Psychological Methods
2002, Vol. 7, No. 1, 19–40

1082-989X/02/$5.00 DOI: 10.1037//1082-989X.7.1.19

On the Practice of Dichotomization of Quantitative Variables

Robert C. MacCallum, Shaobo Zhang, Kristopher J. Preacher, and Derek D. Rucker
Ohio State University

The authors examine the practice of dichotomization of quantitative measures, wherein relationships among variables are examined after 1 or more variables have been converted to dichotomous variables by splitting the sample at some point on the scale(s) of measurement. A common form of dichotomization is the median split, where the independent variable is split at the median to form high and low groups, which are then compared with respect to their means on the dependent variable. The consequences of dichotomization for measurement and statistical analyses are illustrated and discussed. The use of dichotomization in practice is described, and justifications that are offered for such usage are examined. The authors present the case that dichotomization is rarely defensible and often will yield misleading results.

We consider here some simple statistical procedures for studying relationships of one or more independent variables to one dependent variable, where all variables are quantitative in nature and are measured on meaningful numerical scales. Such measures are often referred to as *individual-differences measures,* meaning that observed values of such measures are interpretable as reflecting individual differences on the attribute of interest. It is of course straightforward to analyze such data using correlational methods. In the case of a single independent variable, one can use simple linear regression and/or obtain a simple correlation coefficient. In the case of multiple independent variables, one can use multiple regression, possibly including interaction terms. Such methods are routinely used in practice.

However, another approach to analysis of such data is also rather widely used. Considering the case of one independent variable, many investigators begin by converting that variable into a dichotomous variable by splitting the scale at some point and designating individuals above and below that point as defining two separate groups. One common approach is to split the scale at the sample median, thereby defining high and low groups on the variable in question; this approach is referred to as a median split. Alternatively, the scale may be split at some other point based on the data (e.g., 1 standard deviation above the mean) or at a fixed point on the scale designated a priori. Researchers may dichotomize independent variables for many reasons—for example, because they believe there exist distinct groups of individuals or because they believe analyses or presentation of results will be simplified. After such dichotomization, the independent variable is treated as a categorical variable and statistical tests then are carried out to determine whether there is a significant difference in the mean of the dependent variable for the two groups represented by the dichotomized independent variable. When there are two independent variables, researchers often dichotomize both and then analyze effects on the dependent variable using analysis of variance (ANOVA).

There is a considerable methodological literature examining and demonstrating negative consequences of dichotomization and firmly favoring the use of regression methods on undichotomized variables. Nevertheless, substantive researchers often dichotomize independent variables prior to conducting analyses. In this article we provide a thorough examination of the practice of dichotomization. We begin with numerical examples that illustrate some of the consequences of dichotomization. These include loss of information about individual differences as well as havoc with

Robert C. MacCallum, Shaobo Zhang, Kristopher J. Preacher, and Derek D. Rucker, Department of Psychology, Ohio State University.

Shaobo Zhang is now at Fleet Credit Card Services, Horsham, Pennsylvania.

Correspondence concerning this article should be addressed to Robert C. MacCallum, Department of Psychology, Ohio State University, 1885 Neil Avenue, Columbus, Ohio 43210-1222. E-mail: maccallum.1@osu.edu

regard to estimation and interpretation of relationships among variables. We then examine the dichotomization approach in terms of issues of measurement of individual differences and statistical analysis. We next review current practice, providing evidence of common usage of dichotomization in applied research in fields such as social, developmental, and clinical psychology, and we examine and evaluate justifications offered by users and defenders of this procedure. Overall, we present the case that dichotomization of individual-differences measures is probably rarely justified from either a conceptual or statistical perspective; that its use in practice undoubtedly has serious negative consequences; and that regression and correlation methods, without dichotomization of variables, are generally more appropriate.

Numerical Examples

Example Using One Independent Variable

We begin with a series of numerical examples using simulated data to illustrate and distinguish between the regression approach and the dichotomization approach. Raw data for these numerical examples can be obtained from Robert C. MacCallum's Web site (http://quantrm2.psy.ohio-state.edu/maccallum/). Let us first consider the case of one independent variable, *X,* and one dependent variable, *Y.* We defined a simulated population in which the two variables followed a bivariate normal distribution with a correlation of $\rho_{XY} = .40$. Using a procedure described by Kaiser and Dickman (1962), we then drew a random sample ($N = 50$ observations) from this population. Sample data were scaled such that $M_X = 10$, $SD_X = 2$, $M_Y = 20$, and $SD_Y = 4$. A scatter plot of the sample data is displayed in Figure 1. We assessed the linear relationship between *X* and *Y* by obtaining the sample correlation coefficient, which was found to be $r_{XY} = .30$, with a 95% confidence interval of (.02, .53). The squared correlation ($r^2_{XY} = .09$) indicated that about 9% of the variance in *Y* was accounted for by its linear relationship with *X.* A test of the null hypothesis that $\rho_{XY} = 0$ yielded $t(48) = 2.19$, $p = .03$, leading to rejection of the null hypothesis and the conclusion that there is evidence in the sample of a nonzero linear relationship in the population. Of course, the outcome of this test was foretold by the confidence interval, because the confidence interval did not contain the value of zero.

We then conducted a second analysis of the relationship between *X* and *Y,* beginning by converting *X* into a dichotomous variable. We split the sample at the median of *X,* yielding high and low groups on *X,* each containing 25 observations. The resulting dichotomized variable is designated X_D. A scatter plot of the data following dichotomization of *X* is shown in Figure 2. The relationship between *X* and *Y* was then evaluated by testing the difference between the *Y* means for the high and low groups on *X.* Those means were found to be 21.1 and 19.4, respectively, and a

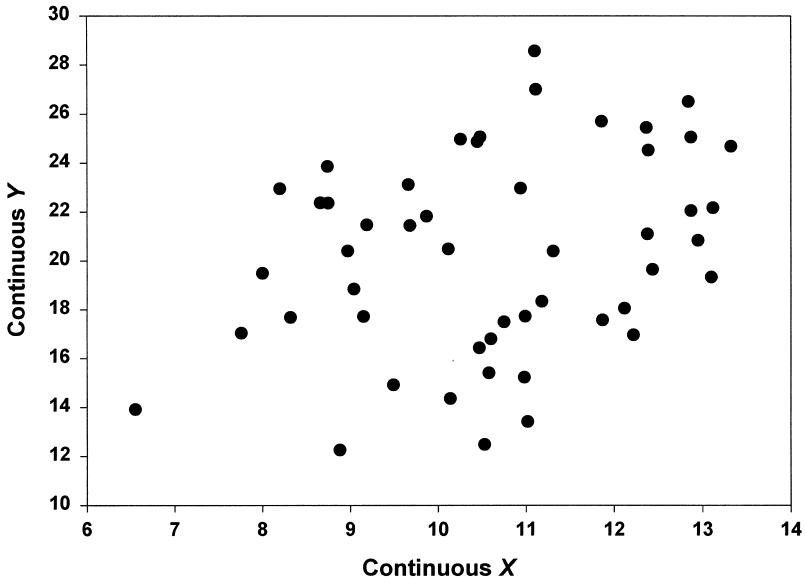

Figure 1. Scatter plot of raw data for example of bivariate relationship.

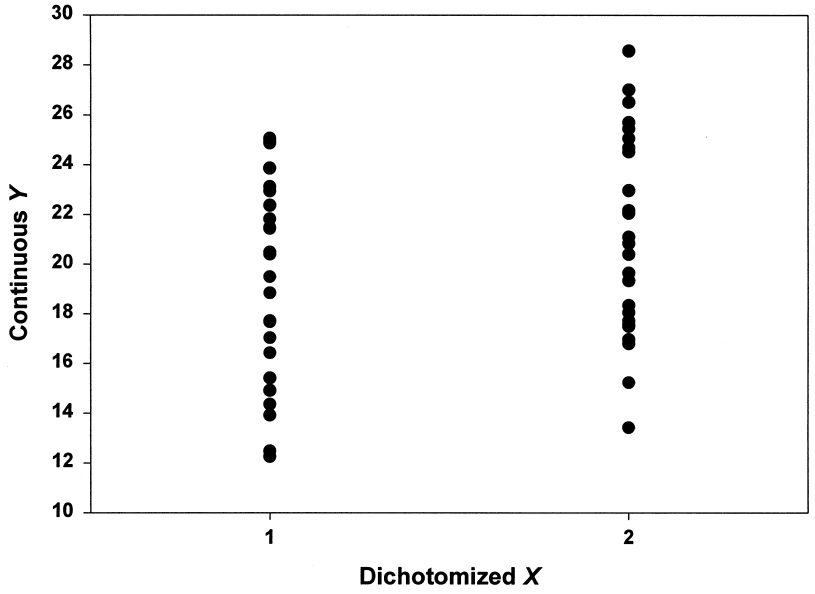

Figure 2. Scatter plot for example of bivariate relationship after dichotomization of *X.* (Note that there is some overlapping of points in this figure.)

test of the difference between them yielded $t(48) = 1.47$, $p = .15$, resulting in failure to reject the null hypothesis of equal population means. From a correlational perspective, the correlation after dichotomization was $r_{X_D Y} = .21$ ($r^2_{X_D Y} = .04$), with a 95% confidence interval of (−.07, .46). Again, the confidence interval implies the result of the significance test, this time indicating a nonsignificant relationship because the interval did include the value of zero. A comparison of the results of analysis of the association between *X* and *Y* before and after dichotomization of *X* shows a distinct loss of effect size and loss of statistical significance; these issues are addressed in detail later in this article.

The dichotomization procedure can be taken one step further by splitting both *X* and *Y,* thereby yielding a 2 × 2 frequency table showing the association between X_D and Y_D. In the present example, this approach yielded frequencies of 13 in the low–low and high–high cells and frequencies of 12 in the low–high and high–low cells. The corresponding test of association $\chi^2_1(1, N = 50) = 0.08$, $p = .78$, showed a nonsignificant relationship between the dichotomized variables, which was also indicated by the small correlation between them, $r_{X_D Y_D} = .06$ with a 95% confidence interval of (−.22, .33). Note that the dichotomization of both *X* and *Y* has further eroded the strength of association between them.

