Correcting the Effect Size of *d* for Range Restriction and Unreliability

PHILIP BOBKO
Gettysburg College

PHILIP L. ROTH
Clemson University

CHRISTOPHER BOBKO
Princeton University

Pearson product moment correlations are often corrected for statistical artifacts such as range restriction and unreliability. Formulas have long existed to make such corrections. However, other effect size estimates are rarely corrected for these artifacts, in spite of the fact that there is an established mathematical link between the correlation and some effect size estimates. Correlations and other effect sizes are therefore vulnerable to the same artifacts. The authors take a common effect size estimate, the standardized mean difference between two groups, and derive (and reaffirm in one instance) correction formulas suitable for use with this statistic. It is demonstrated how these formulas might substantially increase the precision of estimates and decisions made within organizational research and practice, whenever correction factors can be appropriately estimated.

An Example

In addition to the investigation of validity and other test attributes (e.g., cost), suppose a human resource manager is interested in finding a selection test that has less adverse impact than a traditional measure of cognitive ability. The manager knows/ assumes that if a selection test of cognitive ability is used, then Whites will score, on average, about 1.0 standard deviations higher than Blacks on the test (Sackett & Wilk, 1994). The manager sponsors research on recent hires (incumbents selected on the basis of a cognitive ability test) and finds that the use of a biodata measure leads to a smaller difference, say, .3 of a standard deviation, between Whites and Blacks. The manager concludes that the biodata measure has substantially less adverse impact potential for entry-level selection than the cognitive ability test. Is this conclusion accurate?

Authors' Note: A portion of this work was accomplished while the first author was at Syracuse University. We wish to thank the editor and two anonymous reviewers for their thoughtful assistance on an earlier version of this manuscript. Correspondence concerning this article should be addressed to Philip Bobko, Department of Management, Gettysburg College, Gettysburg, PA 17325.

Although some individuals have reached similar conclusions in the face of such data, the actual answer to the above question is, in general, "not necessarily," since the values of 1.0 and .30 are unlikely to be comparable. That is, the value of 1.0 assumed above is generally computed in large national samples of individuals before any explicit selection takes place (Sackett & Ostgaard, 1994). However, the value of .30 for biodata is typically computed on preselected or incumbent samples (Reilly & Warech, 1994). Thus, the effect size of .30 is more "range restricted" than the value of 1.0. If both predictors are candidates for use at the initial stage of selection, then failure to correct for the differential range restriction across the measures makes the values of .30 and 1.0 difficult to compare.

In a later section of this article, we also reaffirm that comparing values of effect sizes can be problematic because of differential reliability. For example, we note that measures of cognitive ability typically have different (generally larger) reliabilities than alternative selection measures such as personality indexes, interviews, situational judgment tests, and so on. These additional corrections for differences in unreliability are therefore important when the focus is on comparisons of latent, population values.

Thus, range restriction and measurement unreliability are important issues in organizational research, as we illustrate throughout this article. Our purpose is to derive correction formulas for a common effect size estimate, apply the formulas using examples (including personnel selection, training, marketing, and risk taking), and generate an awareness on the part of researchers to correct effect sizes for range restriction and attenuation. Effect size estimates will then be relatively more comparable, thereby leading to more accurate organizational decisions when such corrections are appropriate and/or justifiable.

Although the above issues are known to some statisticians, the rarity with which these corrections are made in published work is surprising. We have read more than 2,000 manuscripts in applied psychology and management in the past half decade. We have not seen these corrections applied except as noted below. In addition, comparisons of differentially range restricted estimators appear in the literature. For example, two recent meta-analyses of the selection literature (Bobko, Roth, & Potosky, 1999; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997) directly compare effect sizes across different types of predictors even though these estimates are differentially affected by range restriction. Other examples of such practices are documented below.

The comparison practices in the previous paragraphs, and many that follow, often involve work with covariances in which one variable is dichotomous. For example, as expanded on below, suppose a researcher studies the difference between Black and White (the dichotomous variable) scores on two potential selection tests under consideration, say, cognitive ability and biodata. Suppose that the study is conducted on incumbents (i.e., it is a concurrent study) and the incumbents have been selected using some screen on grade point average (GPA). Differential range restriction will occur on the two potential predictors to the extent that they are differentially related to GPA; hence, ethnic group effect size estimates will be differentially affected. As a second example, suppose a researcher studies the relation between risk taking and entrepreneurial status (manager vs. entrepreneur). In this case, some firms will not have survived at the time of measurement. To the extent that risk taking covaries with survival, then range restriction on risk taking occurs. Finally, it is known that meta-analytic studies can also be affected by these factors; ignoring them could lead to less than accurate results.

Dichotomous Variables and Effect Sizes (*d*)

The investigation of behavioral phenomena in a variety of fields focuses on the influence of a categorical independent variable on a continuous dependent variable. Such questions have a long history in management, psychology, education, and sociology. For example, Galton (1892) theorized about the influence of race on cognitive ability. More recent examples in the business disciplines include the investigation of a training program's influence (vs. a control group) on job performance (e.g., Russell, Wexley, & Hunter, 1984) and the influence of qualitatively different parental styles on a variety of dependent variables in marketing (e.g., Carlson & Grossbart, 1988).

