Butler University LIBRARIES

Butler University
Digital Commons @ Butler University

Scholarship and Professional Work - Business | College of Business

1998

Correcting turnover correlations: A critique

Chuck R. Williams
Butler University, crwillia@butler.edu

L. H. Peters

Follow this and additional works at: http://digitalcommons.butler.edu/cob_papers

 Part of the Organizational Behavior and Theory Commons, and the Other Psychology Commons

Recommended Citation
Williams, Chuck R. and Peters, L. H., "Correcting turnover correlations: A critique" (1998). *Scholarship and Professional Work - Business.* Paper 3.
http://digitalcommons.butler.edu/cob_papers/3

This Article is brought to you for free and open access by the College of Business at Digital Commons @ Butler University. It has been accepted for inclusion in Scholarship and Professional Work - Business by an authorized administrator of Digital Commons @ Butler University. For more information, please contact fgaede@butler.edu.

Correcting Turnover Correlations: A Critique

CHARLES R. WILLIAMS
LAWRENCE H. PETERS
Texas Christian University

In this article, the authors argue that turnover correlations do not need to be corrected. First, they maintain that correction formulas cannot correct for poor construct validity. Second, they discuss the original purposes of turnover correction formulas. Third, the authors describe the logical fallacies of correcting turnover correlations. Finally, they show why turnover correlations are not, as is widely believed, statistically limited to a maximum of .80.

A methodological issue that has been the topic of some controversy in recent turnover literature is whether differences in turnover base rates affect the magnitude of turnover correlations. This concern stems from the common use of the point-biserial correlation (r_{pb}) in turnover studies. Point-biserial correlations represent the relationship between a continuous variable and a dichotomous variable; in this case, turnover. In comparison to Pearson product-moment (PM) correlations between two continuous variables, which have a maximum of ± 1.00, it is widely accepted that the maximum value of r_{pb}s is ± .798 (Nunnally, 1978). Furthermore, because the observed values of r_{pb}s are expected to be largest when the number of stayers equals the number of leavers (i.e., there is a 50-50 split on the dichotomous variable), any divergence from a 50-50 split would lower the size of the r_{pb}. For example, an r_{pb} of .40 obtained from a 50-50 turnover split would theoretically shrink to .37 with a 25-75 split or to .24 with a 95-05 split.

In response to these limitations, Kemery, Dunlap, and Griffeth (1988); Steel, Shane, and Griffeth (1990); and Bass and Ager (1991) proposed formulas and procedures for correcting turnover r_{pb}s. Williams (1990) questioned the use of such procedures and argued that there were sound statistical and methodological reasons for *not* correcting turnover r_{pb}s.

This article reexamines the issue of turnover correlation correction formulas. Although previous arguments for the most part have been about statistical issues, we believe that basic disagreements about correction of turnover r_{pb}s stem from different

construct definitions of turnover. Therefore, we begin by reviewing the different ways in which turnover has been defined as a theoretical construct and by explaining how the construct validity of turnover measures is related to the decision to correct turnover r_{pb}s. Following this, we discuss and extend the methodological issues introduced by Williams (1990).

Construct Validity of Turnover Measures

Scholars have taken a variety of positions regarding the conceptual meaning of turnover. For example, March and Simon (1958, p. 92); Mobley (1982, p. 10); Mowday, Porter, and Steers (1982, p. 124); and Williams (1990) have argued that turnover is a *truly dichotomous construct*. Employees either stay or they leave, and this act reflects their position on an underlying dichotomous conceptual plane. Yet, others have regarded turnover as the observable manifestation of an *underlying continuous construct*. Hunter and Schmidt (1990a) explain that "in many cases, the theoretical variable of interest is the *propensity* to quit (a continuous variable)" (p. 334). McEvoy and Cascio (1987) stated that "turnover is a dichotomization of the continuous variable called tenure" (p. 750). Hulin, Rosse, and colleagues have suggested that turnover is just one part of a *latent withdrawal construct* known as job adaptation that consists of four withdrawal behaviors: voluntary absenteeism, lateness, voluntary turnover, and retirement. These behaviors are said to be adaptive because they allow employees to decrease the amount of time they spend in dissatisfying work environments (Hulin, 1990; Rosse, 1988; Rosse & Hulin, 1985).

By definition, different epistemological positions such as these are difficult if not impossible to reconcile via data or statistics. Furthermore, the scientific process is well served by examining phenomena, such as organizational withdrawal, from a number of different perspectives. Nonetheless, we firmly believe that the methodological and statistical practices used by turnover researchers must, at the very least, be consistent with their espoused, or implied, epistemological positions. Consequently, discussions concerning correction of turnover correlations should begin by examining the construct validity of espoused (or implied) turnover measures.

One definition of construct validity is "the correspondence between a construct (conceptual definition of a variable) and the operational procedure to measure or manipulate that construct" (Schwab, 1980, pp. 5-6). According to this definition, a basic step in determining the construct validity of turnover measures would be to make sure that those measures are appropriately matched to turnover constructs. In many turnover studies, a dichotomous operational measure has been used as a proxy measure for the *dichotomous stay/quit construct*; in other studies, this same measure has been a proxy measure for different *latent continuous constructs* (e.g. propensity to quit, tenure, or job adaptation). That is, the different theoretical withdrawal constructs described above often have been measured using the same operational variable. If these studies were designed to investigate the dichotomous stay/quit turnover construct, then a match exists between specification of the theoretical construct and the operational measure.[1] On the other hand, if these studies were designed to measure turnover propensity, tenure, or job adaptation, one needs to ask whether a dichotomous measure would possess acceptable construct validity, especially because there are more direct ways to measure these variables.

