Some Implications of Distinguishing Between Unexplained Variance That Is Systematic or Random

Educational and Psychological
Measurement

2018, Vol. 78(3) 482–503

© The Author(s) 2017

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164417691573

journals.sagepub.com/home/epm

David Trafimow¹

Abstract

Because error variance alternatively can be considered to be the sum of systematic variance associated with unknown variables and randomness, a tripartite assumption is proposed that total variance in the dependent variable can be partitioned into three variance components. These are variance in the dependent variable that is explained by the independent variable, variance in the dependent variable that is unexplained but systematic (associated with variance in unknown variables), and random variance. Based on the tripartite assumption, classical measurement theory, and simple mathematics, it is shown that these components can be estimated using observable data. Mathematical and computer simulations illustrate some of the important issues and implications.

Keywords

variance due to the independent variable, systematic variance not due to the independent variable, random variance, tripartite assumption

The crucial role that measurement issues have for substantive work in psychology has long been obvious (e.g., Hunt, 1936; Kaplan & Saccuzzo, 2010; Michell, 1999; Thurstone, 1959). It is not an exaggeration to state that substance and measurement are inextricably linked. In substantive work, consistent with recommendations by many authorities (Finch, Cumming, & Thomason, 2001; Thompson, 1999; Vacha-Haase, Nilsson, Reetz, Lance, & Thompson, 2000; Wilkinson & The APA Task

¹New Mexico State University, Las Cruces, NM, USA

Corresponding Author:

David Trafimow, Department of Psychology, New Mexico State University, MSC 3452, P. O. Box 30001, Las Cruces, NM 88003-8001, USA.

Email: dtrafimo@nmsu.edu

Force on Statistical Inference, 1999), researchers report effect sizes to provide readers with a quantitative assessment of the size of the effects. Although this clearly is a positive trend, it is worthwhile to note that effect sizes are based on total variance (or the standard deviation which is the square root of total variance), but total variance, in turn, can be split into three components to be described presently. Arguably, then, total variance confounds these three components. It is interesting to consider the implications of unconfounding the three components and adjusting effect sizes accordingly. The goal of the present article is to explore these possibilities, aided by the minimal assumptions of the classical measurement theory.¹

Partitioning Variance

Researchers are used to partitioning dependent variable variance into variance that is accounted for by the independent variable (or variables) and variance that is considered to be error. When researchers consider the issue at all, they often consider error variance to be an amalgamation of random measurement error variance and nonrandom variance due to factors not manipulated or measured in the experiment. The classical measurement theory is different because it distinguishes between random measurement error variance and systematic variance of any kind (Borsboom, 2005; Lord & Novick, 1968; Trafimow & Rice, 2008). To see that this is so, consider the definition of a “true score,” which is the expectation across a set comprising an infinite number of independent test-taking occasions, as Equation (1) shows.²

$$\text{A Person's True Score} = \varepsilon(Y). \quad (1)$$

These infinite and independent test-taking occasions are hypothetical; there is no expectation that a researcher will attempt to actually accomplish it empirically. Importantly, it follows from this definition that an observed score is the true score plus an error component. However, this error component is truly random and so the correlation between true scores and random error scores, across participants, equals zero. That this correlation equals zero plays a crucial role in many of accomplishments of the classical theory. For example, an important classical theorem is that total variance in a measure of a construct equals true score variance plus random error variance. This theorem, along with another theorem that there is no correlation between true scores and random error scores, is useful for deriving the famous disattenuation formula, according to which observed correlations can be corrected for attenuation due to random measurement error (e.g., Spearman, 1904).

According to the classical theory, variance on a test can be considered to be due to both systematic (true score variance) and random factors. But researchers rarely ask participants to complete tests in isolation. Normally, there is at least one other test with which to correlate the test of interest. Or, there is an experimental manipulation designed to influence the test of interest. This fact implies that systematic variance due to the “independent” variable needs to be distinguished from systematic variance due to “other” variables, as well as from random measurement variance.

Stated another way, it is both possible and useful to perform a tripartite partition. The tripartite assumption (TA) that pushes in this direction assumes that there can be two sources of systematic variance in the dependent variable: there is systematic variance associated with the independent variable but there also is systematic variance that is associated with variables of which the experimenter has no knowledge. Finally, of course, there also is variance in the dependent variable that is random. Thus, the reasoning to be developed based on the TA implies that there are three types of variance rather than the usual two: there is systematic variance in the dependent variable associated with the manipulation (σ_{IV}^2), associated with other (unknown) variables (σ_O^2), and randomness (σ_R^2). Total variance in the dependent variable (σ_Y^2) is the sum of these three types of variance (see Equation 2).

$$\sigma_Y^2 = \sigma_{IV}^2 + \sigma_O^2 + \sigma_R^2. \quad (2)$$

Before continuing, it is important to be clear about an important assumption, which is that the three types of variance are independent of each other. The independence of σ_R^2 from σ_{IV}^2 and σ_O^2 comes directly from the classical theory. Most statistical tests (e.g., analysis of variance [ANOVA]) also assume that the sum of σ_O^2 and σ_R^2 is independent of σ_{IV}^2 . It follows that σ_O^2 is independent of both σ_{IV}^2 and σ_R^2 .

Is it possible to estimate σ_{IV}^2 , σ_O^2 , and σ_R^2 in terms of an observable variable? If not, the TA would not be very useful. In fact, however, it is possible and not even difficult. The derivations of σ_{IV}^2 , σ_O^2 , and σ_R^2 in terms of observables will be performed separately assuming a correlational study or an experimental one. Importantly, to actually carry out calculations based on the equations to be presented, it is necessary to have a good estimate of the reliability of the dependent variable. In turn, this presumes that the requisite reliability data have been collected or at least that the researcher is planning to collect reliability data in the future.

