American Psychological Association
Journal of Applied Psychology

0021-9010/19/$12.00
2019, Vol. 104, No. 4, 593–602
http://dx.doi.org/10.1037/apl0000361

RESEARCH REPORT

The Accuracy of Dominance Analysis as a Metric to Assess Relative Importance: The Joint Impact of Sampling Error Variance and Measurement Unreliability

Michael T. Braun
University of South Florida

Patrick D. Converse
Florida Institute of Technology

Frederick L. Oswald
Rice University

Dominance analysis (DA) has been established as a useful tool for practitioners and researchers to identify the relative importance of predictors in a linear regression. This article examines the joint impact of two common and pervasive artifacts—sampling error variance and measurement unreliability—on the accuracy of DA. We present Monte Carlo simulations that detail the decrease in the accuracy of DA in the presence of these artifacts, highlighting the practical extent of the inferential mistakes that can be made. Then, we detail and provide a user-friendly program in R (R Core Team, 2017) for estimating the effects of sampling error variance and unreliability on DA. Finally, by way of a detailed example, we provide specific recommendations for how researchers and practitioners should more appropriately interpret and report results of DA.

Keywords: relative weight analysis, dominance analysis, predictor importance, multiple regression, Monte Carlo simulation

Supplemental materials: http://dx.doi.org/10.1037/apl0000361.supp

In organizational settings, practical constraints restrict the number of constructs that can be measured, and by fortunate coincidence, good theories tend to be parsimonious, also restricting the number of constructs to measure. Even with a limited number of constructs and measures, however, it is often helpful in the prediction context to determine which predictors appear to be the most empirically useful compared with the others, in terms of explaining the most variance in a single criterion. This relatively simple regression problem has long been known to be mathematically ambiguous, given that predictors are almost always correlated[1] to some extent (i.e., are multicollinear).

A family of well-grounded statistical techniques called *relative importance analysis* serves to address this mathematical and practical problem. *Dominance analysis* (DA; Budescu, 1993) is one of the most utilized relative importance metrics in the organizational sciences, and it is currently the recommended method, due to its mathematically straightforward approach of averaging results across all possible subset regression models (Thomas, Zumbo, Kwan, & Schweitzer, 2014). The DA method allows researchers and practitioners to partition the overall model R^2 into "shares" attributable to each predictor (e.g., Shockley & Allen, 2015), such that predictors with larger shares are considered more important. Even more specifically, researchers will often rank order predictors by their shares and consider the predictors with higher ranks as more important (e.g., Burns, Morris, & Wright, 2014).

In the history of addressing this regression problem, DA has represented a useful advance in providing organizational researchers and practitioners a well-justified method for partitioning the contribution of predicted variance to a set of predictors. Yet it is obvious that estimates from any statistical model are only as good as the data on which they are generated. Two statistical artifacts in particular, *sampling error variance* and *measurement error variance* (i.e., measurement unreliability), are known to add bias and

This article was published Online First October 15, 2018.

Michael T. Braun, Department of Psychology, University of South Florida; Patrick D. Converse, School of Psychology, Florida Institute of Technology; Frederick L. Oswald, Department of Psychology, Rice University.

Correspondence concerning this article should be addressed to Michael T. Braun, Department of Psychology, University of South Florida, 4202 East Fowler Avenue, PCD 4118G, Tampa, FL 33620-7200. E-mail: mtbraun@usf.edu

[1] Predictors are rarely correlated exactly zero, but this would reflect the one case where standardized regression weights are equal to their squared zero-order validity coefficients, and the sum of these weights thus equals the model R^2.

noise to regression parameter estimates (e.g., Cohen, Cohen, West, & Aiken, 2003) and therefore to decrease the interpretability and generalizability of regression coefficients (e.g., Schmidt, 1971). Both sources of error have been shown to have independent and negative effects on results from another type of relative importance analysis (i.e., relative weights analysis; Johnson, 2004). Likewise, sample size has been shown to impact the standard error of dominance weights (Azen & Budescu, 2003). Furthermore, a Monte Carlo demonstration provided in the online supplemental materials showed that the magnitudes of dominance weights can be underestimated by as much as 50% of their true value and that for smaller samples, dominance weights may only be properly rank ordered less than 25% of the time. These known effects, particularly for sampling error variance, have led some organizational methodologists to encourage the reporting of confidence intervals around results from DA and other relative importance methods (e.g., Azen & Budescu, 2003; Tonidandel & LeBreton, 2011, 2015).