Example Using Two Independent Variables

We next consider an example where there are two independent variables, X_1 and X_2. In practice such data would appropriately be treated by multiple regression analysis. However, a common approach instead is to convert both X_1 and X_2 into dichotomous variables and then to conduct ANOVA using a 2 × 2 factorial design. We provide an illustration of both methods. Following procedures described by Kaiser and Dickman (1962), we constructed a sample ($N = 100$ observations) from a multivariate normal population such that the sample correlations among the three variables would have the following pattern: $r_{X_1 Y} = .70$, $r_{X_2 Y} = .35$, and $r_{X_1 X_2} = .50$. To examine the relationship of *Y* to X_1 and X_2, we first conducted multiple regression analyses. Without loss of generality, regression analyses were conducted on standardized variables, Z_Y, Z_1, and Z_2. Using a linear regression model with no interaction term, we obtained a squared multiple correlation of .49 with a 95% confidence interval of (.33, .61).[1] This squared

[1] Confidence intervals for squared multiple correlations are useful but are not commonly provided by commercial software for regression analysis. A free program offered by James Steiger, available from his Web site (http://

multiple correlation was statistically significant, $F(2, 97) = 46.25$, $p < .01$. The corresponding standardized regression equation was

$$\hat{Z}_Y = .70(Z_1) + .00(Z_2).$$

The coefficient for Z_1 was statistically significant, $\beta_1 = .70$, $t(1) = 8.32$, $p < .01$, whereas the coefficient for Z_2 obviously was not significant, $\beta_2 = .00$, $t(1) = 0.00$, $p = 1.00$, indicating a significant linear effect of Z_1 on Z_Y, but no significant effect of Z_2. Inclusion of an interaction term $Z_3 = Z_1Z_2$ in the regression model increased the squared multiple correlation slightly to .50. The corresponding regression equation was

$$\hat{Z}_Y = .72(Z_1) - .01(Z_2) - .07(Z_3).$$

The standardized regression coefficient for Z_1 was statistically significant, $\beta_1 = .72$, $t(1) = 8.38$, $p < .01$, whereas the other two coefficients were not, $\beta_2 = -.01$, $t(1) = -0.12$, $p = .96$, and $\beta_3 = -.07$, $t(1) = -1.11$, $p = .27$, indicating no significant linear effect of Z_2 and no significant interaction.

We then conducted a second analysis on the same data, beginning by dichotomizing X_1 and X_2 by splitting both at the median to create X_{1D} and X_{2D}. Results of a two-way ANOVA yielded a significant main effect for X_{1D}, $F(1, 96) = 42.50$, $p < .01$; a significant main effect for X_{2D}, $F(1, 96) = 5.26$, $p = .02$; and a nonsignificant interaction, $F(1, 96) = 0.19$, $p = .67$. Total variance accounted for by these effects was .40. Of special note in these results is the presence of a significant main effect of X_{2D} that was not present in the regression analyses but arose only after both independent variables were dichotomized. This phenomenon is discussed further later in this article. It is also noteworthy that total variance accounted for was reduced from .50 prior to dichotomization to .40 after dichotomization.

Summary of Examples

From these few examples it should be clear that there exist potential problems when quantitative independent variables are dichotomized prior to analysis of their relationship to dependent variables. The example with one independent variable showed a loss of effect size and of statistical significance following dichotomization of *X*. The example with two independent variables showed a significant main effect following dichotomization that did not exist prior to dichotomization. Although many other cases and examples could be examined, these simple illustrations reveal potentially serious problems associated with dichotomization. If phenomena such as those just illustrated would be common in practice, then dichotomization of variables probably should not be done unless rigorously justified. In the following section we examine these phenomena closely, focusing on the impact of dichotomization on measurement and representation of individual differences as well as on results of statistical analyses.

Measurement and Statistical Issues Associated With Dichotomization

Representing Individual Differences

We first consider the impact of dichotomization on the measurement and representation of individual differences associated with a variable of interest. Suppose we have a single independent variable, *X*, measured on a quantitative scale, and we observe a sample of individuals who vary along that scale, and suppose the resulting distribution is roughly normal. A distribution of this type is illustrated in Figure 3a, with four specific individuals (A, B, C, D) indicated along the x-axis. Assuming approximately interval properties for the scale of *X,* these individuals can be compared with each other with respect to their standing on *X.* For instance, Individuals B and C are more similar to each other than are Individuals A and D. Individuals A and B are different from each other, and that difference is greater than the difference between B and C. When observed in empirical research, such differences and comparisons among individuals would seem to be relevant to the understanding of such a variable and its relationship to other variables. Researchers in psychology typically invest great effort in the development of measures of individual differences on characteristics and behaviors of interest. Instruments are developed so as to have adequate levels of reliability and validity, implying acceptable levels of precision in measuring individuals, and implying in turn that individual differences such as those just described are meaningful.

Now suppose *X* is dichotomized by a median split as illustrated in Figure 3b. Such dichotomization alters the nature of individual differences. After dichotomization, Individuals A and B are defined as equal as are Individuals C and D. Individuals B and C

www.interchg.ubc.ca/steiger/homepage.htm) computes such confidence intervals as well as other useful statistical information associated with correlation coefficients.

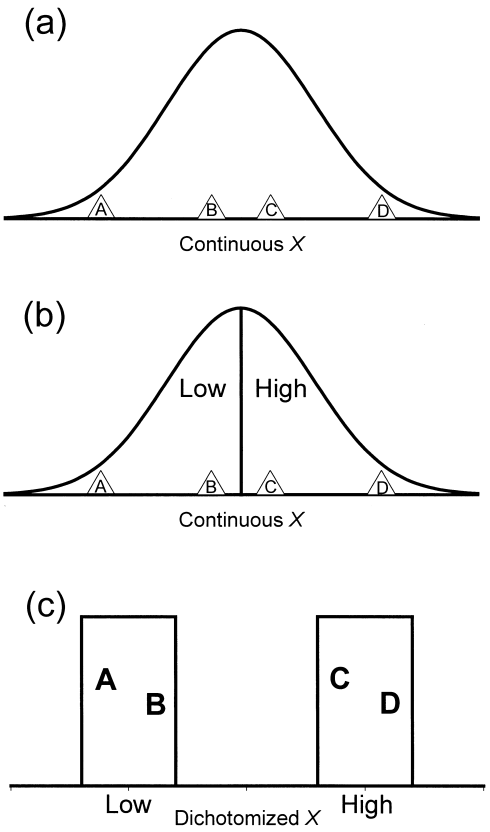

Figure 3. Measurement of individual differences before and after dichotomization of a continuous variable.

are different, even though their difference prior to dichotomization was smaller than that between A and B, who are now considered equal. Following dichotomization, the difference between A and D is considered to be the same as that between B and C. The median split alters the distribution of *X* so that it has the form shown in Figure 3c. Clearly, most of the information about individual differences in the original distribution has been discarded, and the remaining information is quite different from the original. Such an altering of the observed data must raise questions. What was the purpose of measuring individual differences on *X* only to discard much of that information by dichotomization? What are the consequences for the psychometric properties of the measure of *X*? And what is the impact on results of subsequent analyses of the relationship of *X* to other variables?

It seems that to justify such discarding of information, one would need to make one of two arguments. First, one might argue that the discarded information is essentially error and that it is beneficial to eliminate such error by dichotomization. The implication of such an argument would be that the true variable of interest is dichotomous and that dichotomization of *X* produces a more precise measure of that true dichotomy. An alternative justification might involve recognition that the discarded information is not error but that there is some benefit to discarding it that compensates for the loss of information. Both of these perspectives are discussed further later in this article.

Impact on Results of Statistical Analyses

Review of types of correlation coefficients and their relationships. Dichotomization of quantitative variables affects results of statistical analyses involving those variables. To examine these effects, it is necessary to understand the meaning of and relationships among several different types of correlation coefficients. A correlation between two quantitatively measured variables is conventionally computed as the common Pearson product–moment (PPM) correlation. A correlation between one quantitative and one dichotomous variable is a point-biserial correlation, and a correlation between two dichotomous variables is a phi coefficient. The point-biserial and phi coefficients are special cases of the PPM correlation. That is, if we apply the PPM formula to data involving one quantitative and one dichotomous variable, the result will be identical to that obtained using a formula for a point-biserial correlation. Similarly, if we apply the PPM formula to data involving two dichotomous variables, the result will be identical to that obtained using a formula for a phi coefficient. The point-biserial and phi coefficients are typically used in practice for analyses of relationships involving variables that are true dichotomies. For example, one could use a point-biserial correlation to assess the relationship between gender and extraversion, and one could use a phi coefficient to measure the relationship between gender and smoking status (smoker vs. nonsmoker).

Some variables that are measured as dichotomous variables are not true dichotomies. Consider, for example, performance on a single item on a multiple choice test of mathematical skills. The measured variable is dichotomous (right vs. wrong), but the underlying variable is continuous (level of mathematical knowledge or ability). Special types of correlations, specifically biserial and tetrachoric correlations, are used to measure relationships involving such artificial dichotomies. Use of these correlations is based on the assumption that underlying a dichotomous measure is a normally distributed continuous variable. For the case of one quantitative and one dichotomous vari-

able, a biserial correlation provides an estimate of the relationship between the quantitative variable and the continuous variable underlying the dichotomy. For the case of two dichotomous variables, the tetrachoric correlation estimates the relationship between the two continuous variables underlying the measured dichotomies. Biserial correlations could be used to estimate the relationship between a quantitative measure, such as a measure of neuroticism as a personality attribute, and the continuous variable that underlies a dichotomous test item, such as an item on a mathematical skills test. Tetrachoric correlations are commonly used to estimate relationships between continuous variables that underlie observed dichotomous variables, such as two test items.

Note that for the case of one quantitative and one dichotomous variable, one could calculate a point-biserial correlation, to measure the observed relationship, or a biserial correlation, to estimate the relationship involving the continuous variable underlying the dichotomous measure. The biserial correlation will be larger than the corresponding point-biserial correlation, because of the assumed gain in measurement precision inherent in the former. In the population, the relationship between these two correlations is given by

$$\rho_{pb} = \rho_b\left(\frac{h}{\sqrt{pq}}\right), \tag{1}$$

where p and q are the proportions of the population above and below the point of dichotomization, and h is the ordinate of the normal curve at that same point (Magnusson, 1966). Values of h for any point of dichotomization can be found in standard tables of normal curve areas and ordinates (e.g., Cohen & Cohen, 1983, p. 521).