Correlational Study

Any statistics textbook provides the coefficient of determination (e.g., Harris, 1994; Rosenthal & Rosnow, 1991), which gives the proportion of variance that the independent variable and the dependent variable share. Letting X and Y denote the independent and dependent variables, respectively, the coefficient of determination can be denoted as ρ_{XY}^2 . To obtain the amount of variance in the dependent variable associated with variance in the independent variable (σ_{IV}^2), it is necessary merely to multiply ρ_{XY}^2 by the total variance (σ_Y^2) as Equation (3) shows.³

$$\sigma_{IV}^2 = \rho_{XY}^2 \sigma_Y^2. \quad (3)$$

Note that the sample statistics, r_{XY}^2 and s^2 , can be obtained from the data to estimate ρ_{XY}^2 and σ_Y^2 , respectively; therefore Equation (3) produces an estimate of σ_{IV}^2 . However, this leaves open the question of what is to be done about σ_O^2 and σ_R^2 .

To move in this direction, it is useful to recall the classical measurement theory, according to which, the reliability of the dependent variable ($\rho_{YY'}$) indexes the amount of randomness in responding (e.g., Gulliksen, 1987; Lord & Novick, 1968; Spearman, 1904; Trafimow & Rice, 2008, 2009). If the reliability is perfect, there is no randomness whereas if reliability is zero, all is random. A classic definition of reliability is true score variance (systematic variance) divided by total observed variance, as Equation (4) shows (e.g., Gulliksen, 1987).

$$\rho_{YY'} = \frac{\sigma_T^2}{\sigma_Y^2}. \quad (4)$$

According to the classical theory, the reliability ($\rho_{YY'}$) can be considered to be the correlation between two parallel measures of the dependent variable, which is equivalent to true score variance divided by dependent variable variance. Algebraic manipulation renders Equation (5).

$$\sigma_T^2 = \rho_{YY'} \sigma_Y^2. \quad (5)$$

Importantly, another theorem from the classical theory indicates that total variance is the sum of true score variance and random variance as Equation (6) shows (Gulliksen, 1987).

$$\sigma_Y^2 = \sigma_T^2 + \sigma_R^2. \quad (6)$$

And algebraic manipulation involving Equations (5) and (6) renders Equation (7).

$$\sigma_Y^2 - \sigma_R^2 = \rho_{YY'} \sigma_Y^2. \quad (7)$$

Additional algebraic manipulation of Equation (7) renders Equation (8).

$$\sigma_R^2 = \sigma_Y^2 - \rho_{YY'} \sigma_Y^2 = (1 - \rho_{YY'}) \sigma_Y^2. \quad (8)$$

Before continuing, it is worthwhile to pause and consider the very simple case where there is no independent variable, as there already has been presented a sufficient amount of the classical theory to parse variance into that which is due to random measurement error versus that which is due to systematic factors. For a quick worked out example, EXCEL was used to generate 40 random cases for two test-taking occasions under user-defined constraints where the population reliability is .70 and the population standard deviation is 1.00. The sample statistics were as follows: sample reliability ($r_{YY'}$) = .71, Test 1 standard deviation (s_{Y1}) = .96, and Test 2 standard deviation (s_{Y2}) = .88. Taking the average of the two standard deviations renders an estimate of the standard deviation (s_Y) as .92. Instantiating these values into Equation (8) gives an estimate for the variance due to measurement error: $s_R^2 = .92^2 - (.71 \cdot .92^2) = .25$. Subtracting s_R^2 from the total variance estimate gives an estimate for the systematic variance: $s_O^2 = .92^2 - .25 = .60$. Not surprisingly,

these estimates deviate somewhat from the population values, which are .30 and .70 for σ_R^2 and σ_O^2 , respectively.

Equation (8) expresses random variance in terms of observed variance in the dependent variable and the reliability of the dependent variable, both of which can be estimated from observable data. More generally, Equation (3) provides σ_{IV}^2 in terms of observable data and Equation (8) provides σ_R^2 in terms of observable data so all that is left is to obtain σ_O^2 in terms of observable data. To accomplish this, let us substitute Equation (3) and Equation (8) into Equation (2) to obtain Equation (9), and then simplify Equation (9) to express σ_O^2 in terms of observable data (Equation 10).

$$\sigma_Y^2 = \rho_{XY}^2 \sigma_Y^2 + \sigma_O^2 + \sigma_Y^2 - \rho_{YY'} \sigma_Y^2. \quad (9)$$

$$\sigma_O^2 = \rho_{YY'} \sigma_Y^2 - \rho_{XY}^2 \sigma_Y^2 = (\rho_{YY'} - \rho_{XY}^2) \sigma_Y^2. \quad (10)$$

Thus, between Equations (3), (8), and (10), one can estimate σ_{IV}^2 , σ_R^2 , and σ_O^2 , respectively, in terms of observables.⁴

Experiment

Although the foregoing equations would work for an experiment as well as a correlation, researchers are not used to thinking of experiments in correlational terms (but see Rosenthal, Rosnow, & Rubin, 2000 for an exception; also see Rosenthal & Rubin, 1982). Therefore, a separate derivation is provided below, assuming the typical experiment where the independent variable is categorical and comprises an experimental group and a control group. Harris (1994; also see Rosenthal & Rosnow, 1991) conveniently provided an equation to find the proportion of variance in the dependent variable that is associated with the independent variable ($\rho_{YX}^2 = \frac{T^2}{T^2 + df}$) and as was shown in the previous section (Equation 3), multiplication by the total variance in the dependent variable renders the variance in the dependent variable that is associated with variance in the independent variable.^{5,6}

$$\sigma_{IV}^2 = \frac{T^2}{T^2 + df} \sigma_Y^2. \quad (11)$$

After noting the existence of σ_R^2 from Equation (8), it only remains to find σ_O^2 . This can be done easily by substituting Equation (11) and Equation (8) into Equation (2) to obtain Equation (12). And algebraic manipulation renders Equation (13).

$$\sigma_Y^2 = \frac{T^2}{T^2 + df} \sigma_Y^2 + \sigma_O^2 + \sigma_Y^2 - \rho_{YY'} \sigma_Y^2. \quad (12)$$

$$\sigma_O^2 = \rho_{YY'} \sigma_Y^2 - \frac{T^2}{T^2 + df} \sigma_Y^2. \quad (13)$$

As in the previous section, σ_{IV}^2 , σ_O^2 , and σ_R^2 can be estimated in terms of observables.