Unfortunately, a systematic review of the recent organizational literature clearly demonstrates that relative importance statistics are typically relied upon without reporting or considering their accuracy. Instead, the raw magnitude values of DA (and other relative importance) weights or the raw rank orders of predictor weights are reported and interpreted without qualification. Presumably, researchers and practitioners do not appreciate the practical impact that inaccuracy of DA weights has on making appropriate inferences. A straightforward analogy to this situation would be when researchers find a statistically significant *t* test and then jump to reporting *d* values without considering the fact that *d* values have confidence intervals that affect interpretations of practical significance.

The current study (a) examines the detrimental effect that sampling error variance and measurement error variance jointly have on DA statistics and (b) provides researchers and practitioners with a user-friendly tool to estimate these effects in their own data. First, we detail the general rationale for determining relative importance, focusing on DA. Second, we review how DA and other relative importance metrics are currently being used and reported in the literature. Third, we discuss the detrimental effects of sampling error variance and measurement error variance on regression coefficients and DA weights that are derived from them. Using Monte Carlo simulations, we examine the nature and extent of decreases in the accuracy of general dominance weight rank order and magnitude estimates due to statistical artifacts, highlighting the inferential mistakes that could be made by ignoring these effects. Finally, we detail a user-friendly tool in R (R Core Team, 2017) for estimating the effects of sampling error variance and unreliability on DA and provide an example using the tool, while highlighting specific recommendations for how researchers should interpret and report results.

Relative Importance

Oftentimes, researchers and practitioners have a set of variables that predict a criterion of interest using a regression or structural equation model. Whether for informing subsequent theory (MacCallum, Wegener, Uchino, & Fabrigar, 1993; Stelzl, 1986; Vandenberg & Grelle, 2009), or for selecting the most cost- and time-efficient subset of predictors that optimize criterion-related validity (Madden & Bottenberg, 1963), it is frequently of interest to determine the relative strength of predictors. *Relative importance* is a generic term for this relative strength, defined as the proportionate contribution each predictor in a linear regression model makes to the model R^2, considering (in one of many possible ways) its unique contribution to prediction, along with its contribution when combined with other predictors (Hoffman, 1962). When predictor variables are correlated—and they almost always are to some degree—then there are no unambiguous measures of relative importance. However, several quantitative approaches to relative importance are mathematically and conceptually well motivated and provide meaningful results.

By contrast, perhaps the most typical approach to evaluating predictor importance is eyeballing the magnitude of standardized regression weights or their respective *p* values. Unfortunately, regression weights, even standardized regression weights, do *not* directly indicate predictor importance—and in fact they are not intended to. The vector $\boldsymbol{\beta}$ contains *p* regression weights[2] that together minimize the sum of squared errors of prediction. In other words, given an outcome *y* and *p* predictors in the vector $\boldsymbol{x}$, find:

$$\hat{y} = \sum_{i=1}^{p} \beta_i x_i, \text{ such that } \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \text{ is minimized.}$$

These $\boldsymbol{\beta}$ coefficients maximize prediction or minimize error in the least-squares sense; but when predictors are correlated, regression coefficients are ambiguous measures of variable importance and may even be misleading. In addition to the issue of correlated predictors, there is of course the issue of having a large and representative distribution of sample data to ensure that regression weights are stable and accurate (Green, 1977). We will address the issue of sample size below head-on, but note that in practice, detecting and addressing influential outliers in the data is also of importance, as is the issue of dealing with missing data.

