factor results in the conclusion that intercept differences by subgroup are consistently found (i.e., intercept differences in 100% of cases), whereas inclusion of the ASVAB general factor leads to a very different conclusion (i.e., intercept differences in 25% of cases).

A similar, yet more moderate pattern is found in Table 7 for the effort and leadership criterion. When excluding the ASVAB general factor, 11 of the 26 race coefficients are significant; this number reduces to 3 coefficients when ASVAB is added to the model. Thus, including ASVAB in the model changes conclusions about differential prediction in 8 of the 11 instances in which intercept differences are found using the restricted model. The results for the personal discipline criterion in Table 8 show that including the ASVAB general factor does little to change conclusions about differential prediction; 2 of 21 cases show differential prediction by race initially; this number actually increases to 3 cases when the omitted variable is added. In general, the results are very different for the interaction term. Recall that a significant coefficient for the interaction term is interpreted as indicating slope differences by subgroup. When excluding the ASVAB general factor, traditional differential prediction analysis for task proficiency indicates that differential prediction is present in only 1 of 32 cases. When the ASVAB general factor is included in the analysis, conclusions change for that case because the differential prediction disappears. Thus, the exclusion of the ASVAB general factor alters the conclusion that slope differences by subgroup...
exist in only one instance. We note that statistical power is a concern with tests of the interaction term in regression analyses (Aguinis & Stone-Romero, 1997), and thus it is possible that some interaction effects are not detected. Nonetheless, it is informative to see that the rate of detecting slope changes when a previously omitted variable is added is very different from the rate of detecting intercept changes. We did conduct a power analysis by using the approach of Aguinis, Boik, and Pierce (2001), which frames moderator effects in terms of differences in the predictor–criterion correlations, predictor and criterion standard deviations, and predictor and criterion reliabilities for the two subgroups being compared. We estimate power to be approximately .14, .46, and .82 to detect differences of .10, .20, and .30 between the subgroup correlations. Thus, there is high power to detect large differences, but low power to detect small differences. The important point, though, is that the rate of detecting significant interactions is not meaningfully different between the analysis with and without the ASVAB general factor as an omitted variable. If the true state of affairs is that personality consistently interacts with race but that the omission of the ASVAB general factor from the model disguises this fact, then even with power as low as .14 to detect a small effect, one would expect the number of significant interaction terms to increase when adding the ASVAB general factor to the model (i.e., if a true effect is detected 14% of the time and 32 tests are conducted, one would expect 4–5 significant interactions; in fact, when ASVAB is added to the model there are no significant interactions). It remains true, of course, that we cannot be confident of the conclusions for any single interaction given low power, which is one key reason for using the strategy adopted here of basing conclusions on overall rates of occurrence over multiple samples.

Results are similar for effort and leadership and personal discipline. When excluding the ASVAB general factor, slope differences are present in 4 of 26 cases for effort and leadership and 2 of 21 cases for personal discipline. Adding the ASVAB general factor to the differential prediction analysis results in slope differences being found in only 3 cases for effort and leadership and 1 case for personal discipline.

Discussion

This study illustrates how the inclusion of a previously omitted variable can change the results of differential prediction analysis. In 24 of 32 cases for task proficiency, a different conclusion was reached regarding differential prediction when a fully specified model (i.e., a model including the ASVAB general factor) was used. For the omitted variables model (excluding the ASVAB general factor), the coefficient for race was significant in all 32
instances, suggesting intercept differences between the groups for each predictor–criterion combination. However, when the ASVAB general factor was added to the regression analysis, the coefficient for race was significant in only 8 of 32 cases. Even in these 8 cases the coefficients were considerably smaller when the ASVAB general factor was added to the model.

We hypothesized that the omitted variables effect would be greatest for task proficiency due to the strong relationship between ability and task proficiency. Consistent with this hypothesis, we found that including ability in the model changes conclusions less frequently for effort and leadership than for task proficiency and least of all for personal discipline. Thus in addition to illustrating the potential for omitted variables to influence differential prediction findings, we also illustrate that concerns about omitted variables are specific to the domain in question.