An important aspect of the analysis of a categorical variable (Black/White, trained/untrained, etc.) involves not only calculating the significance of the effect but understanding the *magnitude* of the effect. The t statistics used in significance tests in these areas do not tell researchers how "far apart" the relevant groups are or how large the effect of some independent variable is. In contrast, the standardized group difference statistic (d) is often used as a measure of effect size (e.g., Hunter & Schmidt, 1990). This statistic is defined as the difference in means of the two groups divided by the pooled standard deviation. That is,

$$d = \frac{\overline{X}_1 - \overline{X}_2}{S_{pooled\ within}}$$

For example, consistent with the introductory example, a d of 1.00 for cognitive ability and race (White/Black) would suggest that Whites and Blacks differ on average by one standard deviation (Hunter & Hunter, 1984; Sackett & Wilk, 1994).

Need for Corrections

The value of d is a straightforward transformation of the correlation (r) between a dichotomous variable and a continuous one (see Equation [2] below). It is well known that the magnitude of r is affected by range restriction and unreliability (Ghiselli, 1964; Hunter & Schmidt, 1990). These artifacts therefore also affect values of d. Comparing uncorrected values of d can lead to inaccurate conclusions in management and other social science disciplines if the samples are not comparable. As noted earlier, although values of r are often corrected for artifacts, the use of corrected values of d is rare, even though r and d are mathematical transforms of one another.

Corrections for restriction of range (direct or indirect) and measurement reliability are important. Estimates of d will be biased (usually downward) by restriction of range or imperfect measurement reliability on the continuous dependent variable. For example, unreliability in a measure of risk-taking propensity will likely downwardly bias the effect size associated with the categorical variable of entrepreneurial status (e.g., Brockhaus, 1980). Or, range restriction on the criterion (e.g., job performance) will downwardly bias the effect size associated with a dichotomous variable such as training versus no training (e.g., Morrow, Jarrett, & Rupinski, 1997). Furthermore, because differential levels of range restriction and unreliability across studies make the effect size estimates difficult to compare (Guzzo, Jette, & Katzell, 1985), meta-analytic and primary study researchers must be aware of, and correct for, these concerns if appropriate empirical correction parameters (e.g., degree of range restriction) are available.

Overview and Notation

For the most part (the exception is the correction for unreliability), the correction formulas for d that we present below are derived via the following logic. We first convert values of d to values of r, then apply known corrections to the values of r. These corrected values of r are then transformed back to corrected values of d. The derived formulas are grouped into four sections.

1. We consider a formula (from Hunter & Schmidt, 1990) that allows one to compute values of d from r and vice versa. A relatively more efficient version of an earlier formula is provided.
2. We consider direct range restriction on a continuous variable (X). That is, we have a value of d in a restricted sample and will correct the value of d for range restriction. The result is surprisingly straightforward, and the formula can be used to go directly from the restricted value of d to the unrestricted estimate of d with ease.
3. We consider indirect range restriction on a continuous variable (X) due to range restriction on *another* variable (say, Z). A formula is derived that allows one to go from the restricted value of d to the unrestricted estimate of d.
4. We provide a derivation for a formula that allows users to go directly from an attenuated value of d to the estimate of d obtained after correcting for unreliability in the outcome measure. This formula has been derived previously, but we present a straightforward proof to motivate increased use of this correction.

One of the best expositions with regard to corrections of correlations for range restriction can be found in Ghiselli (1964, chap. 11). To assist in the presentation and readability of our equations, we denote statistics that are computed in a restricted sample with capital letters. Thus, R represents a correlation between two variables that has been computed in a restricted sample. Correcting this correlation would lead to a sample estimate simply denoted as r. As another example, one could compute the standard deviation on a variable in a restricted set of scores; this would be denoted by S. The unrestricted standard deviation would simply be s. Indeed, the ratio s/S (unrestricted standard deviation divided by restricted standard deviation) is often used as an index of the degree of range restriction. For example, if X is normally distributed and one takes the top 5% of the scores on X, then s/S is about 3.0 (Bobko, 1995). The ratio of s/S is often denoted as u in range restriction formulas, and we use that notation.

Another important issue concerns the fact that one of the variables of interest will be dichotomous. Such a variable will generally be group membership (e.g., Black vs. White, trained vs. untrained, entrepreneur vs. manager). It will turn out that the proportions in each group affect both the transformations of r to d and the corrections themselves. We therefore denote the proportion in one group as p and the proportion in the other group as q (hence, $p + q = 1$). It will also be critical to remember that p and q may vary depending on whether one is in a restricted or unrestricted sample (so the symbols P and Q will also appear). This makes intuitive sense. For example, if the applicant population is 20% Black (value of p), it cannot readily be assumed that 20% of the selected, restricted sample is also Black (value of P) (Sackett & Ellingson, 1997).