Regardless of the method used in a relative importance analysis, the end goal is to divide the total R^2 into *p* independent components, one for each predictor in the regression model, such that the sum of the relative weights equals either the total R^2 (when weights are components of R^2) or 1.0 (when weights are proportions). Again, there are no unambiguous measures of relative importance when predictors are intercorrelated (Darlington, 1968), and as a result, there are several methods for partitioning R^2 (see Grömping, 2007). For instance, DA is based on an averaging of all subsets of predictors (Budescu, 1993; Kruskal, 1987; Lindeman, Merenda, & Gold, 1980); other methods take other approaches, such as creating a set of maximally correlated orthogonal predictors (relative weights analysis; Genizi, 1993; Johnson, 2000); or simply multiplying the zero-order criterion-related validity by its regression weight (Hoffman, 1960; Pratt, 1987); still other methods have suggested taking a weighted average of each predictor's all-subset contributions to R^2 (Feldman, 2005; Ortmann, 2000).

Dominance Analysis

In the organizational sciences, currently the two most common empirical methods of determining relative importance are DA and

[2] Assuming standardized regression weights here, where there is no intercept. Having a model with an intercept would not change the points made here.

relative weight analysis (RWA). These two approaches consistently produce highly similar statistical results (e.g., Budescu & Azen, 2004; LeBreton, Ployhart, & Ladd, 2004; LeBreton & Tonidandel, 2008), even producing identical values when only two predictors are included in the model (Braun & Oswald, 2011). The present study focuses on DA because it is currently the recommended relative importance statistic due to a fundamental flaw in the mathematical foundation of RWA (Thomas et al., 2014). However, given the similarity often found in DA and RWA results, and given the common use of RWA in prior research, the tool described later in the article can produce either DA or RWA results. DA has been extended from multiple linear regression to a wide variety of statistical situations, including multiple dependent variables (Azen & Budescu, 2006), suppression effects (Azen & Budescu, 2003), logistic regression (Azen & Traxel, 2009), and multilevel modeling (Luo & Azen, 2013).

Chevan and Sutherland (1991) and Budescu (1993) detailed the procedure for calculating relative weights for each of p predictors based on DA, which involves computing each predictor's incremental validity across all $2^p - 1$ submodels that involve that predictor and using the set of incremental validity values that result. *General dominance weights* refer to the predictor's average incremental validity across all possible submodels. That is, for a given predictor, all models containing that predictor are generated. For each of these models, the incremental validity (ΔR^2) of that predictor is calculated when it is entered last in the model. Then, those incremental validities are averaged across all of these submodels. This process is repeated for each predictor in the model. General dominance weights have two appealing properties. First, each general dominance weight reflects the general definition of relative importance previously mentioned: that is, the average contribution of a predictor to a criterion, both on its own and when taking all other predictors in the model into account. Second, general dominance weights always sum to the overall model R^2 or to 1.0 if one divides each weight by the sum of the weights.

General dominance weights are the most commonly reported metric from DA and therefore are the values considered in this study. With general dominance weights, a predictor X_j, is said to dominate another predictor, X_i, when its dominance weight (average incremental R^2 across all submodels) is larger. Relative importance can then be established by rank ordering the p dominance weights. Note that there are two stricter forms of dominance that require a deeper look at the regression submodels. For *conditional dominance*, predictor X_j only dominates predictor X_i when its incremental validity is greater, on average, for *each* family of submodels of the same size. The strictest form of dominance is *complete dominance*, where predictor X_j only dominates predictor X_i when it demonstrates greater incremental validity within *all* submodels containing both predictors (Budescu & Azen, 2004).

Use and Reporting of Relative Importance Statistics

To illustrate recent trends in the use and reporting of relative importance weights, we conducted a literature search using PsycINFO, focusing on the years 2015 through 2017[3]; the terms *dominance analysis*, *dominance analyses*, *relative weight*, and *relative weights*; and five major organizational journals, *Academy of Management Journal*, *Journal of Applied Psychology*, *Journal of Management*, *Organizational Behavior and Human Decision Processes*, and *Personnel Psychology*. This resulted in 12 articles reporting relative importance analysis results. Even within this limited search and set of articles, two themes are noteworthy. First, most of these articles (n = 8, 67%) did not report confidence intervals for the relative importance weights; they were taken at face value. Thus, in most cases, the potential for inaccuracy (variability) in these estimates of importance weights was not directly addressed.