At this point it is useful to consider two possible conceptualizations of the concept of a fully specified model. The conceptualization used in this article and illustrated with the Project A data is to view a fully specified model as a model including all variables included in a selection system. The U.S. military has long used ASVAB, a measure correlated with both performance and with race, in screening recruits, and thus failure to include this measure in an investigation of the use of personality measures could result in an omitted variable problem that could bias the results of differential prediction analysis and thus bias the conclusion one would draw about the applied question of whether adding personality to the selection system would result in differential prediction.

A second conceptualization would be to view a fully specified model not as a model restricted to currently measured predictors but rather as including all possible determinants of the criterion that are also correlated with subgroup membership. Imagine, for instance, that the U.S. military did not have a history of using ASVAB and instead was starting from scratch in considering the use of a personality measure. In that case, a model containing only the personality measure would be a fully specified model under the first conceptualization above. Under the second conceptualization, though, one might hypothesize that cognitive ability would meet the omitted variable criterion (i.e., correlated with both race and the criterion), and include it in differential prediction analysis. The purpose for doing so is very different than under the first conceptualization. In the first conceptualization, the question to be answered is whether the use of the personality measure in conjunction with the ability measure already used in the selection process would result in predictive bias; failure to include ability in the model makes it more likely that one will draw the wrong conclusion.

In the second conceptualization, however, the issue of omitted variables is irrelevant to the applied question of predictive bias: Personality is the only proposed predictor, and differential prediction analysis using only the personality predictor does answer the question of whether differential prediction is present. The issue of omitted variables comes into play if one does find evidence of predictive bias. On finding predictive bias in using the personality measure to predict the criterion of interest, the researcher is unsure of the cause of the bias. One possibility is that the personality

<table>
<thead>
<tr>
<th>Predictor</th>
<th>MOS</th>
<th>Race B</th>
<th>Interaction</th>
<th>Including ASVAB</th>
<th>ΔRace</th>
<th>ΔInteraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ 63B</td>
<td>-0.36*</td>
<td>-0.39*</td>
<td>-4.22*</td>
<td>-0.42*</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>ADJ 94B</td>
<td>-0.65</td>
<td>0.29</td>
<td>-1.16</td>
<td>0.02</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>DEP 31C</td>
<td>0.07</td>
<td>0.05</td>
<td>-1.57</td>
<td>0.05</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>DEP 54B</td>
<td>1.41</td>
<td>-0.09</td>
<td>-0.95</td>
<td>-0.09</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>DEP 55B</td>
<td>5.65*</td>
<td>-0.19</td>
<td>6.66**</td>
<td>-0.19</td>
<td>No</td>
<td>——</td>
</tr>
<tr>
<td>DEP 63B</td>
<td>-1.58</td>
<td>-0.24</td>
<td>-2.60</td>
<td>-0.23</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>DEP 64C</td>
<td>-1.35</td>
<td>0.05</td>
<td>-1.57</td>
<td>0.05</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>DEP 71L</td>
<td>-0.03</td>
<td>0.21*</td>
<td>-2.11</td>
<td>0.19</td>
<td>Yes</td>
<td>——</td>
</tr>
<tr>
<td>DEP 76W</td>
<td>2.24</td>
<td>0.14</td>
<td>-0.27</td>
<td>0.14</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>DEP 76Y</td>
<td>0.46</td>
<td>-0.02</td>
<td>-1.46</td>
<td>-0.03</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>DEP 91A</td>
<td>1.81</td>
<td>0.24</td>
<td>0.93</td>
<td>0.24</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>DEP 94B</td>
<td>0.35</td>
<td>-0.04</td>
<td>0.60</td>
<td>-0.04</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>SUR 31C</td>
<td>-0.49</td>
<td>-0.04</td>
<td>-1.24</td>
<td>-0.05</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>SUR 54E</td>
<td>-0.90</td>
<td>0.02</td>
<td>-3.40</td>
<td>0.02</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>SUR 63B</td>
<td>-3.39</td>
<td>0.01</td>
<td>-4.29*</td>
<td>0.01</td>
<td>Yesb</td>
<td>——</td>
</tr>
<tr>
<td>SUR 64C</td>
<td>-2.28</td>
<td>-0.02</td>
<td>-2.40</td>
<td>-0.02</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>SUR 71L</td>
<td>0.08</td>
<td>0.08</td>
<td>-1.94</td>
<td>0.08</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>SUR 76W</td>
<td>0.87</td>
<td>0.08</td>
<td>-1.46</td>
<td>0.07</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>SUR 76Y</td>
<td>-0.59</td>
<td>-0.09</td>
<td>-2.34</td>
<td>-0.10</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>SUR 91A</td>
<td>2.26</td>
<td>0.11</td>
<td>0.40</td>
<td>0.10</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>SUR 94B</td>
<td>-0.89</td>
<td>0.04</td>
<td>-0.64</td>
<td>0.04</td>
<td>——</td>
<td>——</td>
</tr>
</tbody>
</table>