For the case of two dichotomous variables, one could compute either a phi coefficient to measure the observed relationship or the tetrachoric correlation to estimate the relationship between the underlying dichotomies. Again because of the assumed gain in measurement precision, the tetrachoric correlation is higher than the corresponding phi coefficient. Although the general relationship between a phi coefficient and a tetrachoric correlation is quite complex, it can be defined for dichotomization at the mean as follows:

$$\rho_{phi} = 2[\arcsin(\rho_{tetrachoric})]/\pi \tag{2}$$

(Lord & Novick, 1968, p. 346). If the assumptions inherent in the biserial and tetrachoric correlations are valid, then the corresponding point-biserial correlation and phi coefficient can be seen to underestimate the relationships of interest because of their failure to account for the artificial nature of the dichotomous measures. Given this background on correlation coefficients, we now turn to an examination of how dichotomization of a quantitative variable impacts measures of association between variables.

Analyses of effects of one independent variable. Various aspects of the impact of dichotomization on results of statistical analysis have been examined and discussed in the methodological literature for many years. Here we review and examine in detail the most important issues in this area, beginning with the simplest case of dichotomization of a single independent variable. Basic issues associated with this case were discussed by Cohen (1983). Suppose that X and Y follow a bivariate normal distribution in the population with a correlation of ρ_{XY}; variance in Y accounted for by its linear relationship with X is then ρ^2_{XY}. If X is dichotomized at the mean to produce X_D, then the resulting population correlation between X_D and Y can be designated ρ_{X_DY}. (Note that for a normally distributed variable, dichotomization at the mean and the median are equivalent in the population.) To understand the impact of dichotomization on the relationship between the two variables, we must examine the relationship between ρ_{XY} and ρ_{X_DY}. This relationship can be seen to correspond to the theoretical relationship between a biserial and point-biserial correlation. That is, ρ_{X_DY} corresponds to a point-biserial correlation, representing the association between a dichotomous variable and a quantitative variable, and ρ_{XY} is equivalent to the corresponding biserial correlation, where X is the continuous, normally distributed variable that underlies X_D. The relationship between a point-biserial and biserial correlation was given in Equation 1. Given this relationship, the effect of dichotomization on ρ_{XY} can then be represented as

$$\rho_{X_DY} = \rho_{XY}\left(\frac{h}{\sqrt{pq}}\right). \tag{3}$$

The value of $h/\sqrt{pq}$ can be viewed as a constant, to be designated d, representing the effect of dichotomization under normality. For example, if X is dichotomized at the mean to produce X_D, then $p = .50$, $q = .50$, $h = .399$, yielding $d = .798$. The effect of dichotomization on the correlation is then given by ρ_{X_DY}

$= (.798)\rho_{XY}$, with shared variance being reduced by $\rho^2_{X_DY} = (.637)\rho^2_{XY}$. To represent the effect of dichotomization at points other than the mean, Figure 4 shows the value of $d = h/\sqrt{pq}$ as a function of p, the proportion of the population above the point of dichotomization. It can be seen that as the point of dichotomization moves further from the mean, the impact on the correlation coefficient increases. Clearly there is substantial loss of effect size in the population due to dichotomization at any point. This quantification of the loss is consistent with the subjective impression conveyed by Figures 1 and 2, where the linear relationship between X and Y appears to be weakened by dichotomization of X.

The results in our numerical example reported earlier are consistent with these theoretical results. Our sample was drawn from a population where $\rho_{XY} = .40$ and $\rho^2_{XY} = .16$. Following dichotomization, these population values would become $\rho_{X_DY} = (.798)(.40) = .32$ and $\rho^2_{X_DY} = (.637)(.16) = .10$. In our sample we found that dichotomization reduced $r_{XY} = .30$ to $r_{X_DY} = .21$, and the corresponding squared correlation from .09 to .04. Thus, the proportional reduction of effect size in our sample was slightly larger than would have occurred in the population.

Of course, there would be sampling variability in the degree of change from r_{XY} to r_{X_DY}. If we were to generate a new sample ($N = 50$) for our illustration, we would obtain different values of r_{XY} and r_{X_DY}. The impact of dichotomization would vary from sample to sample. An interesting question is whether dichotomization could cause r_{XY} to increase, even under normality, simply because of sampling error. This point is relevant because researchers sometimes justify dichotomization because of a finding that it yielded a higher correlation. To examine this issue, we conducted a small-scale simulation study. We defined five levels of population correlation, $\rho_{XY} = .10, .30, .50, .70, .90$. We then generated repeated random samples from bivariate normal populations at each level of ρ_{XY}, using six different levels of sample size, $N = 50, 100, 150, 200, 250, 300$. We generated 10,000 such samples for each combination of levels of sample size and ρ_{XY}. In each sample we computed r_{XY}, then dichotomized X at the median and computed r_{X_DY}. For each combination of ρ_{XY} and sample size, we then simply counted the number of times, out of 10,000, that $r_{X_DY} > r_{XY}$, that is, the number of times where dichotomization resulted in an increase in the correlation. Results are shown in Table 1. Of interest is the fact that when sample size and ρ_{XY} were relatively small, it was not unusual to find that dichotomization resulted in an increase in the correlation between the variables, simply due to sampling error. That is, even though, under bivariate normality, dichotomization must cause the population correlation

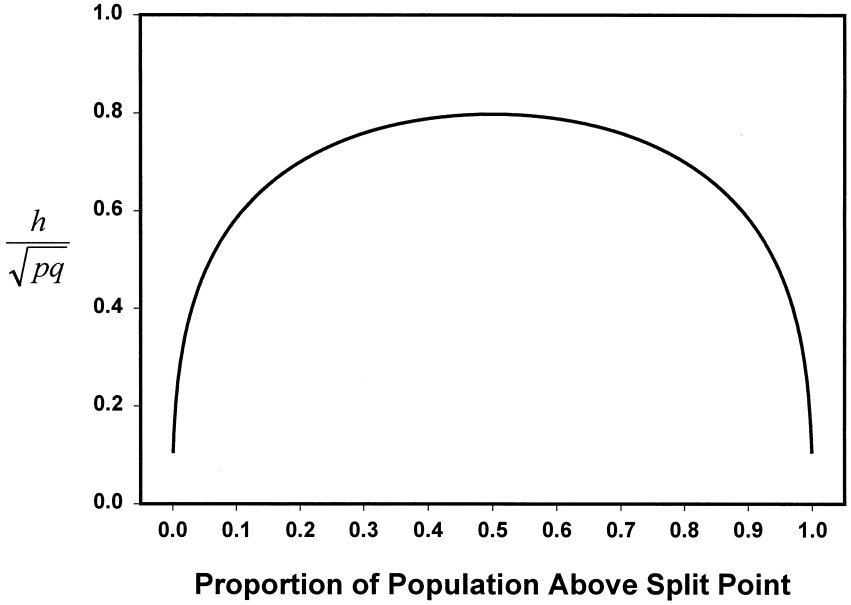

Figure 4. Proportional effect of dichotomization of X on correlation between X and Y as a function of point of dichotomization; p and q are the proportions of the population above and below the point of dichotomization, and h is the ordinate of the normal curve at the same point.

Table 1
Frequency of Increase in Correlation Coefficient After Dichotomization of Independent Variable for Various Levels of Population Correlation (ρ_{XY}) and Sample Size

	N					
ρ_{XY}	50	100	150	200	250	300
.10	4,126	3,760	3,435	3,277	3,015	2,882
.30	2,430	1,620	1,109	776	612	434
.50	1,015	371	127	56	25	2
.70	173	12	3	1	0	0
.90	0	0	0	0	0	0

Note. Entries indicate the number of trials out of 10,000 in which $r_{X_DY} > r_{XY}$.

to decrease, application of this approach in small samples or when ρ_{XY} is relatively small can easily result in an increase in the sample correlation. Undoubtedly in some cases such increases would cause a correlation that was not statistically significant prior to dichotomization to become so after dichotomization. These results must raise a caution about potential justification of dichotomization in practice. A finding that $r_{X_DY} > r_{XY}$ must not be taken as evidence that dichotomization was appropriate or beneficial. In fact, under conditions typically found in psychological research, dichotomization will cause the population correlation to decrease; an observation of an increase in the sample correlation in practice is very possibly attributable to sampling error. Failure to understand this phenomenon could easily cause inappropriate substantive interpretations.

The loss of effect size in the population following dichotomization, and corresponding expected loss in the sample, can affect the outcome of tests of statistical significance. In our earlier example, the *t* statistic for testing the significance of *r* dropped from 2.19 prior to dichotomization to 1.47 after, and statistical significance was lost. This loss of statistical significance can be attributed directly to loss of statistical power. Considering the power of the test of the null hypothesis of zero correlation, we note that in our example, prior to dichotomization power was .84, based on $\rho_{XY} = .40$, $N = 50$, $\alpha = .05$, two-tailed test. After dichotomization of *X,* power was reduced to .63, based on $\rho_{X_DY} = (.798)(.40) = .32$. Such a loss of power would become more severe as the point of dichotomization moves away from the mean, because the loss of effect size would be greater (see Figure 4). As noted by Cohen (1983), the loss of power caused by dichotomization can be viewed alternatively as an effective loss of sample size. For instance, in our example, prior to dichotomization, power of .63 could have been achieved with a sample size of only 32. Thus, the reduction in power from .84 to .63 due to dichotomization was equivalent to reducing sample size from 50 to 32, or discarding 36% of our sample. For a two-tailed test of the null hypothesis of zero correlation, using $\alpha = .05$, this effective loss of sample size resulting from a median split will be consistently close to 36%. It will deviate from this level when any of these aspects is altered, in particular becoming greater when the point of dichotomization deviates from the mean.

Let us next consider the case where both the independent variable *X* and the dependent variable *Y* are dichotomized, thereby converting a correlation question into analysis of a 2 × 2 table of frequencies. We can examine the impact of double dichotomization by focusing on the relationship between ρ_{XY} and $\rho_{X_DY_D}$, the correlations before and after double dichotomization. The relationship between these values corresponds to the relationship between a phi coefficient and the corresponding tetrachoric correlation, assuming bivariate normality. The value $\rho_{X_DY_D}$ is a phi coefficient, a correlation between two dichotomous variables, and the value ρ_{XY} is the corresponding tetrachoric correlation, the correlation between normally distributed variables, *X* and *Y,* that underlie the two dichotomies, X_D and Y_D. For dichotomization at the mean, the relationship between the phi coefficient and the tetrachoric correlation was given in Equation 2. In the present context, this relationship becomes

$$\rho_{X_DY_D} = 2[\arcsin(\rho_{XY})]/\pi \quad (4)$$

and thus represents the impact of double dichotomization on the correlation of interest.[2] This relationship is shown in Figure 5, indicating the association between population correlations obtained before dichotomization (analogous to tetrachoric correlation, on horizontal axis) and after dichotomization (analogous to phi coefficient, on vertical axis). For instance, the value of $\rho_{XY} = .40$ in our example would be reduced to $\rho_{X_DY_D} = .26$. Our sample result showed an even larger reduction, from $r_{XY} = .30$ to $r_{X_DY_D} = .06$.