Converting *r* to *d* and Vice Versa

As suggested earlier, the point-biserial correlation (r) between a dichotomous variable (e.g., Black vs. White) and a continuous variable (e.g., test score) depends, in part, on the mean test score for one group minus the mean test score for the other group. This difference in mean scores between the two groups also partially determines the value of d (i.e., d is the difference in group means divided by the within-group standard deviation).

Hunter and Schmidt (1990, pp. 273-275) provide a discussion of the link between r and d, as well as a variety of other issues including scales of measurement, true versus artificial dichotomization, and so on. For a continuous scale and a true dichotomy, Hunter and Schmidt report the relationship (converting d to r) as

$$r = \frac{\dfrac{d}{1/\sqrt{pq}}}{\sqrt{1+\left[\dfrac{d}{1/\sqrt{pq}}\right]^2}}.$$

A simpler version of this formula (implied in Hunter & Schmidt, 1990) is readily derived by reducing the complex fraction. The result will make the derivation of other formulas (see below) a bit less complicated. Some arithmetic on the above equation leads to

$$r = \frac{d}{\sqrt{\dfrac{1}{pq}+d^2}}. \tag{1}$$

Equation (1) can be reversed by solving for d. The derived result is

$$d = \frac{r}{\sqrt{pq\left(1-r^2\right)}}. \tag{2}$$

Equations (1) and (2) provide ready formulas for converting back and forth from values of r and d when one variable is dichotomously scored. As an example, it is generally assumed that the value of d for cognitive ability tests is about 1.0 when comparing Whites and Blacks. Assuming 20% Blacks and 80% Whites (as in Schmitt et al., 1997), Equation (1) converts the value of $d = 1$ to a value of $r = .37$. That is, the correlation between cognitive ability scores and ethnicity (coded as a 0-1 variable) is .37.

It is also instructive to note the role of p and q in these two equations. First, these parameters are an important part of the relationship between r and d, and the proportions must be accounted for. Second, the direction of the effect of p and q is clear. For example, let the value of d be constant in Equation (1). As p and q become more extreme (e.g., .01 and .99 vs. .50 and .50), the value of r becomes smaller (Hunter & Schmidt, 1990). This makes sense, as the relationship is reduced because the two group means are measured with unequal precision.

We now turn to Equations (1) and (2) to help with our primary purpose—correcting values of d for range restriction and attenuation.

Correction of *d* for Direct Range Restriction

Direct range restriction occurs when a researcher is interested in the covariance between two variables (X and Y) but data are restricted due to selection on one of the two variables (say, X). For example, it is common to analyze the effect of ethnic group status on some selection device in human resource research. In particular, individuals have reported analyses of Black-White differences on tests of cognitive ability. The literature sometimes reports values from incumbents (individuals already hired); on other occasions, the data are from applicants. One recent study (Caretta, 1997) reports statistics from both applicant and incumbent data for the same cognitive ability measure. In that study, $d = .46$ for incumbents and $d = 1.19$ for applicants. This is to be expected of organizations already using selection systems that incorporate cognitive ability. Thus, the use of a correction for direct range restriction is important if one conducts ethnic group difference comparisons across different tests (e.g., cognitive ability tests vs. biodata vs. interviews). To be comparable, statistics should be corrected for relevant artifacts.

We develop our correction formula (for *d*) using Ghiselli's (1964) well-known equation for correcting values of correlations for direct range restriction:

$$r = \frac{Ru}{\sqrt{1 - R^2 + (Ru)^2}}, \tag{3}$$

where R is the correlation in the restricted sample and $u = s/S$. For ease of exposition, we ignore the subscripts on the value of r because there is only one set of subscripts. That is, r always denotes the correlation between the continuous variable/test (X) and the dichotomous variable (e.g., group membership, Y). Similarly, d always denotes the effect size on X across the two categories (or subgroups) of interest.

Before proceeding, it is useful to consider the functional form of Equation (3). The numerator simply takes the restricted value of R and multiplies it by the index of range restriction, u. Because u is generally greater than 1.0, this has the intended effect of increasing the value of R. The denominator helps ensure that the resulting correction remains bounded by –1 and +1. With regard to underlying assumptions, the derivation of Equation (3), and hence our results, assumes that the regression of Y on X is linear and that the errors around that line are homoscedastic (see Ghiselli, 1964). The formula is robust to violations of homoscedasticity, although violation of the linearity assumption may either over- or underestimate the latent correlation depending on the form of the nonlinearity (Boldt, 1973; Brewer & Hills, 1969; Greener & Osburn, 1979, 1980). It is also important to note that the estimator in Equation (3) has a small negative bias, which decreases as the sample size increases, even when all assumptions are met (Bobko & Rieck, 1980). Thus, the corrected correlation will *under*estimate the latent correlation, so the process is conservative, on average. Finally, as noted by Linn, Harnisch, and Dunbar (1981), use of a corrected correlation will generally provide a much better estimate of the latent correlation than an uncorrected value.