Note. White = 1, Black = 2. MOS = military occupational specialties; ASVAB = Armed Services Vocational Battery; ADJ = adjustment; DEP = dependability; SUR = surgency.

*a* Yes = coefficient changed from significant to nonsignificant when ASVAB was added; No = coefficient remained significant when ASVAB was added. Dashes indicate that coefficient remained nonsignificant when ASVAB was added.

*b* Coefficient changed from nonsignificant to significant when ASVAB was added.

*p < .05. **p < .01.
measure itself is biased, and thus there is a need to modify or replace it. Another possibility is that the personality measure is perfectly fine as a measure of its intended construct, and an omitted variable is the cause of the differential prediction. If that is the case, the remedy is to add the omitted variable to the selection system if eliminating differential prediction is the goal. In short, the consideration of omitted variables in the second conceptualization is of value in shedding light on whether the predictor in question is the source of the predictive bias.

One concern about this second conceptualization is that finding an omitted variable, the inclusion of which changes the conclusions about differential prediction, would not be of real value because there may be yet other omitted variables, the inclusion of which might change conclusions yet again. This is a perplexing question because there is no unambiguous way of determining whether all omitted variables have been identified. The best one can do is to be knowledgeable about the research in the domain in question such that one can give a considered answer to the question, Are there predictor constructs that are simultaneously correlated with the criterion and correlated with the subgroup variable under consideration? If so, such variables merit consideration in conducting and interpreting differential prediction analyses.

We wish to make clear that analyses conducted under this second conceptualization of a fully specified model are conducted for research purposes with the goal of understanding the source of the differential prediction. Nothing in this article should be interpreted as suggesting a need for an applied researcher to search for previously excluded predictors on the grounds that their inclusion might change the results of differential prediction analyses. In applied selection settings, the only variables relevant to drawing a conclusion about the presence or absence of differential prediction are the predictor variables that are included in the selection system. Similarly, nothing in this article should be interpreted as critical of the Cleary (1968) formulation of predictive bias or the well-established moderated regression approach to testing for differential prediction. We illustrated the potential problems of testing predictors individually for differential prediction when multiple predictors are, in fact, to be used in combination. If selection is to be conducted on the basis of a composite of predictors, testing for differential prediction using the composite is the appropriate course of action. Although we introduced the possibility of examining previously omitted variables in conjunction with a predictor of interest as a means of shedding light on the causes of a finding of differential prediction, such examinations are of theoretical interest and do not speak to the issue of the presence or absence of differential prediction with the currently used predictors.

Extension to Settings Where the Focal Predictor and the Omitted Variable Are Correlated

When considering the issues discussed in this article in domains other than personality, one encounters issues of potential correlation between measured and omitted variables and questions of where the shared variance between them is to be assigned. Consider, for example, a researcher interested in differential prediction by race when using a particular type of interview, with the intent being to use the interview in conjunction with an ability measure. Assume that in this example there is a positive correlation between scores in the interview and scores on the ability measure. The discussion in this article suggests that omitting ability from the analysis could bias the results, and that there is thus a need to consider the interview and the ability measure in the differential prediction model.