[2] For the case of dichotomization of both variables, Cohen (1983) incorrectly assumed that the effect on ρ_{XY} would be the square of the effect of single dichotomization; for example, $\rho_{X_DY_D} = (.798)^2\ \rho_{XY}$ for dichotomization at the mean. This same error occurs in Peters and Van Voorhis (1940) and was recognized by Vargha et al. (1996).

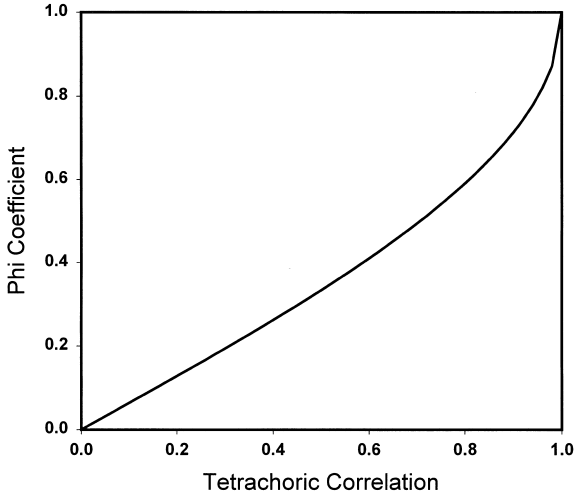

Figure 5. Relationship between phi and tetrachoric correlation for dichotomization of *X* and *Y* at their means.

In general, loss of effect size is greater when both variables are dichotomized than when only one is dichotomized. As the point of dichotomization moves away from the mean, the relationship between the phi coefficient and tetrachoric correlation becomes much more complex (Kendall & Stuart, 1961; Pearson, 1900), and the difference between the two coefficients becomes greater.

As in the case of dichotomization of only *X,* the loss of effect size under double dichotomization can be represented as a loss of statistical power. In our example, power of the test of zero correlation would be reduced from .84 prior to dichotomization to only .45 after dichotomization of both *X* and *Y.* And as before, such loss of power could be represented as an effective discarding of a portion of the original sample. Loss of power, or effective loss of sample size, becomes more severe if either or both of the variables are dichotomized at points away from the mean.

It is important to keep in mind that the effects of dichotomization of *X,* or of both *X* and *Y,* just described are based on the assumption of bivariate normality of *X* and *Y.* It may be tempting to conclude that in empirical populations these effects would not hold. However, Cohen (1983) emphasized that these phenomena would be altered only marginally by nonnormality. It would require extreme skewness, heteroscedasticity, or nonlinearity to substantially alter these consequences of dichotomization, conditions that probably are rather uncommon in social science data. When such conditions are present, that situation simply means that the formulas provided above for the influence of dichotomization on correlations might not hold closely. Such a complication does not provide justification for dichotomization of variables. Rather, it would still be advisable not to dichotomize, but instead to retain information about individual differences and to consider resolving the extreme skewness, heteroscedasticity, or nonlinearity by use of transformations of variables or nonlinear regression.

The issue of nonlinearity merits special attention. Suppose that the original relationship between *X* and *Y* were nonlinear, such that a scatter plot such as that in Figure 1 revealed a clear nonlinear association. Such a relationship could be represented easily using nonlinear regression, and the investigator could obtain and present a clear picture of the association between the variables. However, if *X,* or both *X* and *Y,* were dichotomized, that nonlinear relationship would be completely obscured. Presentation of results based on analyses conducted after such dichotomization would be misleading and invalid. Such errors can easily occur accidentally if researchers dichotomize variables without examining the nature of the association between the original variables.

Analyses of effects of two independent variables. We next examine statistical issues for the case of two independent variables, X_1 and X_2, and one dependent variable, *Y.* The linear relationship of X_1 and X_2 to *Y* can be studied easily using regression methods. Standard linear regression can be extended to investigate interactive effects of the independent variables by introducing product terms (e.g., $X_3 = X_1X_2$) into the regression model (Aiken & West, 1991). However, in practice it is not uncommon for investigators to dichotomize both X_1 and X_2 prior to analysis and to use ANOVA rather than regression. Over a period of more than 30 years, a number of methodological papers have examined the impact of such an approach on statistical results and conclusions. Humphreys and colleagues investigated several issues in this context in a series of articles. Humphreys and Dachler (1969a, 1969b) discussed an approach that they called a *pseudo-orthogonal design* in which individuals are selected in high and low groups on both X_1 and X_2 so as to produce a 2 × 2 design with equal sample sizes in each cell. This approach does not involve dichotomization of X_1 and X_2 after data have been collected but does involve treating individual-differences measures as if they were categorical with only two levels. Such a design had been used by Jensen (1968) in a study of the relationship of intelligence (X_1) and so-

cioeconomic status (X_2) to a measure of rote learning (Y). Humphreys and Dachler (1969a, 1969b) pointed out that this approach forces the independent variables into an orthogonal design when in fact the original X_1 and X_2 may well be correlated. They showed that an ensuing ANOVA would yield biased estimates of differences among means as a result of ignoring the correlation between X_1 and X_2. Humphreys and Fleishman (1974) provided further discussion of potentially misleading results from a pseudo-orthogonal design and also examined the approach wherein measures of X_1 and X_2 are dichotomized after data are gathered. Humphreys and Fleishman focused on the fact that this approach will generally yield unequal sample sizes in the resulting 2×2 design, and they reviewed various ways to analyze such data using ANOVA. They showed how ANOVA results would be related to regression results and described the expected loss of effect size attributable to dichotomization, as well as the potential occurrence of spurious interactions. In yet another article on this matter, Humphreys (1978a), commenting on an applied article by Kirby and Das (1977), again cautioned against dichotomization of independent variables to construct a 2×2 ANOVA design and reiterated the impact in terms of loss of effect size and power as well as distortion of effects. Throughout this series of articles Humphreys and colleagues repeated the general theme of negative consequences associated with dichotomization of continuous independent variables, either by selection of high and low groups or by dichotomization of collected data. They emphasized the loss of information about individual differences and the bias in estimates of effects. They argued that ANOVA methods are inappropriate ("unnecessary, crude, and misleading"; Humphreys, 1978a, p. 874) when independent variables are individual-differences measures and that it is preferable to use regression and correlation methods in such situations so as to retain information about individual differences and avoid negative consequences incurred by dichotomization (Humphreys, 1978b).

The case of dichotomization of two independent variables was examined further by Maxwell and Delaney (1993). After citing numerous empirical studies that followed such a procedure, Maxwell and Delaney showed that the impact of dichotomization on main effects and interactions depends on the pattern of correlations among independent and dependent variables. Although under many conditions dichotomization of two independent variables will result in loss of effect size for main effects and interaction, it was shown that under some conditions dichotomization can yield a spurious main effect. The reader will recall that our numerical example presented earlier exhibited such a phenomenon. Our regression analyses showed a near zero effect for one of the independent variables, but ANOVA using dichotomized independent variables yielded a significant main effect for that same variable. Maxwell and Delaney showed that when the partial correlation of one independent variable with the dependent variable is near zero and the independent variables are correlated with each other, a spurious significant main effect is likely to occur after dichotomization of both predictors. Our earlier numerical example had this property: The two predictors were substantially correlated (.50), and the partial correlation of X_2 with Y was zero. The regression analysis properly revealed no effect of X_2 on Y, whereas ANOVA after dichotomization of X_1 and X_2 yielded a spurious main effect of X_2. Maxwell and Delaney demonstrated that there would be highly inflated Type I error rates for tests of main effects in such situations and that these spurious effects were a result of bias in estimating population effects and were not attributable to sampling error. Finally, Maxwell and Delaney also showed that spurious significant interactions can occur when two independent variables are dichotomized. Such an event can occur when there are direct nonlinear effects of one or both of X_1 and X_2 on Y but no interaction in the regression model. After dichotomization of X_1 and X_2 a subsequent ANOVA will often yield a significant interaction simply as a misrepresentation of the nonlinearity in the effect of X_1 and/or X_2.

Vargha, Rudas, Delaney, and Maxwell (1996) extended the work of Maxwell and Delaney (1993) by further examining the case of two independent variables and one dependent variable. They carefully delineated the loss of effect size or the likely occurrence of spurious significant effects under various combinations of dichotomized and nondichotomized variables, showing that the impact of dichotomization depends on the pattern of correlations among the three variables.

In some instances where effects of two quantitative independent variables are to be investigated, data are analyzed by dichotomizing only one of the two variables, leaving the other intact. Such an approach has been used to study moderator effects. For instance, if it is hypothesized that the influence of X_1 on Y depends on the level of X_2, the researcher might dichoto-

mize X_2 and then conduct separate regression analyses of Y on X_1 for each level of X_2. Moderation is then assessed by testing the difference between the two resulting values of r_{X_1Y}, with a significant difference supposedly indicating a moderator effect. It is important to note that most methodologists would advise against such an approach for investigating moderator effects and would recommend instead the use of standard regression methods that incorporate interactions of quantitative variables (Aiken & West, 1991).

Bissonnette, Ickes, Bernstein, and Knowles (1990) conducted a simulation study to compare the dichotomization approach to the regression approach for examining moderator effects. When no moderator effect was present in the population, they found high Type I error rates under the dichotomization approach, indicating common occurrence of spurious interactions. Under the regression approach, Type I error rates were nominal. When moderator effects were present in the population, they found a higher rate of correct detection of such effects using the regression approach than the dichotomization approach. Given the distortion of information incurred by dichotomization along with the inflated Type I error rates or loss of power, as well as the straightforward capacity of standard regression to test for moderator effects, the use of the dichotomization method in this context seems unwarranted and risky.

Finally, it should be noted that although our presentation here has been limited to the cases of one or two independent variables, the issues and phenomena we have examined are not limited to those cases. The same issues apply for designs with three or more independent variables, although further complexities arise depending on the pattern of correlations among the variables and how many variables are dichotomized.