For example, when the construct of interest is tenure, we would question the practice of representing a continuous tenure distribution with a dichotomous operational measure. Dichotomizing a naturally continuous variable such as tenure only results in a loss of information and measurement precision (J. Cohen, 1983). The consequence, as others have repeatedly pointed out, is that artificially dichotomizing tenure will "systematically understate the actual correlation" between tenure and its predictors (see, e.g., Hunter & Schmidt, 1990a, p. 334). Does this mean that the correction for dichotomization formula should be used to estimate the actual correlation? In our opinion, the answer is clearly "no." First, why use a dichotomous proxy for tenure and then correct "tenure" correlations for that dichotomization when (a) tenure can be directly assessed, and (b) the correction formulas may produce estimation errors that were not in the original data (Nunnally, 1978). Second, at least in cross-sectional designs (see L. H. Peters & Sheridan, 1988), no correction formula can erase the mismatch between the continuous construct of employee tenure and a dichotomous stay/quit measure. This is because in cross-sectional designs, a questionnaire is administered at Time 1 and turnover is assessed later, at Time 2. Because participants in these studies typically begin the study at Time 1 with different amounts of *prior* tenure, the length of employment since Time 1 can only reflect how long employees have stayed after the first point of data collection. That is, with cross-sectional samples, a dichotomous operational measure cannot indicate how long employees have stayed since joining the company (Williams, 1990).[2] In short, correcting turnover correlations does not and cannot correct for mismatches between the construct of employee tenure and a dichotomous stay/quit measure.

The same is true for turnover propensity, which is a continuous variable that indicates the probability that someone will quit. As with employee tenure, why dichotomize this continuous distribution by using a dichotomous operational measure?[3] Furthermore, applying a correction formula cannot change turnover propensity into the dichotomous stay/quit turnover construct. By definition, they are different. In fact, examining the observable relationship between turnover propensity and turnover is a determination of the predictive validity of two conceptually distinct constructs and reflects construct validity only to the extent that turnover propensity is a variable in the nomological net of turnover. Likewise, correction formulas cannot turn a dichotomous operational measure into the latent withdrawal construct of job adaptation. Lisrel or some other covariance-based structural equations technique should be used to model job adaptation, which is determined by the covariances between the manifest behaviors of absenteeism, lateness, turnover, and retirement. Indeed, Hulin (1990) specifically states that the meaning of job adaptation literally "resides in those covariances" (p. 477).

To summarize, it is difficult (and probably inadvisable) to resolve disagreements about the merits of different turnover constructs. However, given a particular epistemological position, it is reasonable to expect that operational variables and statistical practices will be consistent with the theoretical construct that one adopts. Whether researchers choose to study turnover propensity, tenure, or job adaptation, it is incumbent on them to measure them in construct valid ways. Because turnover propensity, employee tenure, or job adaptation can be assessed more directly and because use of a dichotomous stay/quit proxy variable may not directly reflect relevant variance on those underlying continua, we conclude that dichotomous measures of turnover do not possess construct validity when used to measure those underlying

continuous withdrawal constructs. Therefore, turnover correlation correction formulas *cannot* correct for this poor construct validity.

With these arguments as a foundation, we now examine a number of issues associated with correction of turnover r_{pb}s. First, we review and discuss the histories and original purposes of turnover correction formulas. Second, we describe the logical fallacies of correcting turnover r_{pb}s. Third, we explain and demonstrate why r_{pb}s for true dichotomies are not, as is widely believed, statistically limited to a maximum of .80. Finally, we point out that because correlations are unit-free measures, meaningful inferences can be made across different turnover studies even when the rate of turnover differs from study to study.

History and Purpose of Correcting for Dichotomization

An important but overlooked fact in the debate over turnover r_{pb}s is that none of the recommended correction formulas was originally designed to "correct" turnover r_{pb}s. Each was developed under different sets of assumptions and with different purposes in mind.

The correction for dichotomization (see Williams, 1990, p. 733, formula 1) was developed to translate r_{pb}s into *biserial* correlations (r_bs), thereby raising the maximum size of r_{pb} correlations from .798 to 1.00. This correction was developed prior to the widespread availability of calculators (and computers) to shorten the number of mathematical steps, and therefore the time, required to calculate correlations by hand (J. Cohen, 1983; C. C. Peters & Van Voorhis, 1940). For example, Dunlap (1936) published a 13-step, timesaving process in which test scores were "serialized," meaning they were sorted into intervals based on criterion scores.[4] The means or midpoints for each test item were then "used to represent the cases in each interval" (J. Cohen, 1983, p. 249). Item-total correlations could then be estimated from these data; however, they needed to be corrected for "broad categories" because of the loss of accuracy associated with grouping data into artificially created intervals (J. Cohen, 1983; Jaspen, 1946; C. C. Peters & Van Voorhis, 1940). Accordingly, formulas and tables were developed for general serial correlations that could estimate biserial correlations, triserial correlations, quadriserial correlations, and so on, all the way up to 15 intervals (J. Cohen, 1983).

Thus, in the past, continuous data were purposefully grouped into smaller, more manageable intervals but only because it saved computational time. Moreover, by calculating serial correlations, which estimate PM correlations but are not PM correlations (Nunnally, 1978), earlier researchers chose to sacrifice estimation accuracy in exchange for savings in computational time. In most cases, however, the decrease in accuracy was not critical, because serial correlations were primarily used to calculate item-total correlations, which were used, in turn, to make simple decisions about whether to keep or drop specific test items during test construction (J. Cohen, 1983).