Second, almost all of these articles (n = 11, 92%) presented the relative importance results in terms of both (a) the relative importance of the predictors' weights (i.e., ranking the weights or relative proportions of R^2) as well as (b) the magnitude of the predictors' weights (i.e., considering the practical significance of the amount of R^2 reflected in the weights). For example, in a study of emotional intelligence (EI), Joseph, Jin, Newman, and O'Boyle (2015) addressed both of these issues, noting, "the most important predictors of mixed EI, in order, are Emotional Stability (29.5%), Extraversion (26.5%), Conscientiousness (16.1%), self-rated performance (14.2%), general self-efficacy (6.8%), and ability EI (5.5%)" (p. 309). Shockley and Allen (2015) examined work–family conflict decisions using DA and reported that "Although the work variables accounted for more of the variance than did the family variables, the differences are negligible (51.36% vs. 48.64%)" (p. 300). As one other example, Chamberlin, Newton, and LePine (2017) investigated voice and noted, "Personal initiative is the most important antecedent in this category, accounting for 50.6% of the explained variance in voice. The second most important predictor is core self evaluation (CSE), but this variable explains far less unique variance in voice (11.4%)" (p. 30). Given this typical approach to reporting relative importance weights, the present study focuses on both the rank order and magnitude of weights stemming from DA.

Statistical Artifacts and Relative Importance

Methodological researchers have demonstrated either systematically (e.g., Johnson, 2004) or tangentially (e.g., Azen & Budescu, 2003) how the statistical quality of data meaningfully impacts the accuracy of relative importance weights and subsequent inferences made from them. Yet we just demonstrated how studies report and interpret relative importance weights and their rank orders as a "bottom-line" message, without qualifying them by the accuracy of those weights. Even when weights are not statistically or reliably distinguishable from one another, they are almost always treated as distinctive. We believe this phenomenon is due to what relative weights promise: some definitive answer about predictor importance. This promise is a problem endemic to our field; any misuse of relative weights is not the fault of any particular researcher.

Having this context freshly in mind hopefully leads to a greater appreciation of the goal of this article: to highlight the potential inaccuracy of the magnitude and ordering of relative weights due to the joint impact of sampling error variance and measurement error variance, as they are two statistical artifacts that affect every

[3] The year 2015 was chosen as the starting point for our search to allow time for the two most recently published articles (i.e., Tonidandel & LeBreton, 2011, 2015), encouraging researchers to report confidence intervals from relative importance analyses and interpret output with caution.

organizational study. To do so, we review prior research on the effects of sampling error and measurement error variance on regression results and relative importance indices and apply them to discuss their joint effect on DA (as supported by Monte Carlo simulation results available in online supplemental material). Likewise, we provide a tool in R code that calculates the effect that these two artifacts have on DA accuracy; we also detail an example of using the tool and reporting its results so that readers can better and more consistently qualify and otherwise interpret relative importance findings in their research.

Sampling Error Variance

Sampling error variance creates random error in estimating any statistic, including regression weights. Relative weights are derived directly from least-squares regression weights, and it has long been known that regression weights are prone to high levels of inaccuracy when sample sizes are small. In his classic article, Schmidt (1971) found that, for small samples (e.g., $N < 100$ for six predictors), unit weights could outperform sample-derived regression weights in terms of least-squares prediction in the population, because the latter would capitalize on chance fluctuations in the data. Relative weights should be similarly affected, if not more so: To the extent that predictors vary subtly in their relative importance, larger sample sizes will be required to detect those differences. Conversely, smaller sample sizes may be reasonable and adequate for determining variable importance whenever the predictors vary substantially in their importance, and/or when one needs to make simpler practical decisions with regard to predictors in a model (e.g., having to determine the three most important predictors in a five-predictor model, without concern about their specific rank ordering).