Aggregation and comparison of results across studies. Regardless of the number of independent variables, dichotomization raises issues regarding comparison and aggregation of results across studies. Allison, Gorman, and Primavera (1993) cautioned that dichotomization may introduce a lack of comparability of measures and results across studies. For example, groups defined as high or low after dichotomization may not be comparable between studies, especially if the point of dichotomization is data dependent (e.g., the median). That is, the point of dichotomization may vary considerably between studies, thus making groups not comparable. Even when dichotomization is conducted using a predefined scale point, resulting groups may differ considerably depending on the nature of the population from which the sample was drawn.

Hunter and Schmidt (1990) discussed problems caused by dichotomization when results from different studies are to be aggregated using meta-analysis. They showed that dichotomization of one or more independent variables will tend to cause downward distortion of aggregated measures of effect sizes as well as upward distortion of measures of variation of effect sizes across studies. They suggested methods for correcting these distortions in meta-analytic studies. However, those methods do not resolve the problem because they rely on their own assumptions and estimation methods. Such corrections are necessary only because of the persistent use of dichotomization in applied research.

Summary of Impact of Dichotomization on Measurement and Statistical Analyses

In this section we have reviewed literature on the variety of negative consequences associated with dichotomization. These include loss of information about individual differences, loss of effect size and power, the occurrence of spurious significant main effects or interactions, risks of overlooking nonlinear effects, and problems in comparing and aggregating findings across studies. To our knowledge, there have been no findings of positive consequences of dichotomization. Given this state of affairs, it would seem that use of dichotomization in applied research would be rather rare, but that is not at all the case. We now examine such usage in selected areas of applied research in psychology.

The Use of Dichotomization in Practice

Method

We selected six journals publishing research articles in clinical, social, personality, applied, and developmental psychology. The specific journals selected were *Journal of Personality and Social Psychology, Journal of Consulting and Clinical Psychology, Journal of Counseling Psychology, Developmental Psychology, Psychological Assessment,* and *Journal of Applied Psychology.* These journals are known for publishing research articles of high quality. The rationale behind selecting leading journals is simple. Leading journals are known for their high

standards, especially when statistical and methodological considerations are present. If we were to find common use of dichotomization in high-quality journals, this would suggest not only that uses of dichotomization must appear in other journals as well but also that leading researchers, as well as editors and reviewers, may be relatively unaware of the consequences associated with the use of dichotomization.

For each journal, we set out to examine all articles published from January 1998 through December 2000. This time interval was selected so as to reflect current practice. We limited our review to articles containing empirical studies, and we examined each such article to determine if any measured variables had been dichotomized prior to statistical analyses. An examination of articles published in 1998 in these six journals showed relatively frequent use of dichotomization in three of the journals—*Journal of Personality and Social Psychology, Journal of Consulting and Clinical Psychology,* and *Journal of Counseling Psychology*—but relatively rare usage in the other three—*Developmental Psychology, Psychological Assessment,* and *Journal of Applied Psychology.* (Although *Developmental Psychology* contained relatively few uses of dichotomization per se, it did contain an abundance of examples wherein subjects were divided into several groups based on chronological age.) Therefore, the full 3-year literature review was conducted only on the former set of three journals. In our review, we tabulated information about various forms of dichotomization, including median splits, mean splits, and other splits at selected scale points. We also noted other types of splits, such as tertiary splits or selection of extreme groups, although such splits are not examined directly in this article. For each instance of dichotomization we noted the variable that was split. In addition, we noted whether or not a justification for the split was given and what that justification was.

Results

In the three journals examined, there were a total of 958 articles. Of the 958 articles we found that a total of 110 articles, or 11.5%, contained at least one instance of dichotomization of a quantitatively measured variable. For those 110 articles a total of 159 instances of dichotomization were identified. For practical reasons, we did not count multiple median splits performed on the same variable within a given article. Summary information about our literature survey is presented in Table 2.

Table 2
Summary of Literature Search

Journal	No. of articles	No. of articles with splits[a]	Percentage of articles with splits	No. of articles with double splits
JPSP	518	82 (123)	15.8	7
JCCP	312	20 (27)	6.4	1
JCP	128	8 (9)	6.3	1
Total	958	110 (159)	11.5	9

Note. The literature search was conducted on all articles published in JPSP, JCCP, and JCP from January 1998 through December 2000.
JPSP = *Journal of Personality and Social Psychology*; JCCP = *Journal of Consulting and Clinical Psychology*; JCP = *Journal of Consulting Psychology.*
[a] Numbers in parentheses refer to total number of variables split.

As stated earlier, we also examined justifications offered for use of dichotomization. Of the 110 cases in which dichotomization was conducted, only 22 of those cases (20%) were accompanied by any justification. Thus, dichotomization seems to be most often used without any explicit justification. Some of the justifications offered for dichotomization included (a) following practices used in previous research, (b) simplification of analyses or presentation of results, (c) gaining a capability for examining moderator effects, (d) categorizing because of skewed data, (e) using clinically significant cutpoints, and (f) improving statistical power. In most cases, however, as noted above, no justification at all was offered, and it is not clear that this matter was even considered. Variables that were split were most often self-rated psychological scales. Scores on instruments assessing depression, anxiety, marital satisfaction, self-monitoring, attitudes, self-esteem, need for cognition, narcissism, and so forth were frequently dichotomized, although these examples represent only a small subset of the dichotomized variables encountered. Chronological age was the subject of frequent dichotomization, although it was more often segmented into several "age groups" based on arbitrary cutpoints.

Potential Justifications for Dichotomization

We now consider the question of why there seems to be persistent use of dichotomization in applied research in psychology in spite of the methodological case against it. We examine here a variety of possible reasons or justifications for such usage and offer our own assessment of each. Some of these justifications are extracted from explicit statements in published studies, whereas others are drawn from numerous dis-

cussions with colleagues and applied researchers over a period of many years. Many researchers who use dichotomization are eager to defend the practice in such discussions, and in the following sections we try to provide a fair representation of such defenses.

Lack of Awareness of Costs

Some researchers who use dichotomization may simply be unaware of its likely costs and negative consequences as delineated earlier in this article. When made aware of such costs, some of these individuals may be eager to use more appropriate regression methods and thereby avoid the costs and enhance the measurement and statistical aspects of their studies. We hope that the present article will have such an impact on researchers.

Perceiving Costs as Benefits

Some investigators acknowledge the costs but argue that those very costs provide a justification for dichotomization. This argument is that because dichotomization typically results in loss of measurement information as well as effect size and power, it must yield a more conservative test of the relationship between the variables of interest. Therefore, a finding of a statistically significant relationship following dichotomization is more impressive than the same finding without dichotomization; the relationship must be especially strong to still be found even when effect size and power have been reduced. Essentially, this argument reduces to the position that a more conservative statistical test is a benefit if it yields a significant result.

Let us consider this defense carefully. The argument focuses on results of a test of statistical significance and also rests on the premise that dichotomization will make such tests and corresponding measures of effect size more conservative. The focus on statistical significance is unfortunate, and the premise is false. First, regarding a focus on statistical significance, the American Psychological Association (APA) Task Force on Statistical Inference (Wilkinson and the Task Force on Statistical Inference, 1999) urged researchers to pay much less attention to accept–reject decisions and to focus more on measures of effect size, stating that "reporting and interpreting effect sizes . . . is essential to good research" (p. 599). In addition, the *Publication Manual of the American Psychological Association* (5th ed.; APA, 2001) emphasized that significance levels do not reflect the magnitude of an effect or the strength of a relationship and stated the following with respect to reporting results: "For the reader to fully understand the importance of . . . [a researcher's] findings, it is almost always necessary to include some index of effect size or strength of relationship in . . . [the article's] Results section" (p. 25). With regard to the present issue, this perspective implies that it is misguided to focus merely on whether or not a statistical test yields a significant result after dichotomization. Rather, it is important to consider the size of the effect of interest. Our review of methodological issues presented earlier emphasized that dichotomization can play havoc with measures of effect size, generally reducing their magnitude in bivariate relationships and potentially reducing or increasing them in analyses involving multiple independent variables. Second, the belief that statistical tests will always be more conservative after dichotomization is mistaken. Although this will tend to be true in the case of dichotomization of a single independent variable, it is by no means always true. Our earlier results (see Table 1) showed that dichotomization may cause an increase in effect size simply due to sampling error, thus producing a less conservative test. In addition, statistical tests may not be more conservative when two or more independent variables are dichotomized. In that case, as illustrated earlier, spurious significant main effects or interactions may arise depending on the pattern of intercorrelations between the variables. Thus, it is not at all the case that dichotomization will always make a statistical test or a measure of effect size more conservative.

Even if dichotomization did routinely yield more conservative statistical tests and measures of effect size, such a defense of its usage is highly suspect. Researchers typically design and conduct studies so as to enhance power and measures of effect size, as reflected in efforts to use reliable measures, obtain large samples, and use moderate alpha levels in hypothesis testing. If in fact more conservative tests were more desirable and defensible, then perhaps researchers should use less reliable measures, smaller samples, and small levels of alpha. One might then argue that if significant effects were still found, those effects must be quite strong and robust. Of course, such an approach to research would be counterproductive as are most uses of dichotomization. Our general view is that this defense of dichotomization is not supportable on statistical grounds and is inconsistent with basic principles of research design and data analysis.

Lack of Awareness of Proper Methods of Analysis

Over a period of many years, in discussions about dichotomization, we have encountered numerous researchers who simply have been unaware that regression/correlation methods are generally more appropriate for the analysis of relationships among individual-differences measures. This is especially true of individuals whose primary training and experience in the use of statistical methods is limited to or heavily emphasizes ANOVA. Of course, ANOVA is an important tool for data analysis in psychological research. However, difficulties and serious consequences arise when a problem that is best handled by regression/correlation methods is transformed into an ANOVA problem by dichotomization of independent variables. We are convinced that many researchers are not sufficiently familiar with or aware of regression/correlation methods to use them in practice and therefore often force data analysis problems into an ANOVA framework. This seems especially true in situations where there are multiple independent variables and the investigator is interested in interactions. We have repeatedly encountered the argument that dichotomization is useful so as to allow for testing of interactions, under the (mistaken) belief that interactions cannot be investigated using regression methods. In fact, it is straightforward to incorporate and test interactions in regression models (Aiken & West, 1991), and such an approach would avoid problems described earlier in this article regarding biased measures of effect size and spurious significant effects. Furthermore, regression models can easily incorporate higher way interactions as well as interactions of a form other than linear × linear—for example, a linear × quadratic interaction. Thus, we encourage researchers who use dichotomization out of lack of familiarity with regression methods to invest the modest amount of time and effort necessary to be able to conduct straightforward regression and correlational analyses when independent variables are individual-differences measures.