Today, however, correlations are used as inputs for path models, factor and component analyses, multivariate analyses, meta-analyses, and utility analyses. For example, in utility analyses (Schmidt, Hunter, McKenzie, & Muldrow, 1979), correlation coefficients are multiplied by SD_y (i.e., the standard deviation of the dollar value of job performance). Because SD_y estimates can run into "five figures," even small changes in the size of correlation coefficients can dramatically alter the cost/benefit estimates from utility models (McCall & Bobko, 1990). Consequently, in this and

many other areas, estimation accuracy is much more important for modern day researchers and practitioners.

Fortunately, given the nearly instantaneous production of statistical results from computers, researchers need not sacrifice accuracy for computational time (Nunnally, 1978). In fact, as far back as 1978, Nunnally argued that there are "very strong reasons" for not using biserial correlations "in most of the ways that they have been used in the past" (pp. 136-137). Most important, biserial correlations should not be used when the dichotomous variable is inherently dichotomous (Nunnally, 1978) because this violates the assumption that "the dichotomous variable is basically continuous and normally distributed, and that the two dichotomies together form a whole normal distribution" (Jaspen, 1946, p. 23). Accordingly, correcting for dichotomization is inappropriate when turnover is a dichotomous variable and construct.

If this reason were not enough to warrant against correcting turnover correlations for dichotomization, Nunnally (1978) also argued that r_bs can be very poor estimates of PM correlations. He cites an example in which the PM correlation between the two continuous variables was .52. Yet, after artificially dichotomizing one of the continuous variables at its median, the biserial formula estimated the PM correlation to be .71! Statistical textbooks also indicate that r_bs sometimes take on values less than −1 and greater than +1 and, under special circumstances, can produce values greater than 1.25 (Glass & Stanley, 1970; Lord & Novick, 1968). These estimation errors led Nunnally to recommend that r_bs not be used when computing partial correlations, multiple correlations, or other kinds of multivariate analyses. Consistent with Nunnally's warning, we also recommend that r_{pb}s not be transformed into r_bs in individual studies (Steel et al., 1990) or when conducting turnover meta-analyses (Hunter & Schmidt, 1990a, 1990b; Steel et al., 1990). Because r_{pb}s are mathematically equivalent to PM correlations, and because they have the same sampling error variance as PM correlations, there is, in our view, no compelling statistical reason to correct turnover r_{pb}s for dichotomization.[5]

Finally, because this unnecessary correction raises the maximum size of r_{pb}s, this means that correcting for dichotomization will exaggerate the size of turnover r_{pb}s. For example, in a meta-analysis of performance and voluntary turnover, Williams and Livingstone (1994) found that mean, sample-size weighted r_{pb}s were 41% larger after being corrected for dichotomization. Wanous, Poland, Premack, and Davis (1992) found even larger differences in their meta-analysis of met expectations. After correcting for dichotomization, their mean, sample-size weighted turnover r_{pb}s were 46% larger! Indeed, Nunnally (1978) warned that substituting r_bs for r_{pb}s is "illogical" when the dichotomous variable is a real dichotomy and will "only fool one into thinking that the variables have explanatory power beyond that which they actually have" (p. 136).

History and Purpose of Correcting for Unequal *n*s

The correction for unequal *n*s (see Williams, 1990, p. 734, Formula 4) was not designed to raise the maximum size of point-biserial correlations from .798 to 1.00. Instead, because r_{pb}s are typically largest when the dichotomous variable has a 50-50 split, it attempts to estimate what the turnover r_{pb} would be if the turnover rate were .50.[6]

The correction for unequal *n*s has been discussed clearly in the meta-analytic and statistical power literature (J. Cohen, 1988; Glass, 1977), which has focused on two

kinds of effect sizes, *d* and the r_{pb}. The effect size *d* is typically used when cumulating effect sizes from experiments. It is computed by subtracting the mean of the control group from the mean of the treatment group and then dividing that difference by the standard deviation of the control group (Glass, 1977) or by the pooled standard deviation from the control and treatment groups (Hunter & Schmidt, 1990b). Thus, like a *z* score, *d* indicates in standard deviation units how large a difference the treatment made. However, the same treatment effect can also be represented by the r_{pb}, which is mathematically transformable to *d*. When this is done, the dichotomous variable would represent the independent variable (i.e., the treatment group versus the control group), whereas the continuous variable would represent the dependent variable from the experiment. The advantage of converting *d* to the r_{pb} is that "it can be inserted into a correlation matrix in which the intervention is then treated like any other variable" (Hunter & Schmidt, 1990b, p. 268).

Thus, the *d* to r_{pb} transformation makes treatment-based effect sizes comparable with correlation-based effect sizes. However, one of the key assumptions in the development of the transformation formula was that experimental effect sizes were independent of sample sizes (J. Cohen, 1988; Hunter & Schmidt, 1990b). At the conceptual level, experimental effect sizes are determined by the magnitude of the difference between the treatment and control groups and not the relative sample sizes of those groups. Yet, because of attrition or limited resources, experiments sometimes have treatment and control groups with unequal sample sizes (Becker, 1986). But because, theoretically, treatment effect sizes should be independent of these sample sizes, *d* can be transformed to the r_{pb} that would have been obtained if the treatment and control groups had equal *n*s (Becker, 1986; J. Cohen, 1977, 1988; Hunter & Schmidt, 1990b).