The Value of Distinguishing σ_{IV}^2 , σ_O^2 , and σ_R^2

Researchers are used to the importance of σ_{IV}^2 . Therefore, the present focus will be on σ_O^2 and σ_R^2 . The distinction provides useful information with respect to the planning of future research and understanding effect sizes. Further uses will be explored later.

Planning Future Research

Suppose that a researcher obtains data and finds support for the hypothesis that the variance in the independent variable accounts for some of the variance in the dependent variable. What should the next step be? Should the researcher search for additional variables in the hope of accounting for more of the variance in the dependent variable not accounted for by the original independent variable? The foregoing analyses suggest that the researcher would be wise to estimate σ_O^2 and σ_R^2 before deciding. If σ_O^2 is a reasonably large number and σ_R^2 is not, the implication is that there exist additional variables that are capable of accounting for variance in the dependent variable not accounted for by the original independent variable. But suppose that σ_O^2 is a small number whereas σ_R^2 is a large number. In that case, the implication is that the researcher will not find additional variables that will help him or her to account for much additional variance in the dependent variable because the variance that is unaccounted for is mostly random. In this case, the researcher's future efforts might better be devoted to attempting to reduce randomness or perhaps even switching to a different project with more potential.

For researchers in a well-studied area, it might be clear what constitutes a large or small value for σ_O^2 and σ_R^2 . Often, however, this is not clear and so the researcher can use various sorts of ratios to help inform the decision about where to place future research efforts. For example, the researcher might consider σ_O^2/σ_R^2 , which gives the systematic variance in the dependent variable accounted for by other variables divided by random variance in the dependent variable. As the ratio increases, there would be more point in searching for additional variables whereas as this ratio decreases (especially when it is much less than unity), the researcher is unlikely to gain much from such a search. The obvious caveat to this point is that there might be additional effects that are very important, even if very small, that if found would justify the small ratio.

Another interesting ratio is σ_O^2/σ_{IV}^2 , which gives the variance in the dependent variable that can be accounted for by additional systematic factors divided by the variance in the dependent variable that is accounted for by the independent variable. If this ratio is large, it suggests that the independent variable has not exhausted the pool of potentially interesting variables that can account for systematic variance in the dependent variable. In contrast, if this ratio is much smaller than unity, it

suggests that the researcher already has accounted for most of the variance in the dependent variable that can be accounted for in a systematic way. The researcher will not be able to do as well at accounting for further variance in the dependent variable as he or she already has done with the independent variable.

A third ratio that is of potential interest is $\frac{\sigma_W^2 + \sigma_O^2}{\sigma_W^2 + \sigma_O^2 + \sigma_R^2}$, which gives the proportion of total variance in the dependent variable that is due to systematic factors (known and unknown). If this proportion is a small number, it suggests that there is much randomness in the dependent variable, and that effort should be devoted toward considering why this is so. Perhaps the researcher should consider attempting to develop a dependent variable or study design that is less susceptible to random influences. In contrast, if the proportion is a large number, and especially if σ_O^2/σ_R^2 and σ_O^2/σ_W^2 also are large, the search for additional independent variables likely would be fruitful.

Understanding the Effect Size

In a recent editorial, Trafimow and Marks (2015) banned the use of the null hypothesis significance testing procedure but favored the use of effect sizes. Although the disadvantages of null hypothesis significance testing require no further commentary here, the virtues of effect sizes, as currently computed, seem less clear.

Probably, the most commonly used effect size measure is Cohen's d , which is the difference in means divided by the standard deviation (or pooled standard deviation). From the present perspective, it is easiest to consider the standard deviation as the square root of the total variance in the dependent variable to facilitate thinking in terms of variances. Considered in this way, Cohen's d tells us the distance between the means in standard deviation units or square root of variance units $\left(\sigma_Y \text{ or } \sqrt{\sigma_Y^2}\right)$ when there are two groups (e.g., experimental and control conditions). Although this seems to be useful information to have, there are alternative ways in which to think about effect sizes that imply different formulas. And to see why it might be desirable to consider alternatives, it is useful to recall that the (square root of the) variance in the denominator of Cohen's d denotes the total variance in the dependent variable, which confounds σ_O^2 and σ_R^2 .

An obvious way to eliminate the confound is to replace the square root of the total variance with the square root of either σ_O^2 , or σ_R^2 . What are the implications of each?

If the square root of σ_O^2 is used, this decontaminates the denominator from both the independent variable and from randomness and is very much in the spirit of a suggestion by Trafimow (2014) to decontaminate standard deviations from the influence of randomness. In essence, then, such an effect size would render the distance between means in units of the square root of the variance of the dependent variable that is both unaccounted for and systematic $\left(\text{effectsize} = \frac{\text{Mean}_1 - \text{Mean}_2}{\sigma_O}\right)$.

On the other hand, it is possible to argue that Trafimow (2014) got the goal wrong, and that what researchers really should desire is to decontaminate randomness from both kinds of systematic influences. In this case, the denominator should be the square root of σ_R^2 . This way of computing the effect size would render the distance between means in units of the square root of random variance in the dependent variable ($\text{effectsize} = \frac{\text{Mean}_1 - \text{Mean}_2}{\sigma_R}$).

An additional, and completely different way to think about effect sizes is in terms of causation. Arguably, the main reason for conducting an experiment is that the researcher desires to investigate the causal effect of the independent variable on the dependent variable. From this point of view, it is necessary to characterize the meaning of causation and Russo (2009) has recommended that researchers think in terms of the amount of variance in one variable that is attributable to variance in the other variable. From this point of view, σ_{IV}^2 can be considered to be a direct index of the degree of causation, though a potential disadvantage of using σ_{IV}^2 is that this index is influenced by the unit of measurement (e.g., measuring money in pennies would result in a larger value for σ_{IV}^2 than would measuring it in dollars). Alternatively, the reader might consider σ_O^2 or σ_R^2 too. For example, it is possible to consider σ_{IV}^2/σ_O^2 as the ratio of the variance in the dependent variable that is caused by the independent variable to variance in the dependent variable that is caused by other systematic factors. Alternatively, it is possible to consider σ_{IV}^2/σ_R^2 as the ratio of variance in the dependent variable caused by the independent variable to random variance.