A related issue, or underlying cause of this lack of awareness of appropriate statistical methods, involves training in statistical methodology in graduate programs in psychology. Results of surveys of such training (Aiken, West, Millsap, & Taylor, 2000; Aiken, West, Sechrest, & Reno, 1990) indicate that education in statistical methods is somewhat limited for many graduate students in psychology and often focuses heavily on ANOVA methods. Although training in regression methods is certainly available in many programs, it receives less emphasis and is less often part of a standard program. Clearly, enhanced training in basic methods of regression analysis could increase awareness and usage of appropriate methods for analysis of measures of individual differences and help to avoid some of the problems resulting from overreliance on ANOVA, such as the common usage of dichotomization.

Dichotomization Resulting in Higher Correlation

It is not rare for an investigator to dichotomize an independent variable, X, and to find that the correlation between X and the dependent variable, Y, is higher after dichotomization than before dichotomization; that is, $r_{X_D Y} > r_{XY}$. Under such circumstances it may be tempting for the investigator to believe that such a finding justifies the use of dichotomization. It does not. In our earlier discussion of the impact of dichotomization on results of statistical analyses, we reviewed how population correlations would generally be reduced. We also showed by means of a simulation study that this effect will not always hold in samples (see the results in Table 1). That is, simply because of sampling error, sample correlations may increase following dichotomization, especially when sample size is small or the sample correlation is small. Such an occurrence in practice may well be a chance event and does not provide a sound justification for dichotomization.

Dichotomization as Simplification

A common defense of dichotomization involves, in one form or another, an argument that analyses and results are simplified. In terms of analyses, this argument rests on the premise that an analysis of group differences using ANOVA is somehow simpler than an analysis of individual differences using regression/correlation methods. We suggest that proponents of this position may simply be more familiar and comfortable with ANOVA methods and that the analyses and results themselves are not simpler. Both approaches can be applied in routine fashion, and results can be presented in standard ways using statistics or graphics. For instance, from our example of a bivariate relationship presented early in this article, correlational results showed $r_{XY} = .30$, $r^2_{XY} = .09$, $t(48) = 2.19$, $p = .03$. Figure 1 shows a scatter plot of the raw data, which conveys an impression of strength of association. Results after dichotomization would be pre-

sented in terms of a difference between group means. Means of Y for the high and low groups on X were 21.1 and 19.4, respectively. The test of the difference between these means yielded $t(48) = 1.47$, $p = .15$. Both sets of results are straightforward to present, and we see no real gain in simplicity by using ANOVA instead of regression.

For designs with two independent variables, the same principle holds. ANOVA results are typically presented in terms of tables or graphs of cell means. Regression results would be presented in terms of regression coefficients and associated information. An extremely useful mechanism for presenting interactions in a regression analysis is to display a plot of several regression lines. For example, a finding of a significant interactive effect of X_1 and X_2 on Y could be portrayed in terms of three regression lines. The first represents the regression of Y on X_1 for an individual low on X_2 (e.g., 1 standard deviation below the mean), the second represents the regression of Y on X_1 for an individual at the mean of X_2, and the third represents the regression of Y on X_1 for an individual high on X_2 (e.g., 1 standard deviation above the mean). All three lines can be displayed in a single plot, which can provide a simple and clear basis for understanding the nature of the interaction. This approach to presenting interactions is described and illustrated in detail by Aiken and West (1991). Again, we see no loss in simplicity of analysis or results for regression versus ANOVA.

We do recognize that for some individuals it may be conceptually simpler to view results in terms of group differences rather than individual differences. However, this conceptual simplification, to whatever extent it is real, is achieved only at a high cost—loss of information about individual differences, havoc with effect sizes and statistical significance, and so forth. Furthermore, the acceptance of such a simplified perspective will be misleading if the "groups" yielded by dichotomization are not real but are simply an artifact of arbitrarily splitting a sample and discarding information about individual differences. We consider the notion of groups shortly.

In considering the argument that dichotomization provides simplification, Allison et al. (1993) countered with the view that in fact dichotomization introduces complexity. They emphasized complications arising when two or more independent variables are dichotomized. Such a procedure typically creates a nonorthogonal ANOVA design, with its associated problems of analysis and interpretation, and also plays havoc with effect sizes and interpretations as described earlier. In fact, regression analyses are much simpler and allow the researcher to avoid these problems and the associated chance of being misled by results.

One additional perspective regarding the defense of simplification should be considered. Some researchers may dichotomize variables for the sake of the audience, believing that the audience will be more receptive to and will more easily understand analyses and results conveyed in terms of group differences and ANOVA. Although this perspective may be partially valid at times, we urge researchers and authors to avoid this trap. Clearly such a concession has negative consequences that far outweigh any perceived gains, very possibly including the drawing of incorrect conclusions about relationships among variables. No real interests are served if researchers use methods known to be inappropriate and problematic in the belief that the target audience will better understand analyses and results, especially when the results may be misleading and proper methods are simple and straightforward. All parties would be better served by obtaining and implementing the basic knowledge necessary for selection and use of appropriate statistical methods as well as for reading and understanding results of corresponding analyses.

Representing Underlying Categories of Individuals

Perhaps the most common defense of dichotomization is that there actually exist distinct groups of individuals on the variable in question, that a dichotomized measure more appropriately represents those groups, and that analyses should be conducted in terms of group differences rather than individual differences. Such an argument is potentially controversial both statistically and conceptually and must be examined closely.

There seem to be two distinct perspectives regarding the construction or existence of distinct groups of individuals on psychological attributes. These views have been discussed in the personality research literature. Block and Ozer (1982) discussed the *type-as-label* versus *type-as-distinctive form* perspectives. Drawing a similar distinction, Gangestad and Snyder (1985) described *phenetic* versus *genetic* bases for classes and also expressed the view that some psychological variables are dimensional, wherein individual differences are a matter of degree, whereas others are discrete, wherein individual differences

represent discrete classes. The type-as-label or phenetic view is that classes are merely constructed from data and have no underlying meaning. That is, they do not represent distinct categories of individuals but are essentially arbitrary groups defined by partitioning a sample according to scores on variables of interest. The partitioning may be carried out by an arbitrary splitting of one or more scales or may be the result of some more sophisticated approach such as cluster analysis. Regardless, the resulting groups have no substantive meaning as distinct categories. The type-as-distinctive-form or genetic view is based on the notion that there exists a discrete structure of individual differences underlying the observed variation on quantitative measures. The underlying categories have a genetic basis and a causal role with regard to observed measures. Meehl (1992) referred to such classes as *taxons*, and stated that quantitative measures may distinguish among taxons in the sense that those classes differ in kind as well as degree.

In the context of psychological research, it is clear that, to use Gangestad and Snyder's (1985) terms, phenetically defined groups are not as interesting or important as genetically defined groups. The former do not provide any insight about variables of interest, nor about relationships of those variables to others. In fact, basing analyses and interpretation on such arbitrary groups is probably an oversimplification and potentially misleading. Genetically defined groups are clearly more interesting, but also somewhat controversial when their existence is inferred from individual-differences measures. It would seem difficult to make a case that groups defined by dichotomization of measures of such attributes as self-esteem, anxiety, and need for cognition are genetically based and represent a discrete structure of individual differences. This is not to say that such structures do not exist. Meehl (1992; Waller & Meehl, 1998) suggested that there is ample empirical evidence to support the existence of taxons, especially in the areas of personality and clinical assessment.

An interesting case study regarding the potentially controversial nature of this issue began with an article by Gangestad and Snyder (1985), who presented a conceptual argument along with research results attempting to make the case that the variable self-monitoring has a discrete underlying structure. Assuming the observed distribution of scores on a self-monitoring measure to represent distinct classes, and using methods for identifying latent classes or taxons (see Waller & Meehl, 1998), Gangestad and Snyder made a case for the existence of discrete high and low groups that were split at about a 60:40 ratio. Miller and Thayer (1989) attempted to refute this purported class structure for self-monitoring based on various correlational analyses that, they argued, supported the systematic structure of individual differences beyond the differences represented by a simple high–low dichotomy. In a subsequent article, Gangestad and Snyder (1991) argued that the results obtained by Miller and Thayer were not germane to the view that there existed an underlying class structure for self-monitoring.

Regardless of which party had the upper hand in this exchange and regardless of whether or not there exist two distinct categories of individuals with respect to self-monitoring, a critical point is that the existence of such classes is subject to study and analysis and must not be assumed. As emphasized by Meehl (1992), the question of whether a particular phenomenon is most appropriately viewed as dimensional, taxonic, or some mix of these is an empirical question. Such issues can be examined using methods for identifying distinct classes or distributions of individuals, given measures for a sample of individuals on one or more variables. Waller and Meehl (1998) used the generic term *taxometric methods* for such techniques. Taxometric methods include a variety of techniques such as cluster analysis (methods for identifying clusters of individuals from multivariate data; Arabie, Hubert, & De Soete, 1996), mixture models (methods for identifying distinct distributions of individuals that combine to form an observed overall distribution; Arabie et al., 1996; Waller & Meehl, 1998), and latent class analysis (methods for identifying types of individuals based on multivariate categorical data; McCutcheon, 1987). In the present context where we are considering the possible existence of distinct types or groups of individuals on single measured variables, relevant taxometric methods would probably be limited to univariate mixture models and some clustering methods that could be applied to single variables. Regardless, whenever the researcher wishes to investigate or argue for the existence of distinct taxons, types, or classes, he or she should support the argument with taxometric analyses, rather than assume its validity.