It is important to recognize that the rationale for correcting *d* for unequal sample sizes does not apply to r_{pb}s for true dichotomies. That is, despite their algebraic transformability, the assumptions underlying *d* are very different from the assumptions underlying the r_{pb}. Unlike *d*, which captures treatment-based effects in experiments, r_{pb}s were designed to reflect sample size differences in natural environments. Hunter and Schmidt (1990a) underscore this point, by noting the following:

> Conceptually, the effect size is normally thought of as independent of the sample sizes of the control and experimental groups. However, in a natural environment, the importance of a difference depends on how often it occurs. Since the point biserial correlation was originally developed for natural settings, it is defined so as to depend on the group sample sizes. (p. 274)

This is a critical issue because when studying employee turnover researchers must be concerned with naturally occurring frequencies of behavior. The different turnover base rates, across studies, that represent the frequency distribution of employee turnover, are supposed to be reflected in the magnitude of the r_{pb}s.

Indeed, Becker (1986) argued that r_{pb}s should be corrected only "when populations represented by the samples can be assumed to be equally numerous" (p. 5). According to Becker, this occurs in four special conditions: (a) in randomized experiments, (b) in blocked experiments, (c) when one control group serves as the comparison group for a number of different treatment groups, or (d) when equal sample sizes can be assumed in the population (i.e., gender; J. Cohen, 1988). However, none of these

conditions applies to turnover research. The first three conditions are irrelevant because they pertain to transformation of experimental effect sizes. Condition (d) also does not apply because there is no a priori, theoretical rational that would lead one to expect equal numbers of stayers and leavers across studies. Indeed, most research suggests that there should be differences in turnover base rates across studies.

For example, "under nearly all conditions the most accurate single predictor of labor turnover is the state of the economy" (March & Simon, 1958, p. 100). In particular, Eagly (1965) reported a correlation of −.84 between national unemployment and voluntary turnover, whereas Armknecht and Early (1972) demonstrated that 78% of the variance in voluntary quit rates could be attributed to changes in national unemployment. Thus, studies conducted under different levels of national unemployment are likely to have different turnover base rates.

However, it is not just national unemployment that influences quit rates. Steel and Griffeth (1989) argued that "behavioral scientists have implicitly treated the 'labor market' as a homogeneous construct, but in reality 'the labor market' is a heterogeneous mosaic of occupational and regional labor markets" (p. 848). Therefore, even studies conducted under identical national levels of unemployment may have different turnover base rates because of differences in regional, industrial, or occupational labor markets.

Finally, even if unemployment is held constant, real situational differences across organizations will influence turnover base rates. Evidence for this conclusion comes from a longitudinal, multiorganizational study that tracked the level of organizational turnover for a homogeneous group of workers, all drawn from the same labor pool and geographical region. During the 8 years of their study, Miller and van der Merwe (1982) found that although the overall economy accelerated or slowed the absolute level of turnover for each company in their study, there was little change in the rank ordering of companies by annual separation rate. Relative to each other, and regardless of changes in overall economic conditions, the high-turnover companies continued to have high turnover, whereas the low-turnover companies consistently had low turnover. "This replicated finding strongly suggests that the stability of the relative ranking of company turnover is determined by internal institutional factors rather than external organizational forces" (Miller & van der Merwe, 1982, p. 188).

In summary, different rates of study turnover are explainable and understandable. They are the norm not the exception. We conclude that differences in the level of turnover, and concomitant turnover variance across studies, should not be considered statistical artifacts nor should they be controlled for by correcting for unequal *n*s (Williams, 1990).

Logical Fallacies of Correcting Turnover Correlations

In most airplanes, there is a time lag between the instant in which the plane changes its position in airspace and the accurate registration of this change by the plane's instruments. Pilots who do not wait for accurate readings to register on their flight gauges are said to be "chasing gauges" because they end up making flight path adjustments based on incorrect information. In our view, correcting turnover r_{pb}s is akin to chasing gauges. Researchers who rely on corrected turnover r_{pb}s assume they accurately estimate what the r_{pb} would have been if the turnover rate was 50%. Yet, like airplane instruments that have not had time to "catch up," corrected r_{pb}s will

usually be inaccurate. This can be demonstrated by highlighting two logical fallacies inherent in the correction of turnover r_{pb}s.

Logical Fallacy: Turnover Correction Formulas are Predictive

Kemery and Dunlap (1989) state that "the obtained value . . . is the estimated point-biserial correlation that would have obtained *if the researcher had waited until 50% of the sample had terminated employment* [italics added]" (p. 490). This is a logical fallacy. In fact, turnover r_{pb} correction formulas are postdictive not predictive. Mathematical proofs aside, what correction formulas do is tell you what the specific r_{pb} would have been in the sample under study if the turnover rate actually had been 50%, given all other existing differences between leavers and stayers. This is very different from a predictive approach that would involve estimating what the r_{pb} would be when the turnover rate eventually hit 50%. The difference, of course, is that "all other existing differences between stayers and leavers" would have to remain unchanged. In our view, this is a weak assumption and, if not found to be true, would clearly affect the accuracy of the r_{pb}s estimated by turnover correction formulas.

To exemplify how weak this assumption is, it is instructive to look at some of the data from a turnover study by Mobley, Hand, Baker, and Meglino (1979). Mobley et al. found that military recruits who had graduated from recruit training had greater intentions to complete reenlistment ($\bar{x} = 3.07$, $SD = 2.75$, $n = 1{,}345$) than those who eventually dropped out ($\bar{x} = 2.75$, $SD = 1.24$, $n = 176$). In the sample under investigation, the r_{pb} between reenlistment intentions and turnover was –.19. However, Kemery et al.'s (1988) correction formula suggests that the r_{pb} will increase to –.25 when turnover hits 50%. Yet, for this predicted r_{pb} to be accurate, the means and standard deviations for the reenlistment intentions variable that Mobley et al. obtained for 1,345 stayers and 176 leavers must be exactly the same when turnover reaches 50%. This seems highly unlikely, given that 584 of the original 1,345 stayers will have become leavers (761 stayers and 760 leavers) by the time turnover reaches 50%. Thus, the –.25 r_{pb} predicted by the correction formula probably will not be accurate.