If one computes a value of *d* for the variable X in a range-restricted sample, then the magnitude of *d* will also generally be lower than the value of *d* in the unrestricted sample. This statement should be obvious, given the existence of Equation (3) and the relationship between r and d in Equations (1) and (2). However, it is relatively rare to see corrections to values of d in the organizational literature. On the other hand, it seems critical to do so if organizational researchers or decision makers compare values of d to one another or average them across studies. As already noted, data from concurrent validity studies are often used in reporting values of d for ethnic differences on cognitive ability tests, interviews, or other potential selection tests (e.g., Bobko et al., 1999; Schmitt et al., 1997). Thus, some of the values of d will be artificially small (high degree of selectivity and substantial range restriction for incumbent data) whereas other values of d will be relatively large (low degrees of selectively and little range restriction), making explicit comparisons inappropriate.

To derive the range restriction correction formula for values of d, we first start with the restricted value, labeled D. The logic is to take the value of D and convert it to a value of R using Equation (1). The resulting value of R is then corrected for range restriction using Ghiselli's (1964) formula (Equation [3]). The corrected value (r) is then transformed back to a corrected value of d using Equation (2).

More specifically, given a value of D, Equation (1) indicates that the value of R is

$$R = \frac{D}{\sqrt{\frac{1}{PQ} + D^2}}.$$

To simplify the algebra a bit, let $M = \sqrt{\frac{1}{PQ} + D^2}$. Thus, $R = \frac{D}{M}$. Putting this restricted value of R into Equation (3) yields

$$r = \frac{\frac{D}{M}u}{\sqrt{1 - \frac{D^2}{M^2} + \frac{D^2}{M^2}u^2}}.$$

This reduces to

$$r = \frac{Du}{\sqrt{M^2 - D^2 + D^2u^2}}.$$

To then obtain the value of d, the expression for r immediately above is placed in Equation (2), resulting in

$$d = \frac{Du\big/\sqrt{M^2 - D^2 + D^2u^2}}{\sqrt{pq\left[1 - \frac{D^2u^2}{M^2 - D^2 + D^2u^2}\right]}}.$$

After a variety of algebraic manipulations, this eventually simplifies to

$$d = \frac{Du}{\sqrt{pq(M^2 - D^2)}}.$$

It is important to remember that the values of p and q depend on whether one is considering the restricted or unrestricted sample. The values of p and q in the above equation refer to transformations occurring in the unrestricted sample. Hence, they do not have the "capital" notation. However, the value of M is partially dependent on values of P and Q, since these proportions were estimated in the restricted sample.

Finally, substituting for M in the above equation yields

$$d = Du\sqrt{\frac{PQ}{pq}}. \tag{4}$$

This is a remarkably straightforward equation, and we encourage researchers to use it in correcting values of D for direct range restriction on X when appropriate.

As an example, Equation (4) might be applied to a situation in which a researcher is studying the relationship between ethnic status (Black vs. White) and cognitive ability. The researcher might observe that $D = .38$ in a group of hires selected based on cognitive ability. (We took this value from a recent meta-analysis by Roth, BeVier, Bobko, Switzer, and Tyler [2000], who report an overall value of .38 for incumbent samples.) Noting that the value of .38 must be corrected, the researcher gathers information such as $p = .83$ and $q = .17$ (as per Hunter & Schmidt, 1990). Possible values for P and Q should reflect a smaller proportion of minority hires on a cognitive ability test (as per Sackett & Wilk, 1994), so a researcher might find that $P = .90$ and $Q = .10$. Finally, assume, as earlier, that the value of u is approximately 3.0 (low selection ratio). Using Equation (4), the corrected $d = .91$.

Although it will not occur unless $d = 0$ or unless there is no range restriction, it is interesting to note that if $p = P$, the correction formula in Equation (4) reduces further to the simple function $d = uD$. However, as was the case for Equations (1) and (2), the differences in values of p and P can generally have a substantial influence on the correction for d provided in Equation (4). For example, if 20% of the applicants are minorities but only 5% of the hires are minorities, then the last factor in Equation (4) is .545, thereby reducing the degree of correction by this factor.

Correction of *d* for Indirect Range Restriction

We now turn our attention to indirect range restriction. Indirect range restriction occurs when the observed relationship between two variables is influenced by their relationship to a third variable. The third variable in some way reduces the range of scores on the two variables of interest.

For example, if individuals who are given a biodata inventory (X) have been prescreened on some paper-and-pencil test of cognitive ability (Z), then to the extent that X and Z are related, scores on biodata (X) will be indirectly restricted in range. In turn, the correlation between X and any other measure (e.g., ethnicity) will be artificially reduced in magnitude. That is, the value of d for the biodata score will be too small relative to its value in the applicant population. Suppose, as we have seen in practice, a human resource decision maker contemplates replacing a test of cognitive ability with a biodata measure because the value of d for biodata is smaller than the value of d for cognitive ability. Of course, we strongly recommend including other aspects of the measurement process in this decision, such as validity, choice of criterion measure, cost, and so on. In any event, as noted earlier, any comparison of *d*s may be faulty due to differential range restriction. Indeed, summary reviews of the biodata literature note that most ethnic group difference assessments have occurred in incumbent populations, and statistical estimators are subject to the effects of range restriction (Reilly & Warech, 1994).