Perhaps the most crucial point is that the correlation between the interview and the ability measures is not an issue for estimating the subgroup coefficient used to identify intercept differences. The predictors are entered at Step 1 of differential prediction analysis, with the subgroup variable (e.g., race) entered at a subsequent step. Thus, there is no need to sort out the partitioning of variance between the predictors to test for intercept differences. In addition, intercept differences are the dominant issue in differential prediction analyses, with slope differences rarely found.

There is also a ready solution to the issue of shared variance in settings where the intent is to use both variables operationally in a selection setting. In the current example, if the intent is to use both the interview and the ability measure, one can create a composite of the two and conduct differential prediction analysis on the composite. For answering the operational question of whether a proposed multipredictor selection system is predictively biased, there is no need to attempt to partition variance between the predictors. In fact, it is the analysis of each predictor individually, rather than jointly, that produces the omitted variables problem.

The issue of shared variance between predictors does, however, pose an interesting problem in testing for slope differences in the setting where one is proposing to use a single focal predictor, and one is interested in determining whether findings from differential prediction analysis with that predictor are influenced by a hypothesized omitted variable that is correlated with the focal predictor. If one were to include the focal predictor and a previously omitted variab at Step 1, subgroup at Step 2, and the focal predictor–subgroup interaction at Step 3, one is in essence assigning all shared variance between the focal predictor and the previously omitted variable to the omitted variable when conducting the Step 3 analysis.

Our recommendation in this instance involves creating a composite of the focal predictor and the previously omitted variable. One would compare the results of several differential prediction analyses. First, use both the focal predictor and the previously omitted variable in differential prediction analysis, creating the interaction term using the focal predictor (i.e., focal predictor and omitted variable at Step 1, then subgroup, then the focal predictor by subgroup interaction). In this analysis, only the variance unique to the focal predictor (i.e., not shared with the omitted variable) drives the interaction. Then do the converse: Conduct a second differential prediction analysis using the omitted variable to create the interaction term. In this analysis, only the variance unique to the omitted predictor (i.e., not shared with the focal predictor) drives the interaction. Finally, create a composite of the focal predictor and the previously omitted predictor, and use this composite in differential prediction analysis. In this analysis, shared variance is included in the interaction. By comparing the results of these analyses, one can differentiate between interaction effects because of the unique variance in the focal predictor, because of the unique variance in the previously omitted predictor, and because of shared variance.
The Omitted Variables Problem in the Literature

Given the potential for omitted variables to influence conclusions about differential prediction and given that this issue has been raised as early as 1971, one might think that omitted variables would be routinely considered in applications of differential prediction analysis. This, however, is not true. As a way of obtaining a rough gauge of degree to which attention was paid to the omitted variables problem, we drew a sample of 33 applications of differential prediction analysis by race or gender from the published literature in education and psychology, based on an electronic search for articles up to the year 2000 containing the terms differential prediction or predictive bias in the titles. This sampling included articles published in 22 journals, with the two most frequent being Personnel Psychology and Journal of Educational Measurement. Although the omitted variables issue was mentioned in 6 of the 33 studies, only 2 of the 33 studies actually included variables beyond test, subgroup and the test–subgroup interaction in their analysis. Thus attention to this issue is a rare exception rather than the norm.

We offer two examples from the literature. First, te Nijenhuis and van der Flier (2000) examined differential prediction on the basis of native versus immigrant status in a sample of applicants for truck driver positions in the Netherlands. The selection system included a general cognitive ability measure, and measures of the personality traits of Extraversion and Neuroticism. They examined each of these three predictors separately, and reported finding differential prediction for the two personality measures for one of the study’s criterion variables. In contrast, Gael, Grant, and Ritchie (1975) examined differential prediction by race in using a battery of cognitive and perceptual measures to predict the performance of telephone operators. Preliminary work identified four tests from a larger battery as candidates for a final selection system. Differential prediction analysis by race was then conducted on the four-test composite. Thus, these studies illustrate differences in the literature in how predictors are treated when differential prediction analyses are performed in the setting in which multiple predictors are being used in a selection system.