An Example of Alternative Conclusions Based on Different Variances

Thus far, the presentation of the three types of component variances; σ_{IV}^2 , σ_O^2 , and σ_R^2 ; has been abstract. The present section provides an example where the same amount of error variance (σ_E^2) implies different conclusions depending on σ_O^2 , and σ_R^2 .

Suppose that $\sigma_{IV}^2 = 5$, $\sigma_O^2 = 20$, and $\sigma_R^2 = 1$ or that $\sigma_{IV}^2 = 5$, $\sigma_O^2 = 1$, and $\sigma_R^2 = 20$. Note that in both cases, via Equation 2, the traditional error variance is 21 because $\sigma_E^2 = 20 + 1 = 21$ or $\sigma_E^2 = 1 + 20 = 21$. From the traditional point of view, there is no difference between these cases and the proportion of explained variance is $\frac{\sigma_{IV}^2}{\sigma_{IV}^2 + \sigma_E^2} = \frac{5}{5+21} = .19$. However, invoking the TA suggests that important differences nevertheless exist. Let us consider first σ_O^2/σ_R^2 which is either $20/1 = 20$ or it is $1/20 = .05$. The large ratio in the former case suggests that there are other variables whose variances are systematically associated with dependent variable variance and so a search for these variables is likely to be a good investment of the researcher's time and effort. In contrast, the small ratio in the latter case suggests that most of the variance not accounted for by the original independent variable is random and so a search for additional variables that matter in a systematic way is unlikely to be a good investment of the researcher's time and effort. The researcher might profitably

consider ways to decrease randomness in the experiment (e.g., by creating a more reliable dependent measure).

It is interesting that the proportion of systematic variance is $\frac{\sigma_W^2 + \sigma_o^2}{\sigma_W^2 + \sigma_o^2 + \sigma_R^2}$ which translates to .96 in the first case but only .23 in the second case. In general, all else being equal, the search for additional variables that are systematically related to the dependent variable makes more sense when there is a larger proportion of systematic variance than when there is a smaller proportion of systematic variance.

It also is interesting to consider the effect size according to the two cases. Two common ways to assess effect size are Cohen's d and the proportion of variance accounted for, both of which will give the same numbers for the two cases and which would push in the direction of the conclusion that there is no difference. In addition, if σ_{IV}^2 is used to index the effect sizes, there again would be no difference between the two cases. But let us consider the use of σ_{IV}^2/σ_O^2 , which implies that the effect size for the first case is $5/20 = .25$ whereas the effect size for the second case is $5/1 = 5$. The interpretation is that the independent variable accounts for a much more of the known systematic variance, relative to the systematic variance that is unknown, in the second case than in the first case. In contrast, if σ_{IV}^2/σ_R^2 is used, then the two ratios are 5 and .25, respectively, and so the conclusion would be that the independent variable performs much better relative to randomness in the first case than in the second case.

Mathematical and Computer Simulations

The foregoing analyses bring up the following issues. First, what is the effect of reliability on error variance and variance due to the independent variable? Second, given that sample-derived estimates of the variance in the dependent variable due to the independent variable, based on sample correlations or t -values are biased, how important is this and what is the effect of increasing the sample size and using a standard correction formula? Relatedly, it may matter if the parent distribution is normal or non-normal. Third, what is the effect of sample size on the range of sample reliability coefficients one is likely to observe? The following simulations address each of these questions, in turn.

The Effect of Reliability on Variance due to the Independent Variable

Figure 1 illustrates the results of a mathematical simulation where reliability was allowed to range from .4 to 1.0 along the horizontal axis for a reason to be explained in the subsequent paragraph. One line shows the effect on random error variance, which, according to Equation (8), is a function solely of total variance (set at 1) and reliability. As reliability increases, error variance decreases. The other lines give the variance in the dependent variable due to variance in the independent variable, setting the systematic variance due to other factors at arbitrary levels of .1, .2, .3, or .4. Figure 1 shows that, keeping total variance constant, variance due to the independent

Figure 1. Variance is represented along the vertical axis as a function of reliability along the horizontal axis and as a function of random error variance and variance in the independent variable not due to variance in the dependent variable (other variance).

variable increases as reliability increases and also as systematic variance due to other factors decreases.

It is now convenient to explain why the range in reliabilities was from .4 to 1.0. When the reliability is .4, the random error variance is .6, according to Equation (8). And when the variance due to other variables is set at .4, the total of these two variances is 1.0. Since total variance was set at 1, variance in the dependent variable due to variance in the independent variable has to be 0. That is, $1 - .6 - .4 = 0$. And if the reliability were allowed to range below the .4 level, Equation (2) implies that in the case where variance due to other variables is set at .4, variance in the dependent variable due to variance in the independent variable would have to take on negative values. Because this is not possible, it makes sense to set a lower limit of .4 for reliability coefficients along the horizontal axis in Figure 1.

The Effects of Correction and Distribution Type on Sample Variances

The foregoing equations have limitations. One limitation is the assumption of normal distributions. To address this limitation, uniform and right triangular distributions also were tested. Another limitation is that the equation that provides the variance in the dependent variable due to variance in the independent variable is biased at the sample level. Hays (1994) suggested using the following to estimate variance in the dependent variable due to variance in the independent variable at the sample level:

$$\frac{t^2 - 1}{t^2 + n_1 + n_2 - 1}$$

where n_1 and n_2 are the sample sizes of the two conditions, respectively. When using this correction, if the sample t -value is less than or equal to 1, the estimated variance in the dependent variable that is due to the independent variable is set at zero.