Logical Fallacy: Maximum r_{pb}s Occur When the Base Rate Hits 50%

The second logical fallacy follows from the first and suggests that turnover r_{pb}s are largest when the turnover rate hits .50. This fallacy arises because turnover correction formulas ignore important effects due to the length of the study measurement window (see L. H. Peters & Sheridan, 1988). In particular, considerable time would have to pass in most studies before the sample turnover base rate increased to .50. This increased time span would weaken, as opposed to strengthen, observed r_{pb}s. To illustrate, Mobley, Horner, and Hollingsworth (1978) reported a r_{pb} of .49 between intentions to quit and turnover for a 47-week study in an organization that had 10% turnover. When corrected for unequal ns, this r_{pb} rises to .68. Thus, the corrected correlation suggests that the relationship between intention to quit and turnover will strengthen if turnover increases from 10% to 50%. Yet, assuming a constant rate of attrition, it would take approximately 5 years for the turnover rate to reach this level. And because most turnover studies are about a year in length, and because the average rate of turnover is approximately 21% (Steel et al., 1990), the average turnover study would have to be 2½ years long for the turnover rate to reach 50%.

When this is the case, why would variables like job satisfaction, organizational commitment, or intentions to quit be better predictors of stay/leave decisions after 2 or more years than after 1 year? In fact, as study times increase, changes in organizational circumstances (e.g., the introduction of new technology, management systems, work design, or reward systems) or changes in employees themselves (e.g., persons moving from one career stage to another) might make such measures (taken at Time 1) increasingly irrelevant to later turnover decisions and behavior (made at a considerably more distant Time 2).

Indeed, results from three turnover meta-analyses support these arguments. Steel and Ovalle (1984, pp. 682-683) found that the length of a turnover study was negatively related to the size of the correlation between turnover and intentions to quit, overall satisfaction, work satisfaction, and organizational commitment. Carsten and Spector (1987) found similar results, and concluded that "the intention-turnover and particularly the job satisfaction-turnover relations weaken as the time period of turnover data collection becomes longer" (p. 379). A. Cohen (1993) found that the mean r_{pb} between organizational commitment and voluntary turnover was –.33 for studies lasting 6 months or less, whereas the mean r_{pb} was –.20 for studies lasting longer than 6 months. So, despite the predicted increase in turnover r_{pb}s suggested by turnover correction formulas, these arguments and data clearly suggest that turnover r_{pb}s would most likely shrink, not grow, by the time turnover eventually reached 50% in most studies.

Turnover Correlations and .80: A False Ceiling

In comparison to PM correlations between two continuous variables, which have maxima at ±1.00, it is generally accepted that the maximum value of r_{pb}s is ±.798 (Nunnally, 1978). However, it is not widely understood that the .798 limitation only applies to r_{pb}s between two continuous variables when *one of the continuous variables has been artificially dichotomized.* When this is the case, the maximum size of the r_{pb} is limited to .798 because there are mathematical limits on how far apart the two group means can be.

> The largest difference occurs if the continuous variables are perfectly correlated. In this case, the two groups are the top and bottom halves of a normal distribution. The largest difference for such groups occurs if the split is at the mean, a difference of 1.58 standard deviations. (Hunter & Schmidt, 1990a, p. 270)

There is, however, no such mathematical limitation when the r_{pb} is between a continuous variable and a truly dichotomous variable (Hunter & Schmidt, 1990b). Instead, just like with any PM correlation, the size of the r_{pb} is limited only by the similarity or dissimilarity of the distributions of these variables. In fact, Karabinus (1975) reasoned that the presumed .798 limitation on r_{pb}s would not hold when the continuous and dichotomous distributions are made more similar:

> While it is recognized that it is impossible to obtain a perfect correlation with the r-point biserial (this can occur only with two continuous variables or two dichotomous variables), it is possible to more nearly approach ±1.00 by making the shape of the continuous variable more like the shape of the dichotomous one. This can be done by having a bimodal but symmetrical distribution on the continuous variable, which for each part of the dichotomous variable would be as peaked as possible. (p. 278)

Table 1, from Karabinus (1975, p. 280), shows r_{pb}s for ns of 10, 30, and 100; with p values (i.e., turnover base rates) of .50 and .60; with overlap and no overlap on the continuous variables; and with five different distributions on the continuous variable (i.e., normal, rectangular, bimodal-normal, bimodal-peaked, and bimodal-peaked and skewed), each of which was constructed to most closely approximate the distribution of the dichotomous variables. An examination of the data in Table 1 does not support the widely cited mathematical limitation of .798 for r_{pb}s. Indeed, the average correlation in Table 1 was .86, 50 out of the 60 r_{pb}s were greater than or equal to .798, 25 were greater than or equal to .90, and 1 was as high as .978.

Although these results support our position, we are not claiming that they are typical. Clearly, they are not, given that the data were expressly manufactured to maximize the similarity between the distributions of the continuous and dichotomous variables. Nonetheless, it is worth taking a closer look at the results obtained from the more typical turnover data in Table 1. For example, 4 of the 12 correlations derived from normally distributed data (Table 1, columns 3 and 4) exceeded .798. However, Karabinus (1975) explained that "those coefficients > .798 under the 'Normal distribution' occurred because the distributions were not perfectly normal" (p. 279). This seems to suggest that .798 is a practical, but not mathematical, limit for r_{pb}s.