Examples of indirect range restriction are also found in many other domains. For example, marketing research has examined the influence of qualitatively different parental styles on variables such as the amount of disposable income of children and parental attitudes (e.g., Carlson & Grossbart, 1988; Walsh, Laczniak, & Carlson, 1998). If individuals at both the higher and lower ends of the income spectrum are less likely to respond to surveys (e.g., see Roth, 1994), then the third variable of parental income (which is likely related to parental styles and children's disposable income) may artifactually restrict the range of the primary variables of interest.

Or, in the domain of human resources training programs, some studies will compare trained employees to untrained employees on a measure of job performance and then calculate an effect size (Guzzo et al., 1985; Morrow et al., 1997). Results of such studies may be influenced by indirect range restriction, since many employees are initially hired based on general cognitive ability or other constructs related to job performance.

Because a third variable has been introduced, some additional notation is needed for our derivation. Let X continue to be the variable of interest (e.g., biodata or interview measure). Let Z be the variable that causes the indirect range restriction (e.g., a prescreening measure of cognitive ability or GPA). Let E be the dummy variable used to code group membership. (We use E to reflect the selection example and to serve as a reminder that ethnicity is a frequent dichotomous variable of interest, although the results generalize to any other dichotomous variable, such as trained/untrained.)

There are now three possible correlations: r_{xz}, r_{xe}, and r_{ze}; the latter two are point biserial correlations. There are also two possible values of d: d_x and d_z. And, all of these statistics can be considered in both the unrestricted form and the restricted form.

We start with Ghiselli's (1964, p. 367) formula for correcting a value of r for indirect range restriction:

$$r_{xe} = \frac{R_{xe} - R_{zx}R_{ze} + R_{zx}R_{ze}u^2}{\sqrt{\left[1 - R_{zx}^2 + R_{zx}^2u^2\right]\left[1 - R_{ze}^2 + R_{ze}^2u^2\right]}}, \tag{5}$$

where u is defined on the variable causing the range restriction (i.e., $u = s_z/S_z$).

As in Equation (3), note that the restricted correlation of interest (R_{xe}) appears in the numerator. However, in Equation (3), the restricted value of r was multiplied by the value of u. In indirect range restriction, there is no multiplicative correction. Rather, there is a subtraction and an addition (involving u^2). Furthermore, because there are three correlations, the formula is clearly more complicated than the correction for direct range restriction.

To remind the reader, the purpose of this section is to take the observed value D_x that is restricted because of indirect range restriction on Z, and estimate the unrestricted value, d_x. Therefore, we start by rewriting Equation (5) in terms of values of D_x. For example, from Equation (1) we know that

$$R_{xe} = \frac{D_x}{\sqrt{\frac{1}{PQ} + D_x^2}}.$$

Substituting for R_{xe} and R_{ze} in Equation (5) and rearranging terms leads to

$$r_{xe} = \frac{D_x\sqrt{\frac{1+PQD_z^2}{1+PQD_x^2}} + (u^2-1)R_{xz}D_z}{\left[1+(u^2-1)R_{zx}^2\right]\left[\frac{1}{PQ}+D_z^2u^2\right]}.$$

Note that R_{xz} is not converted to a value of D because neither respective variable (X or Z) is dichotomous. We assume that only the group membership variable (E) is dichotomous. The above expression for r_{xe} can then be placed in Equation (2) to derive the equation for d_x. The derivation requires some tedious, but straightforward, algebra. To make the result a bit less complex, we define the intermediate statistic, M, as

$$M = \sqrt{\frac{1+PQD_z^2}{1+PQD_x^2}}.$$

The end result is that d_x can be expressed in terms of D_x as

$$d_x = \frac{D_xM + (u^2-1)R_{zx}D_z}{\sqrt{pq}\sqrt{\frac{1}{PQ} + u^2D_z^2 + (u^2-1)R_{zx}^2\left(\frac{1}{PQ}+D_z^2\right) - D_x^2M^2 - 2(u^2-1)MR_{zx}D_zD_x}}, \tag{6}$$

where M is defined as above.

It is important to note that the values of p and q in the beginning of the denominator of this expression are proportions taken from the unrestricted (applicant) population. In contrast, the values of P and Q are taken from proportions in the restricted sample. Unfortunately, the expression for d_x does not simplify as much as in the case for direct range restriction.