Figure 2 illustrates the results of computer simulations where there are two conditions, and the user-defined values for effect size and variance within each condition were arbitrarily set at .5 and 1, respectively. Sample sizes within each condition were set at 10, 20, 30, 40, or 50 (so total sample size was 20, 40, 60, 80, or 100) in the hope of covering a set of sample sizes that could reasonably be expected to be representative of the ones seen in published research. In addition, the computer simulations were run using either the uncorrected or corrected formulas for estimating the variance in the dependent variable due to variance in the independent variable. The uncorrected or corrected findings are represented by two panels in Figure 2.

Most important, the computer simulations were run with different populations in mind. That is, there were 10,000 cases per simulation assuming that the parent distributions were normal, uniform (rectangular), or to introduce skewness, right triangular. As there were 5 sample sizes within each parent distribution, and also there was correction or not, the total number of cases represented in Figure 2 is $10,000 \times 3 \times 2 = 60,000$. Figure 2 illustrates the estimated variance in the dependent variable due to variance in the independent variable along the vertical axis, as a function of the sample size per condition along the horizontal axis, and with curves representing the assumption of a normal, uniform, or right triangular distribution.

It is instructive to inspect the uncorrected panel first. When the parent distributions really are normal, the uncorrected formula performs well if the sample size per condition equals or exceeds 30, though there is a slight degree of overestimation even when these conditions are met. However, if the sample size is reduced, or if the parent distributions are not normal, there is more overestimation. The corrected panel shows the worth of correcting for bias at the sample level. In the best-case scenario, when the parent distributions are normal and the sample size per condition equals or exceeds 30, there is no overestimation, though at the cost of slight underestimation. Although there is overestimation when the sample size is less than 30 or when the parent distributions are not normal, the overestimation is less than in the uncorrected panel. In general, if the correction is used, the overestimation due to insufficient sample sizes and wrong assumptions about parent distributions are mitigated. However, if the researcher has a large sample size and is confident that the parent distribution is normal, it may be worthwhile to consider not correcting for bias to avoid underestimation.

Sample Size and Sample Reliability Ranges

One problem with reliability coefficients is that, like any other parameter, they have to be estimated from sample data. Figure 3 illustrates the difficulty by presenting the range of reliability coefficients that might reasonably be expected to be produced by

Figure 2. Variance in the dependent variable due to variance in the independent variable is represented along the vertical axis as a function of the sample size per condition of the experiment along the horizontal axis and type of parent distribution (normal, uniform, or right triangular).

Figure 3. The range of sample reliability coefficients that researchers can reasonably be expected to obtain is represented along the vertical axis as a function of the population reliability coefficient represented along the horizontal axis and sample size (10, 20, 30, 40, or 50).

various sample sizes, if the researcher were to conduct a reliability study. Because reliability coefficients are correlations, and distributions of correlation coefficients are skewed, Figure 3 was produced via mathematical simulation using Fisher's r to z transformations as follows (see Rosenthal & Rosnow, 1991).

Population reliability coefficients were allowed to vary and ranged along the horizontal axis from 0 to 1. Very low reliability coefficients are not representative of published research and so these arguably need not have been included. Nevertheless, it was not inconvenient to include them in Figure 3, so they were included. In any event, reliability coefficients were converted into a z -scores, using the following:

$$\frac{1}{2} \log_e \frac{1+r}{1-r}.$$

And standard deviations of z -scores were obtained; assuming a sample size for a reliability study of 10, 20, 30, 40, or 50; using the following: $\frac{1}{\sqrt{n-3}}$. Minimum and maximum z -scores were obtained by using $z \pm \frac{1}{\sqrt{n-3}}$ and then these were back-converted into reliability (correlation) coefficients using $\frac{e^{2z}-1}{e^{2z}+1}$. Subsequently, what might be considered to be a reasonable range of sample reliability coefficients was obtained, for each sample size, by subtracting the minimum reliability coefficient from the maximum reliability coefficient. That is, the reasonable range of reliability coefficients goes from one standard deviation below the mean to one standard deviation above the mean, based on the process of conversion to z -scores and back-conversion to reliability coefficients. These ranges are displayed along the vertical axis of Figure 3, as a function of the actual reliability coefficient and the sample size.

There is one complication. Specifically, when the population reliability is low, some of the sample reliability coefficients can be expected to be negative. Because of this possibility, it is necessary to make a decision about whether to set negative sample reliability coefficients at zero or whether to allow them to remain negative. Arguably, reliability coefficients below zero are possible, and so they should be included. Consistent with this position, the first panel of Figure 3 does not limit sample reliability coefficients to positive numbers. In contrast, it seems likely that many researchers would refuse to entertain reliability coefficients below zero, which implies that zero sets an effective limit. Consistent with this position, the second panel sets a lower limit at zero. Thus, the reader can observe the implications of both positions.

Let us first consider the panel where there is no limit, which illustrates the following. First, as the sample size increases, the range of sample reliability coefficients decreases. Second, as the population reliability increases along the horizontal axis, the range of sample reliability coefficients decreases. Third, there is an "interaction" whereby the importance of sample size on the range of reliability coefficients decreases as the population reliability increases. A practical implication of this interaction for the substantive researcher is that effort devoted to obtaining reliable measures may be quite economical because fewer participants are needed to substantially

reduce the range of sample reliability coefficients. To my knowledge, Figure 3 is the first demonstration of this interaction between sample size and size of the population reliability coefficient on the range of sample reliability coefficients. The range of sample reliability coefficients is an important issue because estimates of random variance, as well as variance due to other systematic factors, can be underestimates or overestimates depending on the degree to which sample reliability coefficients underestimate or overestimate true values. In turn, such underestimation or overestimation may be more pronounced when the range of sample reliability coefficients is large, whereas small ranges substantially mitigate the potential problem.