Implications for Research and Practice

These findings have considerable implications for research and practice. We return to the scenarios introduced at the outset of the article. First, a researcher has assembled a trial test battery for use in a criterion-related validity study and is selecting a smaller number of predictors for inclusion in an operational battery. The researcher is concerned about predictive bias by race and gender and decides to examine each predictor separately for predictive bias. The conceptual framework and empirical results of the present article suggest that such a strategy would be in error: Omitted variables can lead to erroneous conclusions if each predictor is examined in isolation. Instead, the researcher should examine predictive bias for any predictor composite being considered for operational use.

Second, consider the scenario in which an organization is considering adding a new predictor to an existing selection system. A decision rule is proposed specifying that the new predictor only be added if an analysis reveals no evidence of predictive bias. Again, examining the new predictor in isolation would be a mistake because other variables that are part of the selection process might function as omitted variables in such an analysis. The empirical portion of the present article illustrates this: The U.S. Army has long used the ASVAB in its selection system, and the Project A research examined new predictors as possible supplements to the existing system. Thus, including ASVAB in models examining whether proposed personality predictors produce differential prediction is a step toward avoiding an omitted variables problem.

Third, consider the researcher attempting to meta-analytically cumulate findings from examinations of slope and intercept differences by subgroup for a particular test type or construct (Raju, Fralix, & Steinhaus, 1986). Absent the perspective developed in this article, there would be a tendency to interpret cumulative evidence of slope or intercept differences as offering a general statement about bias in the predictor under examination. In the context of the omitted variables problem, any interpretations of such findings would require careful attention of whether the predictor in question was examined in isolation or whether other predictors were also included in the model, and any conclusions would need to carefully specify whether one was discussing bias in the context of the predictor used alone or as part of a broader selection system.

We note that the effect of an omitted variables problem in the settings examined in this article is to produce differential prediction that results in the overprediction of minority performance. Some may take the perspective that differential prediction is a practical concern only if minority performance is underpredicted, and thus there is no reason to consider the omitted variables issue. We offer a number of responses. First, we note that there is no consensus on the issue of when differential prediction is a concern; the competing position is that differential prediction is undesirable, regardless of whether it is the majority or the minority group that would be adversely affected by the use of a common regression line. Second, although the data in this article show differential prediction resulting in overprediction, we encountered data that we cannot share here in which the finding was underprediction of minority group performance. In those data, the minority group in question (Asian Americans) had a higher ability mean than the majority sample, in contrast to the direction of the majority–minority difference in the present article. The result was that the use of personality alone underpredicted minority performance, with the underprediction eliminated when ability was included in the model. Third, we also feel that the article has value in addressing potential confusion on the part of some who would posit that because personality measures produce little or no subgroup mean differences, one would subsequently expect no differential prediction. This article clarifies that the presence or absence of predictor subgroup differences has no necessary implications for the presence or absence of differential prediction.

Narrowing the Criterion as a Route to Minimizing the Omitted Variables Problem

Kehoe (2000) noted that the lack of direct correspondence between the constructs underlying the predictor and the criterion is at the heart of the issue of reaching misleading conclusions from
differential prediction analyses. The omitted variables problem is linked to the use of multidimensional criteria: If a criterion measured a single unidimensional construct and a predictor reliably measured that same construct, there would be no possibility of an omitted variables problem. This is also the central idea behind Terris’s (1997) critique of differential prediction methodology. He frames it as an inquiry into whether the predictor and criterion load equally on a common factor.