As an example, consider the selection test situation described earlier—initially comparing a value of d (White vs. Black) for biodata (X) with a value of d for cognitive ability (Y). A recent meta-analysis (Bobko et al., 1999) suggested that the d for biodata is .33. However, consistent with Reilly and Warech's (1994) observation, this value of d is from incumbent populations (a check of the citations used in the meta-analysis confirmed this suspicion). We assume that many of the selection systems used to select those incumbents have used some indicator of cognitive ability. To the extent that biodata measures/items reflect cognitive ability (e.g., items with regard to academic achievement), the range of scores on biodata (X) will be indirectly restricted via preselection on cognitive ability (Z). Thus, the value of .33 should be corrected for range restriction. This restriction is further reflected in our notation as $D_x = .33$. From Bobko et al. (1999), we also take the value $R_{xz} = .19$. Furthermore, similar to the findings previously reported from Caretta (1997), we use $D_z = .46$. To complete the use of Equation (6), we assume (a) a value of $u = 2.5$ to model range restriction when the selection ratio is 10% (Schmidt, Hunter, & Urry, 1976), (b) the proportion of White applicants is .8 ($p = .8$), and (c) the proportion of White incumbents is .9 ($P = .9$) because of some indirect selection on cognitive ability.

Application of Equation (6) in this situation yields an estimate of $d_x = .52$. This effect size is 58% greater than the initially reported value of .33. Thus, the corrected value of *d* indicates that there will be a greater potential for adverse impact than originally anticipated (although still not as much as a test of cognitive ability, assuming the generally accepted value of 1.00) if a biodata measure is used in place of cognitive ability across job applicants.

Correction of *d* for Attenuation

Measurement reliability has long been known to downwardly bias the magnitude of correlations between two variables. As recently noted by several researchers (Hunter & Schmidt, 1990; Schmitt, Clause, & Pulakos, 1996), the uncorrected effect size of *d* will also underestimate its population counterpart (again, the same artifacts affect both *r* and *d*).

For example, in the organizational productivity literature, it is fairly common to calculate effect sizes of interventions on job performance without correcting for unreliability in the index of job performance (Guzzo et al., 1984; Morrow et al., 1997; Pritchard, Jones, Roth, Stuebing, & Ekeberg, 1988). We were able to find only one empirical study (Burke & Day, 1986) that corrected effect size estimates (training versus no training) for unreliability in the criterion measure, and we commend the authors for making this correction. Burke and Day (1986) report that they took the effect size and divided it by the square root of the reliability of the criterion measure. In fact, that correction factor has been derived in Hedges's (1981) structural equations analysis of the distribution theory of *d* (see also Hedges & Olkin, 1985), and it is restated (without proof) in Hunter and Schmidt (1990). We provide a straightforward proof in this article to motivate use of this correction factor in future research.

The derivation proceeds by applying classical test theory to the definition of *d*. We again use the capital symbol *D* to denote the observed, attenuated value of *D*. Thus,

$$D = \frac{\bar{X}_1 - \bar{X}_2}{S_{pooled\ within}}, \tag{7}$$

where *S* denotes the within-group standard deviation in the observed, attenuated database.

In classical test theory, unreliability is introduced via random error (assumed to have mean zero) by the following model: observed scores (X) = true scores (T) + random error (e). Reliability (*rel*) in this model is defined as

$$rel = s^2_{true} \big/ S^2_{observed},$$

which leads to

$$s_{true} = \left(\sqrt{rel}\right) S_{observed}.$$

There is no effect of unreliability on the numerator of Equation (7), since the random error is assumed to have a mean of zero. Thus, there is no influence on the group means or their difference. On the other hand, as random error is increased, the variance on X (see the denominator of *D*) will increase. In turn, the value of *D* is reduced.

If one is interested in the true scores, the value of the observed standard deviation in Equation (7) must be reduced by a factor equivalent to $\sqrt{rel}$. Thus,

$$d = \frac{\overline{X}_1 - \overline{X}_2}{\left(\sqrt{rel}\right) S_{pooled\ within}}.$$

Substituting for *D* from Equation (7), we get

$$d = \frac{D}{\sqrt{rel}}, \tag{8}$$

which is consistent with Burke and Day (1986).

As an example, some researchers (e.g., Brockhaus, 1980; Carland, Hoy, Boulton, & Carland, 1984) have studied the relationship between entrepreneurial status (entrepreneur vs. manager) and risk-taking propensity (via a personality scale from the Jackson Personality Inventory). A study in this area might result in $D = .50$. However, such values are almost never corrected for unreliability. An internal consistency estimate of the reliability of the Jackson Personality Inventory is .76 (Stewart, Watson, Carland, & Carland, 1999). Correcting *D* using Equation (8) results in an estimated *d* of .57, a 14% increase in the estimated standardized group difference. Indeed, this correction is probably conservative, since we used an available internal consistency estimate of reliability. Such estimates are usually large given that they do not include variance due to time or alternative forms of a test, and they may "undercorrect" measures of relationship (Hunter & Schmidt, 1990).