However, we direct readers to the column marked "Bimodal-Normal." A bimodal-normal distribution assumes that the continuous variable is normally distributed for each value of the dichotomous variable. For example, if leavers are less satisfied than stayers, imagine two normal curve distributions that are separated by the mean difference in satisfaction scores. As the difference becomes larger, the overall distribution of satisfaction scores deviates from normality, but the separate satisfaction distributions for stayers and leavers remain normally distributed (Bass & Ager, 1991). We believe that dual normal distributions are representative of typical turnover data. Walker and Lev (1953, p. 265), Karabinus (1975, p. 277), and Bass and Ager (1991, p. 595) went even further, claiming that r_{pb}s and conventional significance tests for r_{pb}s were developed under the assumption of dual normal distributions. All of the correlations derived from the bimodal-normal distribution exceeded .798. Moreover, eight of those correlations exceeded .90. These data suggest that the widely claimed .798 limit is neither a mathematical limit nor a practical limit for turnover r_{pb}s.

Again, the key issue is not how often observed r_{pb}s will exceed .798. In behavioral research, uncorrected correlations larger than .40 are rare. The issue is, What is the mathematical limit of r_{pb}s? Because the justification for correcting r_{pb}s in turnover research is predicated on a mathematical limit of .798, and because the evidence shows that the limitation is closer to .978, there appears to be no compelling statistical reason to apply correction formulas to turnover r_{pb}s. Karabinus's (1975) data not only show that .798 is a false ceiling but also that the magnitude of r_{pb}s is determined by the similarity of the continuous distributions across the dichotomous groups and not just by the base rate. Furthermore, Karabinus's results are neither isolated nor extraordinary, as other studies have also found point-biserial correlations greater than .80 (e.g., Adams, 1960; Bowers, 1972). In conclusion, r_{pb}s have no "limitation that is not present in any of the other members of the Pearson family (except for the natural limitation of the variables themselves)" (Karabinus, 1975, p. 282).

Table 1
Point Biserial Coefficients With Different Shapes of Continuous Variable

		Normal		*Rectangular*		*Bimodal-Normal*		*Bimodal-Peaked*		*Bimodal-Peaked Skewed*	
n	Y *on* X	p = .5	p = .6	p = .5	p = .6	p = .5	p = .6	p = .5	p = .6	p = .5	p = .6
10	No overlap	.809	.539	.878	.866	.921	.919	.921	.919	.882	.896
	Overlap	.730	.745	.849	.722	.845	.840	.845	.840	.781	.813
30	No overlap	.805	.817	.870	.853	.927	.919	.951	.934	.936	.916
	Overlap	.696	.772	.857	.833	.889	.889	.926	.909	.904	.871
100	No overlap	.802	.706	.870	.853	.950	.948	.978	.977	.960	.960
	Overlap	.796	.769	.861	.803	.934	.934	.970	.968	.946	.946

Source. Karabinus (1975).

Inferences From Point-Biserial Correlations

Karabinus's (1975) results are also important because they address the most critical inferential issue in the debate about correcting turnover correlations: whether correlations obtained from different turnover base rates are directly comparable. Those who recommend correcting turnover r_{pb}s believe that they are not comparable, because different turnover base rates produce different limits on the maximum size of turnover r_{pb}s. Steel et al. (1990, pp. 180-181), for example, explain,

> Suppose two different studies report correlations of .58 between organizational commitment and turnover. May a reader safely conclude that the correlations are identical estimates of the underlying relationship between the two latent variables, organizational commitment and turnover? That conclusion is warranted if the studies feature comparable base rates, ceteris paribus. But if they do not, the correlations may not be directly comparable. Suppose the two studies have base rates of .06 and .50, respectively. A correlation of .58 from a study with a base rate of .06 signifies a far stronger relationship than a correlation of .58 from a study with a base rate of .50. Although the former relationship corresponds to the maximum feasible correlation, given a base rate of .06 (Thorndike, 1978), the latter statistic represents a moderate relationship since the absolute limits of point-biserial correlations normally range from .00 and ±.80. (Thorndike, 1978)

Thus, Steel et al. (1990) clearly suggest that the strength of turnover correlations should not be judged by the *absolute* size of the r_{pb}s but by the *relative* size of the observed correlation at base rate p compared to the maximum correlation that could be obtained at p. Although slightly different, other authors also have recommended specific r_{pb}/r_{pbmax} procedures (Bass & Ager, 1991; Kemery, Dunlap, & Bedeian, 1989). Using the relative r_{pb}/r_{pbmax} criterion, the .58 r_{pb} obtained at a .06 base rate is said to be a stronger index of relationship (r_{pb}/r_{pbmax} = .58/.58 = 1.00) than the .58 r_{pb} obtained at a .50 base rate (r_{pb}/r_{pbmax} = .58/.80 = .725).

However, there is no reason to employ the relative r_{pb}/r_{pbmax} criterion when judging and comparing the strength of turnover r_{pb}s obtained under different turnover base rates. This is because correlations are expressed in standard deviation units and, as such, are by definition unit-free measures of linear relationship (J. Cohen, 1983; Lord & Novick, 1968; Nunnally, 1978). Therefore, regardless of the measurement scales used, and regardless of differences in the variability of scores (i.e., turnover base rates), correlations have the same statistical interpretation from study to study (J. Cohen, 1983). For example, no matter what the turnover base rate is (.06 or .50), a turnover correlation of .58 means that a 1.00 standard deviation change in a turnover antecedent (e.g., organizational commitment) will be associated with a .58 standard deviation change in turnover. Furthermore, the relationship is not stronger in the .06 base rate than in the .50 base rate. The relationship, expressed in standard deviation units, is equally strong in both studies.