Given this background, let us consider the defense of dichotomization as a technique for representing types of individuals. It should be immediately clear that dichotomization is not a taxometric method but is rather a simple method for grouping individuals into

arbitrarily defined classes. Even if discrete latent classes actually exist in a given case, the groups resulting from dichotomization may bear little or no resemblance to those latent classes. Dichotomization assumes that the number of taxons is two and does not allow for the possibility that there are more than two. Furthermore, dichotomization at the median assumes that the base rate for each of those two taxons is 50%. Dichotomization at some other scale point is probably no less arbitrary in terms of unjustifiable assumptions about the number of taxons and their base rates. Although Meehl (1992) argued for the existence of taxons in some cases, he clearly saw little value in dichotomization, cautioning that categories concocted by dichotomization of some measure are simply arbitrary classes and are not of interest to psychologists. In short, even if a researcher believes that there exist distinct groups or types of individuals and an underlying taxonic structure, dichotomization is not a useful technique. It is based on untenable assumptions and defines arbitrary classes that are unlikely to have much empirical validity. Rather, in such situations, a researcher should apply taxometric methods to assess the presence of latent classes and classify individuals into groups for subsequent analyses.

Dichotomization and Reliability

In questioning colleagues about their reasons for use of dichotomization, we have often encountered a defense regarding reliability. The argument is that the raw measure of X is viewed as not highly reliable in terms of providing precise information about individual differences but that it can at least be trusted to indicate whether an individual is high or low on the attribute of interest. Based on this view, dichotomization, typically at the median, would provide a "more reliable" measure. Cohen (1983) mentioned this possible justification but claimed that dichotomizing would in fact reduce reliability by introducing errors of discreteness. He did not elaborate on the effect of dichotomization on measurement reliability. We examine this issue in detail here and show that, under principles of classical measurement theory, dichotomization of a continuous variable does not refine the original measurement, but on the contrary, makes reliability substantially worse.

In classical measurement theory, an observed variable score, X, is assumed to be the sum of the true score, T, and random error, e:

$$X = T + e. \tag{5}$$

Under classical assumptions, reliability, ρ_{XX}, is then defined as the ratio of true variance to total variance (Allen & Yen, 1979; Magnusson, 1966):

$$\rho_{XX} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2} = \frac{\sigma_T^2}{\sigma_X^2}. \tag{6}$$

To examine points of interest in the present context, it is especially useful to consider the correlation between observed and true scores, ρ_{XT}, which has a simple relationship to reliability. The correlation between observed and true scores can be shown to be equal to the square root of reliability:[3]

$$\rho_{XT} = \sqrt{\rho_{XX}}. \tag{7}$$

The value of ρ_{XT} is called the reliability index and is designated here as $\tilde{\rho}_{XX}$, so that

$$\tilde{\rho}_{XX} = \rho_{XT} = \sqrt{\rho_{XX}}. \tag{8}$$

When the observed variable X is dichotomized to yield X_D, the value of the reliability index will change, and our interest is in the nature and degree of such change. After dichotomization we can define the reliability index as ρ_{X_DT}. This definition is problematic, however, because there are several ways in which T can be defined after X has been dichotomized. We consider here three different definitions of T for a dichotomized X, and for each definition of T we define a corresponding reliability index. These three versions of the reliability index will be designated $\tilde{\rho}_{XX(1)}$, $\tilde{\rho}_{XX(2)}$, and $\tilde{\rho}_{XX(3)}$.

The first such index is based on the notion that the true score is unchanged by dichotomization of X. Then the reliability index for X_D could be defined as

$$\tilde{\rho}_{XX(1)} = \rho_{X_DT}. \tag{9}$$

Note that if X is normal, the relationship between $\tilde{\rho}_{XX(1)}$ and $\tilde{\rho}_{XX}$ (or between ρ_{X_DT} and ρ_{XT}) is effectively the relationship between a point-biserial correlation and a biserial correlation, respectively. That is, T is a continuous variable, and X is a continuous, normally distributed variable underlying X_D. The relationship between the point-biserial and biserial correlations was mentioned earlier in another context and was

[3] The relationship between reliability and the correlation of observed and true scores can be shown as follows:

$$\rho_{XT} = \frac{\sigma_{XT}}{\sigma_X\sigma_T} = \frac{E(XT)}{\sigma_X\sigma_T} = \frac{E(T+e)T}{\sigma_X\sigma_T}$$
$$= \frac{E(T^2) + E(Te)}{\sigma_X\sigma_T} = \frac{E(T^2) + 0}{\sigma_X\sigma_T} = \frac{\sigma_T^2}{\sigma_X\sigma_T} = \frac{\sigma_T}{\sigma_X} = \sqrt{\rho_{XX}},$$

where E is the expected value operator.

specified in Equation 1. Adapting Equation 1 to the present context yields

$$\tilde{\rho}_{XX(1)} = \tilde{\rho}_{XX}\left(\frac{h}{\sqrt{pq}}\right). \quad (10)$$

For any specific dichotomizing point, $\tilde{\rho}_{XX(1)}$ is a constant proportion of the corresponding $\tilde{\rho}_{XX}$, where the constant is given by $d = h/\sqrt{pq}$. For dichotomization at any specified point, the effect on the reliability index is then given by $\tilde{\rho}_{XX(1)} = d\tilde{\rho}_{XX}$. For dichotomization at the mean, $d = .798$ (based on $p = .50$, $q = .50$, $h = .399$). For dichotomization at 0.5 standard deviations from the mean, $d = .762$ (based on $p = .691$, $q = .309$, $h = .352$). For more extreme dichotomization at 1.0 standard deviation, $d = .662$ (based on $p = .841$, $q = .159$, $h = .242$). These values indicate the reduction in reliability due to dichotomization, considering T as unaltered. Figure 4, presented earlier in another context, can now be viewed as portraying the reduction in the reliability index caused by dichotomization at any specified point on a normally distributed X.

We now consider a second way to conceptualize the reliability of X_D. Suppose we conceive of the dichotomized measure X_D as a measure of a dichotomized true variable, T_D. That is, a perfectly reliable X_D would indicate without error whether each individual is above or below the dichotomization split point on the true score distribution. From this perspective, the reliability index for X_D could be defined as

$$\tilde{\rho}_{XX(2)} = \rho_{X_D T_D}. \quad (11)$$

Note that $\tilde{\rho}_{XX(2)}$ is a correlation between two dichotomous variables and is thus a phi coefficient. From this perspective, if X and T are both normal, then the original reliability index, or $\tilde{\rho}_{XX}$, is effectively the corresponding tetrachoric correlation. The effect on reliability of dichotomization of both X and T is then given by the relationship between a phi coefficient and a corresponding tetrachoric correlation. For the case when both variables are dichotomized at the mean, this relationship was specified in Equation 2 in another context. Adapting Equation 2 to the present context yields

$$\tilde{\rho}_{XX(2)} = 2[\arcsin(\tilde{\rho}_{XX})]/\pi. \quad (12)$$

Figure 5, presented earlier, can also be applied to the present context as showing the impact on reliability when both X and T are dichotomized at the mean. For dichotomization away from the mean, the relationship between a phi coefficient and the corresponding tetrachoric correlation becomes very complex with the difference increasing as the split point deviates further from the mean, implying a greater impact on reliability in the present context.

A third definition of the reliability index after dichotomization is based on defining the true score, according to classical measurement theory, as the expected value of the observed measurement. If an individual takes the same measurement repeatedly under appropriate conditions, or if he or she takes parallel tests, his or her true score would be equal to the mean of an infinite number of such repeated measures or parallel tests. This mean over an infinite number of testings is represented formally using the expected value operator, E, which indicates more generally the mean of a random variable over an infinite number of samplings. In the present context, we express the true score for individual i as the expected value of the observed score:

$$T_i = E(X_i). \quad (13)$$

If the observed variable X is dichotomized, the corresponding true score T^* becomes

$$T_i^* = E(X_{Di}). \quad (14)$$

For dichotomization at the split point c, if the dichotomized observed variable X is coded as 1 if it is greater than c or as zero if not, the true score can be further expressed as

$$T_i^* = p(X_{Di} = 1) = p(X_i > c \mid T_i). \quad (15)$$

This true score for an observed variable after dichotomization is the conditional probability of the observed score being greater than c given the original undichotomized true score. The conditional probability density function of X_i given T_i is normal with mean T_i and variance σ_e^2. Therefore, the true score of individual i after dichotomization is represented by the area under the normal curve, with mean T_i and variance σ_e^2, above the cutpoint c. Formally, this is written using the normal probability density function as

$$T_{Di}^* = \int_c^{\infty}\left(\frac{1}{\sqrt{2\pi\sigma_e^2}}\right)\exp\left[\frac{-(X - T_i)}{2\sigma_e^2}\right]dX. \quad (16)$$

Note that T_D^* is a continuous variable. The correlation between the dichotomized observed score X_D and the newly defined true score T_D^* can be computed, and a third version of reliability index is defined as this correlation:

$$\tilde{\rho}_{XX(3)} = \rho_{X_D T_D^*}. \quad (17)$$

It would be desirable to derive the functional relationship between this newly defined index of reliability,

$\tilde{\rho}_{XX(3)}$, and the original reliability index, $\tilde{\rho}_{XX}$. However, such an expression may not be tractable. A simulation study to be presented shortly demonstrates the relationship for some conditions.

In summary, we have presented three alternative definitions of the index of reliability when the measure X is dichotomized. The first index, $\tilde{\rho}_{XX(1)}$, is the correlation between the dichotomized observed variable and the original true score; the second, $\tilde{\rho}_{XX(2)}$, is the correlation between the dichotomized observed variable and the dichotomized true score; and the third, $\tilde{\rho}_{XX(3)}$, is the correlation between the dichotomized observed variable and the corresponding classically defined true score, which is the expected value of the dichotomized observed score. The relationships of $\tilde{\rho}_{XX(1)}$ and $\tilde{\rho}_{XX(2)}$ to the original reliability index $\tilde{\rho}_{XX}$ are specified in Equations 10 and 12, respectively. The relationship of $\tilde{\rho}_{XX(3)}$ to $\tilde{\rho}_{XX}$ probably does not have a tractable algebraic representation. These relationships define the impact of dichotomization on reliability under classical measurement theory.

To demonstrate and further examine this effect, we conducted a simple simulation study. Artificial data were generated under the classical measurement model using four levels of the reliability coefficient: .60, .70, .80, and .90. Transformed to the reliability index, these values correspond to .775, .837, .894, and .949, respectively. For each of these four conditions, a single large random sample ($N = 10{,}000$) was generated based on the classical measurement model (Equation 5). First, true scores and random errors were generated independently. The true score was assumed unit normally distributed, and the random error score was normally distributed with a mean of zero and a variance of $\sigma_e^2 = 1/\rho_{XX}$. True scores and random errors were then added to create observed scores for the 10,000 simulated individuals within each level of reliability.