Conclusion

We have provided formulas for correcting effect sizes based on the *d* statistic. These formulas (particularly Equations [4], [6], and [8]) are useful and appropriate in the face of range restriction or unreliability. We encourage their use in organizational research when accurate, empirical estimates of the parameters in the above equations are available.

Without these corrections, analyses of effect sizes may be misleading. For example, within human resources management, we demonstrated that noncognitive predictors of entry-level performance might have more adverse impact than previously believed—due to both direct and indirect range restriction. As another example, in research on risk taking and entrepreneurs, we noted that estimates of effect sizes might be underestimated by substantial amounts.

We also suggest that thinking about these corrections facilitates the investigation of previously neglected research questions/issues. For example, although there has been more than a century of research on this issue (Galton, 1892; Herrnstein & Murray, 1994), the vast majority of concern with regard to ethnic differences in cognitive ability has been with observed effects (e.g., Herrnstein & Murray, 1994) rather than with asking the question, "What is the difference at the population level (assuming perfectly reliable measurement or lack of range restriction)?" The lack of focus on the population level of analysis inhibits the formation of accurate relationships between variables at this level and could retard accurate theory development. We suggest that increased implementation of corrections for the *d* statistic will facilitate asking important questions, more accurate population analysis, and better theory development.

These formulas are also the starting point from which a great deal of research might flow. We suggest research to systematically assess the level of influence these artifacts have across organizational domains (marketing, training, etc.). In many areas, this will require gathering estimates of restricted and unrestricted variances on a variety of variables and understanding the dynamics involved in the restriction. For example, entrepreneurship researchers may need to examine whether important individual difference variables such as risk propensity (or need for achievement) undergo restriction of range due to the success of the business venture. If successful businesses are those with higher or lower risk propensities, then effect sizes may need to be adjusted using Equation (4) or Equation (6). Likewise, corrections for attenuation should be considered when appropriate.

In other business disciplines, there may be enough information to begin to understand and model the influence of artifacts. Human resources management may represent such an area, since researchers have estimates of some of the key statistics involved in restriction of range or attenuation due to measurement unreliability. For example, cognitive ability tests (or GPAs) are often used to screen individuals before they are allowed to enter the second stage of selection in which an interview might be used. And, the published literature contains estimates of the correlations between variables such as cognitive ability, interview evaluations, and so on. Researchers interested in personnel selection could use Monte Carlo analyses to understand likely levels of indirect range restriction in such situations. Monte Carlo analyses might be an excellent place to start for two reasons. First, there is a need to understand the role of selection ratios in such research, since selection ratios are very important in analyzing ethnic group effect sizes (Sackett & Ellingson, 1997). Second, we suspect some of the statistics needed in individual primary studies are not reported. (For example, we note the difficulty of finding full information for corrections in Huffcutt and Roth's [1998] study of ethnic group differences in interview evaluations.) Indeed, as a reviewer suggested, the rarity of corrections of *d* may be due to the lack of information available to perform the corrections. As a corollary to the need for Monte Carlo analyses, we call on researchers to report relevant artifact-related information when available.

We also call for organizational research at the meta-analytic level to incorporate these concerns. Meta-analytic studies that examine restricted and unrestricted groups (e.g., applicants vs. hires) using the *d* statistic may wish to use the degree of range restriction as a moderator variable. Other researchers may wish to correct *d*s for measurement unreliability. Many meta-analysts may also want to renew their search for information to correct their estimates of effect sizes for these artifacts.

As a cautionary note, use of our formulas assumes that accurate, empirical estimates of the parameters in the above equations are available, since the correction formulas are only as good as the available estimates and underlying assumptions. With regard to assumptions, as one reviewer noted, the robustness studies cited earlier assume continuous variables, yet one of our variables is always a true dichotomy. If p or q is not close to 0 or n is large, the implied dichotomous, binomial variable is well approximated by a continuous normal distribution. However, we encourage research on the properties of our estimators when n is small and p or q is close to zero.

We also made two mathematical choices in deriving and presenting our results. First, we chose several examples and some meta-analytic values to indicate the variety of applications of our formulas, as well as place the examples in realistic and substantiated values for r, d, u, and unreliability. Second, we wanted to derive closed, analytic solutions to the problem because the formulas for correlational corrections were available and the solution was tractable. Thus, the formulas are, by definition, accurate but only to the extent that assumptions hold. These assumptions were outlined earlier, and the research to date has shown that correction formulas tend to be quite robust to the assumptions. On the other hand, we encourage Monte Carlo research with regard to the accuracy of the formulas under violations of combinations of the assumptions.

In sum, we hope that our formulas encourage researchers to consider correcting their estimates of effect size when calculating the d statistic when it is appropriate to their research question. We also hope that consideration of the corrections will help facilitate asking many important research questions. We look forward to a great deal of future research that the examination of these underlying issues should encourage.

References

Bobko, P. (1995). *Correlation and regression: Principles and applications for industrial/organizational psychology and management*. New York: McGraw-Hill.