What others might meaningfully imply when they state that turnover correlations are not comparable is that, even when correlations are the same, the magnitude of that change in the absolute rate of turnover will be different under different turnover base rates. This is true because predicted turnover scores are a function of both the turnover r_{pb} and the variance in the turnover criterion, and the criterion variance with regard to turnover, in turn, is a function of its base rate. In other words, if the correlation is the

same in two settings, the larger absolute change will occur in the setting that has the larger criterion variance.

To illustrate, the average correlation between intentions to quit and turnover is .45 (Steel & Ovalle, 1984). When the turnover rate is 90%, the turnover standard deviation is .30 (i.e., $\sqrt{.90 \times .10}$). A 1.00 standard deviation decrease in intentions to quit should reduce turnover .135 (i.e., $r_{pb} \times SD = .45 \times .30$), from .90 to .765. Yet, when the turnover rate is 50%, the turnover standard deviation will be .50 (i.e., $\sqrt{.50 \times .50}$), and a 1.00 standard deviation decrease in intentions to quit would reduce turnover .225 (i.e., $r_{pb} \times SD = .45 \times .50$), from .50 to .275. So, although the correlations are the same, the larger turnover standard deviation yields the larger absolute change.

However, as unit free measures, correlations do not estimate linear relationships in terms of the original scale of measurement (i.e., the standard deviation of the criterion variable). The appropriate statistic for this purpose is an unstandardized regression coefficient (J. Cohen & Cohen, 1983). By contrast, correlations, which are expressed in standard deviation units, have the same statistical interpretation from study to study (J. Cohen, 1983).

Conclusions

We have argued that (a) turnover correlation correction formulas cannot correct for poor construct validity when a dichotomous operational variable is used to measure turnover; (b) correction of turnover r_{pb}s violates the original assumptions and purposes of correction formulas; (c) corrected correlations are not predictive but postdictive; (d) in most instances, turnover r_{pb}s should get smaller rather than larger by the time the turnover rate eventually reaches 50%; (e) r_{pb}s are not limited to a maximum of .798; and (f) because correlations are expressed in standard deviation units, differences in turnover base rates do not affect the study-to-study statistical interpretation of turnover r_{pb}s. Consequently, we conclude that there are serious logical, statistical, and inferential problems associated with correcting turnover r_{pb}s for dichotomization and unequal *n*s.

In short, there is simply no need to apply correction formulas to turnover r_{pb}s. As PM correlations, turnover r_{pb}s have the same meaning, limitations, and sampling errors as do PM correlations between continuous variables.

Notes

1. We are not suggesting that measurement specification alone allows one to conclude that there is construct validity between constructs and measures. Indeed, numerous attempts have been made to refine the match between dichotomous turnover measures and the stay/quit turnover construct. For example, because voluntary turnover has been defined as a "choice behavior" (Mobley, 1982), Abelson (1987) found much better support for turnover models when employees who had to quit (i.e., unavoidable turnover due to sickness, spouse moving because of job changes, etc.) were removed from the analysis. We just want to emphasize the importance of clear, precise definitions for constructs and variables (Schwab, 1980, pp. 12-13).

2. Of course, this is not an issue in studies that use cohort samples to investigate how organizational entry, socialization, recruitment sources, or realistic job previews influence employee tenure.

3. We also found that the terms *turnover propensity*, or *propensity to leave* have been used in turnover studies to represent *intentions to turnover*. We view this as another example of the mismatch between turnover constructs and turnover variables.

4. These faster methods were still incredibly slow by today's standards. Even with his improved procedures, Dunlap (1936) estimated that it still took 250 working hours to compute 3,480 correlations!

5. One widespread misunderstanding among researchers is that r_{pb}s are different from standard PM correlations. In fact, r_{pb}s *are* PM correlations! According to Nunnally (1978), "the numerical result obtained by applying the regular PM formula is exactly the same as that which would be obtained from the shortcut version r_{pb}" (p. 134).

6. Statistical artifacts do not account for why point-biserial correlations tend to be larger when the dichotomous variable has a 50-50 split. There is a much simpler explanation, namely, that "the shape of a dichotomous distribution is most similar to that of a normal distribution when p is .50" (Nunnally, 1978, p. 145). However, if the continuous variable has a non-normal distribution, then r_{pb}s will be largest at points other than a 50-50 split (Bass & Ager, 1991). In other words, r_{pb}s, like PM correlations between continuous variables, just reflect the similarity between two variable distributions.

References

Abelson, M. A. (1987). Examination of avoidable and unavoidable turnover. *Journal of Applied Psychology, 72*, 382-386.

Adams, J. F. (1960). The effect of non-normally distributed criterion scores on item analysis techniques. *Educational and Psychological Measurement, 20*, 317-320.

Armknecht, P. A., & Early, J. F. (1972). Quits in manufacturing: A study of their causes. *Monthly Labor Review, 95*, 31-37.

Bass, A. R., & Ager, J. (1991). Correcting point-biserial turnover correlations for comparative analysis. *Journal of Applied Psychology, 76*, 595-598.

Becker, G. (1986). Correcting the point-biserial correlation for attenuation owing to unequal sample size. *Journal of Experimental Education, 55*, 5-8.