For each of these four large samples, the measure X was then dichotomized at the median, and sample values of $\tilde{\rho}_{XX(1)}$, $\tilde{\rho}_{XX(2)}$, and $\tilde{\rho}_{XX(3)}$ were computed. Note that values of $\tilde{\rho}_{XX(1)}$ and $\tilde{\rho}_{XX(2)}$ could be predicted by Equations 10 and 12, respectively. Table 3 shows squared values of these indices obtained after dichotomization, along with predicted values for $\tilde{\rho}^2_{XX(1)}$ and $\tilde{\rho}^2_{XX(2)}$. Squared values of the reliability index correspond to classically defined reliability. Note first that the values obtained from the simulated data were very close to the predicted values. More generally, results show clearly that dichotomization caused substantial decreases in reliability regardless of how the true score is defined. The impact of dichotomization varied across levels of reliability and among the different indexes examined but was always substantial.

Table 3
Effect of Dichotomization at Mean on Reliability

Reliability before dichotomization	Reliability after dichotomization	
	Observed	Predicted
	Reliability index (squared): $\tilde{\rho}^2_{XX(1)}$	
.600	.375	.382
.700	.442	.446
.800	.514	.509
.900	.572	.573
	Reliability index (squared): $\tilde{\rho}^2_{XX(2)}$	
.600	.307	.318
.700	.386	.398
.800	.500	.497
.900	.642	.632
	Reliability index (squared): $\tilde{\rho}^2_{XX(3)}$	
.600	.399	
.700	.487	
.800	.596	
.900	.716	

Note. Definitions of alternative versions of the reliability index, $\tilde{\rho}_{XX(1)}$, $\tilde{\rho}_{XX(2)}$, and $\tilde{\rho}_{XX(3)}$, are provided in Equations 10, 12, and 17, respectively.

This simulation study was extended to examine the effect of dichotomization at points away from the mean. As expected, as the point of dichotomization deviated further from the mean, the impact on reliability increased. Detailed results are not presented here but may be obtained from Robert C. MacCallum.

In summary, the foregoing detailed analysis shows that dichotomization will result in moderate to substantial decreases in measurement reliability under assumptions of classical test theory, regardless of how one defines a true score. As noted by Humphreys (1978a), this loss of reliable information due to categorization will tend to attenuate correlations involving dichotomized variables, contributing to the negative statistical consequences described earlier in this article. To argue that dichotomization increases reliability, one would have to define conditions that were very different from those represented in classical measurement theory.

General Comments on Defenses of Dichotomization

We have considered a variety of defenses and justifications that have been offered for the practice of dichotomization of individual-differences measures.

We believe that under close examination each of these defenses breaks down. Without valid justification, researchers who dichotomize individual-differences measures are in a position of using a method that does considerable harm to the analysis and understanding of their data, with no clear beneficial effects.

Legitimate Uses of Dichotomization?

Given that we have called into question a wide range of justifications that are often offered for dichotomization, it is interesting to consider the question as to whether dichotomization is ever justified. Although we believe that there may be occasional cases where it is justified, we emphasize that we believe such cases to be very rare. Two such cases are considered here. One would be a situation in which taxometric analyses provided clear support for the existence of two types or taxons within the observed sample, along with a clear scale point that differentiated the classes. Of importance, the existence of types must be supported by such analyses and must not be assumed or theorized. Even when such support is obtained, the classes in question almost certainly could not be identified by the commonly used median split approach because of the arbitrariness of the point of dichotomization. Furthermore, the researcher must recognize that, even given clear evidence for the existence of distinct groups, dichotomization would likely result in the loss of reliable information about individual differences within the groups, as well as misclassification of some individuals. Claims of the existence of types, and corresponding dichotomization of quantitative scales and analysis of group differences, simply must be supported by compelling results from taxometric analyses.

A second possible setting in which dichotomization might be justified involves the occasional situation where the distribution of a count variable is extremely highly skewed, to the extent that there is a large number of observations at the most extreme score on the distribution. For example, suppose research participants were asked how many cigarettes they smoked per day. A large number of people would give a response of "zero," and the remainder of the sample would show a distribution of nonzero values. Such a distribution indicates the presence of two groups of people, smokers and nonsmokers. Corresponding dichotomization of the measured variable would yield a dichotomous indicator of smoking status, which may be useful for subsequent analyses. However, an investigator following such a procedure must recognize that such dichotomization involves loss of all information about variation among those individuals not at the extreme scale point (e.g., variation in smoking frequency among smokers). If such information is to be retained and the variable is left intact, it is essential to use specialized regression methods for analysis of such variables; readers are referred to a discussion of regression models for count variables in Long (1997).

Although there may be an occasional situation in which dichotomization is justified, we submit that such circumstances are very rare in practice and do not represent common usage of dichotomization. In common usage, dichotomization is typically carried out without apparent justification and without serious awareness or regard for its consequences. Such ad hoc procedures are simply inappropriate and incur substantial costs.

Summary and Conclusions

The methodological literature shows clearly and conclusively that dichotomization of quantitative measures has substantial negative consequences in most circumstances in which it is used. These consequences include loss of information about individual differences; loss of effect size and power in the case of bivariate relationships; loss of effect size and power, or spurious statistical significance and overestimation of effect size in the case of analyses with two independent variables; the potential to overlook nonlinear relationships; and, as shown in this article, loss of measurement reliability. These consequences can be easily avoided by application of standard methods of regression and correlational analysis to original (undichotomized) measures.

Despite these circumstances, many researchers continue to apply dichotomization to measures of individual differences. This practice may be due to factors such as a lack of awareness of the consequences or of appropriate methods of analysis, a belief in the presence of types of individuals, or a belief that dichotomization improves reliability. Such justifications have been examined in the present article and, we believe, have been shown to be invalid.

Cases in which dichotomization is truly appropriate and beneficial are probably rare in psychological research and probably could be recognized only through taxometric analyses. Thus, we urge researchers who are in the habit of dichotomizing individual-differences measures to carefully examine whether such a practice is justified. We also urge journal edi-

tors and reviewers to caution authors against the use of dichotomization in most situations, and we caution consumers of research to be skeptical regarding results of studies wherein individual-differences measures have been dichotomized. We hope that such efforts will drastically reduce the practice of dichotomization and its negative impact on our scientific endeavors.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Aiken, L. S., West, S. G., Millsap, R., & Taylor, A. (2000, October). *Don't shoot the messengers: First results from the survey of the quantitative curriculum of doctoral programs in psychology.* Paper presented at the meeting of the Society of Multivariate Experimental Psychology, Saratoga Springs, NY.

Aiken, L. S., West, S. G., Sechrest, L., & Reno, R. R. (1990). Graduate training in statistics, methodology, and measurement in psychology. *American Psychologist, 45,* 721–734.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory.* Monterey, CA: Sage.

Allison, D. B., Gorman, B. S., & Primavera, L. H. (1993). Some of the most common questions asked of statistical consultants: Our favorite responses and recommended readings. *Genetic, Social, and General Psychology Monographs, 119,* 155–185.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Arabie, P., Hubert, L. J., & De Soete, G. (1996). *Clustering and classification.* Singapore: World Scientific.

Bissonnette, V., Ickes, W., Bernstein, I., & Knowles, E. (1990). Personality moderating variables: A warning about statistical artifact and a comparison of analytic techniques. *Journal of Personality, 58,* 567–587.

Block, J., & Ozer, D. J. (1982). Two types of psychologists: Remarks on the Mendelsohn, Weiss, and Feimer contribution. *Journal of Personality and Social Psychology, 42,*1171–1181.

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

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Gangestad, S. W., & Snyder, M. (1985). "To carve nature at its joints": On the existence of discrete classes in personality. *Psychological Review, 92,* 317–349.

Gangestad, S. W., & Snyder, M. (1991). Taxonomic analysis redux: Some statistical considerations for testing a latent class model. *Journal of Personality and Social Psychology, 61,* 141–146.

Humphreys, L. G. (1978a). Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. *Journal of Educational Psychology, 70,* 873–876.

Humphreys, L. G. (1978b). Research on individual differences requires correlational analysis, not ANOVA. *Intelligence, 2,* 1–5.

Humphreys, L. G., & Dachler, H. P. (1969a). Jensen's theory of intelligence. *Journal of Educational Psychology, 60,* 419–426.

Humphreys, L. G., & Dachler, H. P. (1969b). Jensen's theory of intelligence: A rebuttal. *Journal of Educational Psychology, 60,* 432–433.

Humphreys, L. G., & Fleishman, A. (1974). Pseudo-orthogonal and other analysis of variance designs involving individual-differences variables. *Journal of Educational Psychology, 66,* 464–472.

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

Jensen, A. R. (1968). Patterns of mental ability and socioeconomic status. *Proceedings of the National Academy of Sciences, 60,* 1330–1337.

Kaiser, H. F., & Dickman, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. *Psychometrika, 27,* 179–182.

Kendall, M. G., & Stuart, A. (1961). *The advanced theory of statistics: Vol. 2. Inference and relationship.* New York: Hafner.

Kirby, J. R., & Das, J. P. (1977). Reading achievement, IQ, and simultaneous–successive processing. *Journal of Educational Psychology, 69,* 564–570.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

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

Magnusson, D. (1966). *Test theory.* Reading, MA: Addison-Wesley.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median-splits and spurious statistical significance. *Psychological Bulletin, 113,* 181–190.

McCutcheon, A. C. (1987). *Latent class analysis.* Beverly Hills, CA: Sage.

Meehl, P. E. (1992). Factors and taxa, traits and types, dif-

ferences of degree and differences in kind. *Journal of Personality, 60,* 117–174.

Miller, M. L., & Thayer, J. F. (1989). On the existence of discrete classes in personality: Is self-monitoring the correct joint to carve? *Journal of Personality and Social Psychology, 57,* 143–155.

Pearson, K. (1900). On the correlation of characters not quantitatively measurable. *Royal Society Philosophical Transactions, Series A, 195,* 1–47.

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

Vargha, A., Rudas, T., Delaney, H. D., & Maxwell, S. E. (1996). Dichotomization, partial correlation, and conditional independence. *Journal of Educational and Behavioral Statistics, 21,* 264–282.

Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua.* Thousand Oaks, CA: Sage.

Wilkinson, L., and the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594–604.

Received July 10, 2001
Revision received October 25, 2001
Accepted October 25, 2001 ■