Bobko, P., & Rieck, A. (1980). Large sample estimators for standard errors of functions of correlation coefficients. *Applied Psychological Measurement*, *4*, 385-398.

Bobko, P., Roth, P., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology*, *52*, 561-589.

Boldt, R. (1973). *Range restriction assumptions assuming quadratic conditional variances* (Research Bulletin 73-59). Princeton, NJ: Educational Testing Service.

Brewer, J., & Hills, J. (1969). Univariate selection: The effects of size of correlation, degree of skew, and degree of range restriction. *Psychometrika*, *34*, 347-361.

Brockhaus, R. (1980). Risk taking propensity of entrepreneurs. *Academy of Management Journal*, *23*, 509-520.

Burke, M., & Day, R. (1986). A cumulative study of the effectiveness of managerial training. *Journal of Applied Psychology*, *71*, 232-245.

Caretta, T. (1997). Group differences on US Air Force pilot selection tests. *Sex and Ethnic Group Differences*, *5*, 115-126.

Carland, J., Hoy, F., Boulton, W., & Carland, J. (1984). Differentiating small business owners from entrepreneurs. *Academy of Management Review*, *9*, 354-359.

Carlson, L., & Grossbart, S. (1988). Parental style and consumer socialization of children. *Journal of Consumer Research*, *15*, 77-94.

Galton, F. (1892). *Hereditary genius*. London: Macmillan.

Ghiselli, E. (1964). *Theory of psychological measurement*. New York: McGraw-Hill.

Greener, J., & Osburn, H. (1979). An empirical study of the accuracy of corrections for restriction in range due to explicit selection. *Applied Psychological Measurement*, *3*, 31-41.

Greener, J., & Osburn, H. (1980). Accuracy of corrections for restriction in range due to explicit selection in heteroscedastistic and non-linear distributions. *Educational and Psychological Measurement*, *40*, 337-345.

Guzzo, R., Jette, R., & Katzell, R. (1985). The effects of psychologically based intervention programs on worker productivity: A meta-analysis. *Personnel Psychology*, *38*, 275-291.

Hedges, L. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107-128.

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Herrnstein, R., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Huffcutt, A., & Roth, P. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology*, *83*, 179-189.

Hunter, J., & Schmidt, F. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

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

Linn, R., Harnisch, D., & Dunbar, S. (1981). Corrections for range restriction: An empirical investigation of conditions resulting in conservative corrections. *Journal of Applied Psychology*, *66*, 655-663.

Morrow, C., Jarrett, M., & Rupinski, M. (1997). An investigation of the effect and economic utility of corporate wide training. *Personnel Psychology*, *50*, 91-119.

Pritchard, R., Jones, S., Roth, P., Stuebing, K., & Ekeberg, S. (1988). Effects of group feedback, goal setting, and incentives on organizational productivity. *Journal of Applied Psychology*, *73*, 337-358.

Reilly, R., & Warech, M. (1994). The validity and fairness of alternatives to cognitive tests. In L. Wing & B. Gifford (Eds.), *Policy issues in employment testing* (pp. 131-224). Boston: Kluwer.

Roth, P. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, *47*, 537-559.

Roth, P., BeVier, C., Bobko, P., Switzer, F., & Tyler, P. (2000, August). *Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis*. Paper presented at the annual meeting of the Academy of Management, Toronto.

Russell, J., Wexley, K., & Hunter, J. (1984). Questioning the effectiveness of behavior modeling training in an industrial setting. *Personnel Psychology*, *37*, 465-481.

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

Sackett, P., & Ostgaard, D. (1994). Job specific applicant pools and national norms for cognitive ability tests: Implications for range restriction corrections in validation research. *Journal of Applied Psychology*, *79*, 680-685.

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

Schmidt, F., Hunter, J., and Urry, V. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology*, *61*, 473-485.

Schmitt, N., Clause, C., & Pulakos, E. (1996). Subgroup differences associated with different measures of some common job-relevant constructs. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 115-139). New York: John Wiley.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology*, *5*, 719-730.

Stewart, W., Watson, W., Carland, J., & Carland, J. (1999). A proclivity for entrepreneurship: A comparison of entrepreneurs, small business owners, and corporate managers. *Journal of Business Venturing*, *14*, 189-214.

Walsh, A., Laczniak, R., & Carlson, L. (1998). Mother's preferences for regulating children's behavior. *Journal of Advertising*, *27*, 23-36.

Philip Bobko is a professor of management and psychology at Gettysburg College. His research interests and publications are in statistics, research methods, personnel selection, and organizational behavior. He has served as editor of the Journal of Applied Psychology.

Philip L. Roth is a professor of management at Clemson University. His research interests include employee selection and research methods. His selection interests center on interviewing, utility analysis, and grades as predictors of organizational outcomes. His research methods interests include missing data, meta-analysis, and influential cases.

Christopher Bobko is a B.S.E. candidate in civil engineering at Princeton University. Along with statistics, he is interested in the aesthetics of structures in civil engineering and architecture.