Bowers, J. (1972). A note on comparing r-biserial and r-point biserial. *Educational and Psychological Measurement, 32*, 771-775.

Carsten, J. M., & Spector, P. E. (1987). Unemployment, job satisfaction, and employee turnover: A meta-analytic test of the Muchinsky model. *Journal of Applied Psychology, 72*, 374-381.

Cohen, A. (1993). Organizational commitment and turnover: A meta-analysis. *Academy of Management Journal, 36*, 1140-1157.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249-253.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J., & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Dunlap, J. W. (1936). Note on computation of bi-serial correlations in item evaluation. *Psychometrika, 1*, 51-58.

Eagly, R. V. (1965). Market power as an intervening mechanism in Phillips Curve analysis. *Economics, 32*, 48-64.

Glass, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education, 5*, 351-379.

Glass, G. V., & Stanley, J. C. (1970). *Statistical methods in education and psychology.* Englewood Cliffs, NJ: Prentice Hall.

Hulin, C. L. (1990). Adaptation, persistence, and commitment in organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, pp. 445-505). Palo Alto, CA: Consulting Psychologists Press.

Hunter, J. E., & Schmidt, F. L. (1990a). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology, 75,* 334-349.

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

Jaspen, N. (1946). Serial correlation. *Psychometrika, 11,* 23-30.

Karabinus, R. A. (1975). The r-point biserial limitation. *Educational and Psychological Measurement, 35,* 277-282.

Kemery, E. R., & Dunlap, W. P. (1989). A table of correction factors for estimating point-biserial correlations with equal category proportions. *Psychological Reports, 64,* 487-492.

Kemery, E. R., Dunlap, W. P., & Bedeian, A. G. (1989). The employee separation process: Criterion-related issues associated with tenure and turnover. *Journal of Management, 15,* 417-424.

Kemery, E. R., Dunlap, W. P., & Griffeth, R. W. (1988). Correction for variance restriction in point-biserial correlations. *Journal of Applied Psychology, 73,* 688-691.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA.: Addison-Wesley.

March, J. G., & Simon, H. A. (1958). *Organizations.* New York: John Wiley.

McCall, M. W. Jr., & Bobko, P. (1990). Research methods in the service of discovery. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, pp. 381-418). Palo Alto, CA: Consulting Psychologists Press.

McEvoy, G. M., & Cascio, W. F. (1987). Do good or poor performers leave? A meta-analysis of the relationship between performance and turnover. *Academy of Management Journal, 30,* 744-762.

Miller, S., & van der Merwe, R. (1982). Consistency in inter-organizational labour turnover: A follow-up. *Journal of Occupational Psychology, 55,* 185-189.

Mobley, W. H. (1982). *Employee turnover: Causes, consequences, and control.* Reading, MA: Addison-Wesley.

Mobley, W. H., Hand, H. H., Baker, R. L., & Meglino, B. M. (1979). Conceptual and empirical analysis of military recruit training. *Journal of Applied Psychology, 64,* 10-18.

Mobley, W. H., Horner, S. O., & Hollingsworth, A. T. (1978). An evaluation of precursors of hospital employee turnover. *Journal of Applied Psychology, 63,* 408-414.

Mowday, R. T., Porter, L. T., & Steers, R. M. (1982). *Employee-organization linkages: The psychology of commitment, absenteeism, and turnover.* New York: Academic Press.

Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw-Hill.

Peters, L. H., & Sheridan, J. E. (1988). Turnover research methodology: A critique of traditional designs and a suggested survival model alternative. In K. M. Rowland & G. R. Ferris (Eds.), *Research in personnel and human resources management* (Vol. 6, pp. 231-262). Greenwich, CT: JAI.

Peters, C. C., & Van Voorhis, W. R. (1940). *Statistical procedures and their mathematical bases.* New York: McGraw-Hill.

Rosse, J. G. (1988). Relations among lateness, absence, and turnover: Is there a progression of withdrawal? *Human Relations, 36,* 324-347.

Rosse, J. G., & Hulin, C. L. (1985). Adaptation to work: An analysis of employee health, withdrawal, and change. *Organizational Behavior and Human Decision Processes, 36,* 324-347.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). Impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology, 64,* 609-626.

Schwab, D. P. (1980). Construct validity in organizational behavior. *Research in Organizational Behavior, 2,* 3-43.

Steel, R. P., & Griffeth, R. W. (1989). The elusive relationship between perceived employment opportunity and turnover behavior: A methodological or conceptual artifact? *Journal of Applied Psychology, 74,* 846-854.

Steel, R. P., & Ovalle, N. K. (1984). A review and meta-analysis of research on the relationship between behavioral intentions and employee turnover. *Journal of Applied Psychology, 69,* 673-686.

Steel, R. P., Shane, G. S., & Griffeth, R. W. (1990). Correcting turnover statistics for purposes of comparative analysis. *Academy of Management Journal, 33,* 179-187.

Thorndike, R. L. (1978). *Correlational procedures for research.* New York: Gardner.

Walker, H. M., & Lev, J. (1953). *Statistical inference.* New York: Holt, Rinehart & Winston.

Wanous, J. P., Poland, T. D., Premack, S. L., & Davis, K. S. (1992). The effects of met expectations on newcomer attitudes and behaviors: A review and meta-analysis. *Journal of Applied Psychology, 77,* 288-297.

Williams, C. R. (1990). Deciding when, how, and if to correct turnover correlations, *Journal of Applied Psychology, 75,* 732-737.

Williams, C. R., & Livingstone, L. P. (1994). Another look at the relationship between performance and voluntary turnover. *Academy of Management Journal, 37,* 269-298.