Changes in Component Variances

It is possible to imagine an ideal case where a dependent measure that is not perfectly reliable is made to be so, thereby removing all randomness (new $\sigma_R^2 = 0$). In this case, assuming that total variance remains constant, some of the former random variance will go into σ_{IV}^2 and some will go into σ_O^2 . If the steps taken to remove randomness have no influence on the relative contributions of the two types of systematic effects, then σ_{IV}^2 and σ_O^2 should increase according to proportionality in absolute terms. That is, σ_{IV}^2 should increase by $\frac{\sigma_{IV}^2}{\sigma_{IV}^2 + \sigma_O^2} \sigma_R^2$ and σ_O^2 should increase by $\frac{\sigma_O^2}{\sigma_{IV}^2 + \sigma_O^2} \sigma_R^2$.⁷ It is possible to consider the sum of the original σ_{IV}^2 and the increase due to the removal of randomness to be the “potential performance” (*PPIV*) of the independent variable, as expressed by Equation (14).

$$PPIV = \sigma_{IV}^2 + \frac{\sigma_{IV}^2}{\sigma_{IV}^2 + \sigma_O^2} \sigma_R^2. \quad (14)$$

Alternatively, researchers might be interested in the potential performance of other variables that are not yet known but nevertheless are systematically related to the dependent variable. Equation 15 renders the potential performance of these “other” variables.

$$PPO = \sigma_O^2 + \frac{\sigma_O^2}{\sigma_{IV}^2 + \sigma_O^2} \sigma_R^2. \quad (15)$$

Equations 14 and 15 are interesting because they imply an important qualification of the ratios presented earlier. That is, although eliminating randomness increases σ_{IV}^2 and σ_O^2 , in the ideal case from whence these equations were derived, an interesting ratio, such as σ_O^2/σ_{IV}^2 , does not necessarily change. Normally, if the ratio is a small number, this militates against searching for additional variables as suggested earlier. However, Equation (15) suggests that even given that the ratio remains a small number after randomness is removed, the absolute value of *PPO* nevertheless might increase sufficiently to make the search for new variables worthwhile.

Suppose that rather than decreasing σ_R^2 to zero, it decreases it to some other level. What will be the effect on σ_{IV}^2 and on σ_O^2 ? Similar reasoning that led to Equation (14) applies, but this time to the difference between the original σ_R^2 and the new σ_R^2 . To see that this is so, consider that in Equation (14), where the new σ_R^2 is zero, the difference is the original σ_R^2 minus zero, which equals the original σ_R^2 . If the new σ_R^2 is allowed to take on any value greater than zero, it follows that Equation 14 should be adjusted via Equation 16 below. In computing the new performance of the *IV* (*NPIV*), the original σ_R^2 is designated as σ_{R-1}^2 and the new σ_R^2 as σ_{R-2}^2 .

$$NPIV = \sigma_{IV}^2 + \frac{\sigma_{IV}^2}{\sigma_{IV}^2 + \sigma_O^2} (\sigma_{R-1}^2 - \sigma_{R-2}^2). \quad (16)$$

By similar reasoning, Equation (15) can be adjusted to render Equation (17), thereby providing the new performance of other variables that systematically influence the dependent variable (*NPO*) when random variance is changed.

$$NPO = \sigma_O^2 + \frac{\sigma_O^2}{\sigma_{IV}^2 + \sigma_O^2} (\sigma_{R-1}^2 - \sigma_{R-2}^2). \quad (17)$$

Thus far (Equations 14-17), the assumption has been that increasing the reliability of the measure, and thereby decreasing the variance attributable to randomness, is done without changing the relative contributions of the independent variable and variables with other systematic effects. But it is not necessary to assume this. Speaking more generally, there are two ways of increasing σ_{IV}^2 while keeping σ_Y^2 constant. One way is to reduce σ_R^2 and a second way is to reduce σ_O^2 . (If σ_Y^2 is not kept constant, increasing this value can increase σ_{IV}^2 too.) Suppose that the researcher finds a way to reduce σ_O^2 without having any influence on σ_R^2 or σ_Y^2 . In that case, whatever reduction occurs with respect to σ_O^2 transfers directly into σ_{IV}^2 as described by Equation 18 below. In Equation 18, it is convenient to denote the original variance accounted for by other variables as σ_{O-1}^2 and the new variance accounted for by other variables as σ_{O-2}^2 . It is similarly convenient to denote the original variance accounted for by the independent variable as σ_{IV-1}^2 and the new variance accounted for by the independent variable as σ_{IV-2}^2 .

$$\sigma_{IV-2}^2 = \sigma_{IV-1}^2 + (\sigma_{O-1}^2 - \sigma_{O-2}^2). \quad (18)$$

Equations (14)-(17) assume that random variance decreases but that this decrease increases the other two kinds of variances proportionally. Equation (18) assumes that any change in systematic variance in the dependent variable accounted for by other variables goes into the variance in the dependent variable accounted for by the independent variable. In contrast, suppose that there are two levels of a second independent variable whereby the first level corresponds to original values ($\sigma_{IV-1}^2, \sigma_{O-1}^2, \sigma_{R-1}^2$); the second level corresponds to new values ($\sigma_{IV-2}^2, \sigma_{O-2}^2, \sigma_{R-2}^2$); and this second independent variable is allowed to influence variance

in the dependent variable explained by variance in the independent variable, random variance and systematic variance accounted for by other variables, simultaneously. In that case, it would be possible to recycle equations presented earlier to estimate σ_{IV}^2 , σ_O^2 , and σ_R^2 but to use them separately for Level 1 of the second independent variable and Level 2 of the second independent variable. Importantly, the estimation cannot proceed unless there is a measure of the reliability of the dependent variable under each of the two levels of the second independent variable. Let us denote these as $\rho_{YY'-1}$ and $\rho_{YY'-2}$, respectively. Let us also denote the total variance in the dependent variable under each of these two levels as σ_{Y-1}^2 and σ_{Y-2}^2 , respectively. Also, assuming an interest in correlations under each of the two levels of the second independent variable, these can be denoted as ρ_{XY-1}^2 and ρ_{XY-2}^2 , respectively. If the first independent variable is manipulated, it is possible to denote the T -value under each of the two levels of the second independent variable as T_1 and T_2 , respectively; and the degrees of freedom associated with each T -value as df_1 and df_2 , respectively. Given these designations, it is possible to obtain the following.