Thus, an alternative to seeking a fully specified model in the prediction of a complex criterion is to seek a narrower, focused criterion that focuses exclusively on the workplace manifestation of the construct tapped by the predictor. If this could be accomplished, the omitted variables problem would be eliminated, simplifying the conduct and interpretation of differential prediction analysis. The operational feasibility of such an approach is unclear because even criteria in contemporary research that are much narrower than overall performance measures tend to be multidimensional. The Project A data used in this research, for example, contained separate measures in the task proficiency domain and in the effort and leadership domain, although, both ability and personality predictors are related to both of these criteria. However, it should logically be the case that the narrower the criterion, the less likely it is that omitted variables are of concern.

This approach might at first glance appear to run contrary to the generally acknowledged strategy that one should start by identifying the criterion of interest and then select predictors hypothesized to be related to this criterion. It would appear backward to start with a predictor and then seek to identify criteria linked closely to that predictor. In fact, however, the process may be one in which one does begin by identifying the criterion construct of interest (e.g., detail orientation and perseverance on the job), then turning to selecting a predictor with a close conceptual link to this criterion (e.g., a Conscientiousness measure), and finally identifying an operational criterion measure for use in a validation study. This operational criterion may end up broader than the original intent (e.g., an overall measure of performance rather than one focused on detail orientation and perseverance). The present article advises against such slippage in criterion measurement.

Kehoe (2000) offered the proposition that differential prediction analyses be conducted only with criteria that are intended to be theoretically relevant to the predictor. Overall criterion measures would be appropriate only for prediction composites intended to reflect the full range of constructs represented by the criterion measure. His perspective reflects the assumption that differential prediction is an assessment of bias based on the intended meaning of the predictor constructs, and he concluded that evidence of differential prediction of a construct irrelevant criterion is not an indication of bias in the development of the predictor. We believe it is useful to add the observation that following Kehoe’s advice and using narrower criteria when conducting differential prediction analysis does not preclude the use of overall performance criteria for other purposes in validation research, such as documenting the relationship between a predictor or set of predictors and overall performance. For purposes of understanding predictor–criterion relationships, the careful matching of predictors and criteria is quite useful and informative. For broader applied selection system design concerns, however, the use of overall performance criteria are certainly of central importance.

Conclusion

In conclusion, this article calls attention to the omitted variables problem in studying differential prediction and illustrates how the use of a more fully specified model can change conclusions about the presence and causes of predictive bias. We note that the specific issue illustrated in this article, namely, ability as an omitted variable in validation work using personality measures, is but one instance of the more general issue of omitted variables bias, and thus potential omitted variables are worthy of consideration when conducting differential prediction research in other domains as well.

References


Received April 5, 2002
Revision received March 19, 2003
Accepted March 27, 2003

---

**Call for Nominations**

The Publications and Communications (P&C) Board has opened nominations for the editorships of *Comparative Psychology, Experimental and Clinical Psychopharmacology, Journal of Abnormal Psychology, Journal of Counseling Psychology*, and *JEP: Human Perception and Performance* for the years 2006–2011. Meredith J. West, PhD, Warren K. Bickel, PhD, Timothy B. Baker, PhD, Joloda C. Hansen, PhD, and David A. Rosenbaum, PhD, respectively, are the incumbent editors.

Candidates should be members of APA and should be available to start receiving manuscripts in early 2005 to prepare for issues published in 2006. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. Self-nominations also are encouraged.

Search chairs have been appointed as follows:

- **Comparative Psychology**, Joseph J. Campos, PhD
- **Experimental and Clinical Psychopharmacology**, Linda P. Spear, PhD
- **Journal of Abnormal Psychology**, Mark Appelbaum, PhD, and David C. Funder, PhD
- **Journal of Counseling Psychology**, Susan H. McDaniel, PhD, and William C. Howell, PhD
- **JEP: Human Perception and Performance**, Randi C. Martin, PhD

To nominate candidates, prepare a statement of one page or less in support of each candidate. Address all nominations to the appropriate search committee at the following address:

Karen Sellman, P&C Board Search Liaison
Room 2004
American Psychological Association
750 First Street, NE
Washington, DC 20002-4242

The first review of nominations will begin December 8, 2003. The deadline for accepting nominations is **December 15, 2003**.