If the study comprises correlations under Level 1 and Level 2 of a second independent variable, the equations for obtaining all of the component variances are given in Equations (19)-(24).

$$\sigma_{IV-1}^2 = \rho_{XY-1}^2 \sigma_{Y-1}^2. \quad (19)$$

$$\sigma_{R-1}^2 = \sigma_{Y-1}^2 - \rho_{YY'-1} \sigma_{Y-1}^2. \quad (20)$$

$$\sigma_{O-1}^2 = \rho_{YY'-1} \sigma_{Y-1}^2 - \rho_{XY-1}^2 \sigma_{Y-1}^2. \quad (21)$$

$$\sigma_{IV-2}^2 = \rho_{XY-2}^2 \sigma_{Y-2}^2. \quad (22)$$

$$\sigma_{R-2}^2 = \sigma_{Y-2}^2 - \rho_{YY'-2} \sigma_{Y-2}^2. \quad (23)$$

$$\sigma_{O-2}^2 = \rho_{YY'-2} \sigma_{Y-2}^2 - \rho_{XY-2}^2 \sigma_{Y-2}^2. \quad (24)$$

In contrast, if the first independent variable is manipulated under Level 1 and Level 2 of a second independent variable, the equations for obtaining all of the component variances are given in Equations (25)-(30).

$$\sigma_{IV-1}^2 = \frac{T_1^2}{T_1^2 + df_1} \sigma_{Y-1}^2. \quad (25)$$

$$\sigma_{R-1}^2 = \sigma_{Y-1}^2 - \rho_{YY'-1} \sigma_{Y-1}^2. \quad (26)$$

$$\sigma_{O-1}^2 = \rho_{YY'-1} \sigma_{Y-1}^2 - \frac{T_1^2}{T_1^2 + df_1} \sigma_{Y-1}^2. \quad (27)$$

$$\sigma_{IV-2}^2 = \frac{T_2^2}{T_2^2 + df_2} \sigma_{Y-2}^2. \quad (28)$$

$$\sigma_{R-2}^2 = \sigma_{Y-2}^2 - \rho_{YY'-2}\sigma_{Y-2}^2. \quad (29)$$

$$\sigma_{O-2}^2 = \rho_{YY'-2}\sigma_{Y-2}^2 - \frac{T_2^2}{T_2^2 + df_2}\sigma_{Y-2}^2. \quad (30)$$

Once the component variances have been computed under each level of the second independent variable, it is easy to obtain the effect of that second independent variable on each of the component variances by subtraction. Equations (31)-(33) give the effect of the second independent variable on each of the component variances in terms of differences. Note that if the manipulation increases a component variance when going from Level 1 to Level 2, the difference score will be negative whereas if a component variance decreases, the difference score will be positive.

$$\sigma_{IV-Diff}^2 = \sigma_{IV-1}^2 - \sigma_{IV-2}^2. \quad (31)$$

$$\sigma_{R-Diff}^2 = \sigma_{R-1}^2 - \sigma_{R-2}^2. \quad (32)$$

$$\sigma_{O-Diff}^2 = \sigma_{O-1}^2 - \sigma_{O-2}^2. \quad (33)$$

An Example

Suppose that a measure of the independent variable and a measure of the dependent variable are taken under two levels of a second independent variable and that the sample variances under levels 1 and 2 of the second independent variable are 25 and 25, respectively; the sample reliability coefficients of the dependent variable are .70 and .90, respectively, and the sample correlation coefficients are .24 and .35, respectively. Based on Equations 19-24, the following component variances are obtained under each level of the second independent variable.

$$\text{estimated } \sigma_{IV-1}^2 = (.24^2)(25) = 1.44.$$

$$\text{estimated } \sigma_{R-1}^2 = 25 - (.70)(25) = 7.5.$$

$$\text{estimated } \sigma_{O-1}^2 = (.70)(25) - (.24^2)(25) = 16.06.$$

$$\text{estimated } \sigma_{IV-2}^2 = (.35^2)(25) = 3.0625.$$

$$\text{estimated } \sigma_{R-2}^2 = 25 - (.90)(25) = 2.5.$$

$$\text{estimated } \sigma_{O-2}^2 = (.90)(25) - (.35^2)(25) = 19.4375.$$

In this example, the second independent variable worked in two ways—one by decreasing random variance in the dependent variable (from 7.5 to 2.5) and the other by increasing unexplained systematic variance (from 16.06 to 19.4375). The first change is beneficial for explained variance whereas the second change is not. The

net effect, in this example, was to increase the variance in the dependent variable explained by variance in the independent variable from 1.44 to 3.0625. Expressed as a change in the percentage of variance in the dependent variable explained by the independent variable, the increase is from $1.44/25 = 5.76\%$ to $3.0625/25 = 12.25\%$. It is also possible to express the change in systematic but unexplained variance as a proportion under both levels of the second independent variable: $16.06/25 = 64.24\%$ and $19.4375/25 = 77.75\%$. So it can be seen that the second independent variable increased the proportion of explained variance and the proportion of unexplained but systematic variance, while it decreased random variance (and the proportion of random variance). There is what, according to normal statistical thinking, would seem to be a paradox. Researchers normally assume that if the proportion of “explained variance” increases then the amount of “unexplained” variance decreases. This is a natural consequence of the usual bipartite assumption and keeping the total variance constant. That is, if there are only two sources of variance and total variance is kept constant, any increase in one kind of variance necessitates a corresponding decrease in the other kind of variance. In contrast, the TA leaves open the possibility shown in the example. Even keeping total variance constant, both explained and unexplained (but systematic) variance can increase simultaneously, if both increases come out of a decrease in random error variance.

Conclusion

The first part of the present argument, based on the TA, is that it is possible to derive equations for obtaining σ_{IV}^2 , σ_O^2 , and σ_R^2 in terms of observable data. Researchers have long known how to estimate σ_{IV}^2 . In addition, to estimate σ_O^2 , and σ_R^2 , it also is useful to obtain the unbiased sample variance and to have conducted a reliability study so as to obtain a reasonable value for the reliability of the dependent variable. None of these requirements constitutes an insurmountable obstacle for the researcher.

The second part of the present argument is that it is useful to keep σ_{IV}^2 , σ_O^2 , and σ_R^2 distinct from each other; or more precisely, rather than settle for σ_E^2 , it is useful to decompose it into σ_O^2 and σ_R^2 (see Equation 2). The decomposition can aid the researcher in three general ways. First, it provides researchers with a procedure for making good decisions about the placement of future research efforts; one of the examples indicates that even when σ_E^2 is exactly the same in two cases, σ_O^2 and σ_R^2 nevertheless can suggest dramatically different conclusions about the likely value of future research designed to discover additional independent variables that explain variance in the dependent variable. The increased ability to inform decisions about whether the search for additional variables is likely or unlikely to be fruitful, engendered by the TA, should be a boon not only to basic researchers but also to applied researchers. Second, the decomposition provides additional ways of conceptualizing effect sizes. Depending on which effect size measure is used, the foregoing examples demonstrate how radically different conclusions can be supported.

A third benefit of the TA is that it allows researchers to test the effect of a second independent variable not only on the relationship between the dependent variable and the original independent variable but also on the relationship between the dependent variable and unknown variables that systematically account for some of its variance. The ability to draw conclusions, though limited, about variables of which the researcher has no knowledge, constitutes a nontrivial gain for the social sciences.

The goal in suggesting the TA and the decomposition of traditional “error” variance into σ_O^2 and σ_R^2 was not to argue for the superiority of any one kind of ratio over another, or of any one kind of effect size index over another. On the contrary, researchers might have a variety of goals and different ratios and different sorts of effect size indices are useful for different purposes. The TA and its implications provide researchers with a more expansive set of descriptive tools for the interpretation of the data than does the current set of tools based on an overly simplistic bipartite assumption of dependent variable variance into that which is or is not accounted for by the independent variable.

Acknowledgments

I thank Justin MacDonald for his helpful comments on an earlier version of the article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Lord and Novick (1968) pointed out that relative to other measurement theories, such as item response theory, the classical theory has minimal or weak assumptions. They also pointed out that the disadvantage to minimal assumptions is that less can be accomplished. The advantage of minimal assumptions is that they are less likely to be wrong. Thus, it is an advantage to use the minimal assumptions of the classical theory if these are sufficient to accomplish one’s goal, which they are at present.
2. One way of conceptualizing this is to think of a scenario where a person takes a test, has his or her mind wiped to return it to its previous state, takes the test again, and so on for infinite iterations (Lord & Novick, 1968).
3. A caveat is that the estimate is biased when used at the sample level, which is a topic that will be considered later.
4. Equation (10) is consistent with the famous attenuation formula from the classical theory that shows that reliability places an upper limit on validity. Clearly, $\rho_{YY'}$ must equal or exceed ρ_{XY}^2 to avoid a negative value for σ_O^2 .

5. Like Equation (3), Equation (11) is biased when used at the sample level and this topic will be addressed later.
6. Rosenthal and Rosnow (1991) have emphasized the value of the equation $\left(\rho_{XY}^2 = \frac{T^2}{T^2 + df}\right)$ for converting between t -scores and correlation coefficients.
7. It also is possible for the decrease in randomness not to have proportional effects on the two types of systematic variance but that is not the present topic.

References

- Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge, England: Cambridge University Press.
- Finch, S., Cumming, G., & Thomason, N. (2001). Reporting of statistical practices in the *Journal of Applied Psychology*: Little evidence of reform. *Educational and Psychological Measurement*, 61, 181-210. doi:10.1177/00131640121971167
- Gulliksen, H. (1987). *Theory of mental tests*. Hillsdale, NJ: Lawrence Erlbaum.
- Harris, R. J. (1994). *ANOVA: An analysis of variance primer*. Itasca, IL, F. E. Peacock.
- Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace.
- Hunt, T. (1936). *The value of measurement in psychology*. New York, NY: Prentice Hall.
- Kaplan, R. M., & Saccuzzo, D. P. (2010). *Psychological testing: Principles, applications, and issues* (8th ed.). Belmont, CA: Wadsworth/Cengage Learning.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Michell, J. (1999). *Measurement in psychology*. Cambridge, England: Cambridge University Press.
- Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York, NY: McGraw-Hill.
- Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. New York, NY: Cambridge University Press.
- Rosenthal, R., & Rubin, D. B. (1982). A simple general purpose display of magnitude and experimental effect. *Journal of Educational Psychology*, 74, 166-169. doi: 10.1037/0022-0663.74.2.166
- Russo, F. (2009). *Causality and causal modeling in the social sciences* (Methods Series 5). New York, NY: Springer.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101. Retrieved from <http://webpace.ship.edu/pgmarr/Geo441/Readings/Spearman%201904%20-%20The%20Proof%20and%20Measurement%20of%20Association%20between%20Two%20Things.pdf>
- Thompson, B. (1999). If statistical significance tests are broken/misused, what practices should supplement or replace them? *Theory and Psychology*, 9, 165-181. doi: 10.1177/095935439992006
- Thurstone, L. L. (1959). *The measurement of values*. Chicago, IL: University of Chicago Press.
- Trafimow, D. (2014). Estimating true standard deviations. *Frontiers in Quantitative Psychology and Measurement*, 5, 235. doi:10.3389/fpsyg.2014.00235
- Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, 37, 1-2. doi:10.1080/01973533.2015.1012991

- Trafimow, D., & Rice, S. (2008). Potential performance theory (PPT): A general theory of task performance applied to morality. *Psychological Review*, *115*, 447-462. doi: 10.1037/0033-295X.115.2.447
- Trafimow, D., & Rice, S. (2009). Potential performance theory (PPT): Describing a methodology for analyzing task performance. *Behavior Research Methods*, *41*, 359-371. doi:10.3758/BRM.41.2.359
- Vacha-Haase, T., Nilsson, J. E., Reetz, D. R., Lance, T. S., & Thompson, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. *Theory and Psychology*, *10*, 413-425. doi:10.1177/0959354300103006
- Wilkinson, L., & The APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594-604. doi:10.1037/0003-066X.54.8.594