With respect to relative importance specifically, Azen and Budescu (2003) noted that smaller samples will likely increase the size of standard errors around the weights from DA, leading to less stability for determining complete dominance, as defined above. In the context of RWA, Johnson (2004) bootstrapped the standard errors of the relative importance weights of ratings of specific dimensions of job performance predicting overall job performance for two types of jobs: secretarial and staff support and customer service and sales jobs. Not surprisingly, smaller samples produced larger standard errors around RWA estimates, leading to unstable estimates of the magnitudes and rank orderings of the weights.

Measurement Unreliability

Measurement unreliability (error variance) is also an omnipresent and nontrivial factor in psychological measurement; it attenuates the strength of observed correlations between variables and also can distort observed patterns of underlying convergent and discriminant validities between constructs. Psychometric corrections for measurement unreliability can help provide more accurate statistical estimates of underlying relationships on average (Schmidt & Hunter, 1996). The correction does not come for free, however; it comes at the cost of larger standard errors and confidence intervals in the resulting estimates (Hunter & Schmidt, 2004; Oswald, Ercan, McAbee, Ock, & Shaw, 2015). Of course, accurate correction for measurement unreliability depends critically on having the correct conception and measurement of reliability in the first place. Reliability is conceptualized and modeled not as one phenomenon but several (see LaHuis & Avis, 2007; Le, Schmidt, & Putka, 2009; Murphy & DeShon, 2000): Alpha reliability reflects the coherence or dimensionality of items at a single time point, test–retest reliability reflects the stability of a specific measure over time (independent of dimensionality), and alternate-forms reliability estimates the stability of construct measurement (independent of specific content). Similarly, when considering the reliability of observer ratings, factors such as transient error, rater variance, and Rater × Ratee interactions may be quantified and considered.

Focusing on relative importance indices outside of DA, Johnson (2004) evaluated the effect of measurement error variance on results from RWA. Using bootstrapping, he corrected the observed sample correlation matrix by the estimated reliability for both predictors and criteria and then examined differences in the corrected and uncorrected weights (after they were transformed to sum to 1 in each RWA). Differences in the relative ranking between corrected and uncorrected weights were found (i.e., low rank-order correlations) as well as differences in the magnitudes of the weights (i.e., high squared Euclidean distances), with these differences being largely a joint function of the mean validity, number of predictors, and mean reliability. Inspired by the spirit of this work, we examined the joint effects of sampling and measurement error variance on results from DA and subsequent inferences, as will be described.

The Joint Impact of Sampling Error Variance and Measurement Unreliability

Turning to DA, sampling error variance impacts the stability of DA weight estimates. As is the case with standardized regression coefficients, smaller samples produce more variability in estimated relationships (e.g., Schmidt, 1971) and limit the ability of the model to determine statistical significance (e.g., Cohen et al., 2003). As such, sampling error variance has little effect on the *average* weights produced by DA across samples, but can substantially impact the *stability* in the rank order of the weights, which is directly related to the ability to determine whether weights are significantly different from zero or from one another. Although some researchers (e.g., LeBreton et al., 2004) have prescriptively argued that the magnitude of the weights is most important, our previously described literature review revealed that descriptively, many researchers uniquely attend to the relative rank ordering of weights (e.g., Burns et al., 2014). As such, the variability of rank ordering produced by sampling error variance can drastically impact the substantive interpretations from DA output. For example, in the output presented in the online supplemental material, even for large samples and perfect reliability, DA failed to rank order predictors perfectly in 25% of all cases.

Although sampling error variance affects the variability in DA weights, so that they may be higher or lower, measurement error variance lowers (attenuates) correlations (Murphy & Davidshofer, 2005) and usually the DA weights that result, such that the more measurement error variance in each corresponding instrument, the more the DA weight will be underestimated. In the simulated example in the online supplemental materials, the DA weights were underestimated by as much as 50% when criterion reliability was rather low. This can be particularly problematic when trying to determine whether weights significantly differ from zero (e.g., Tonidandel, LeBreton, & Johnson, 2009) or one another (e.g., Johnson, 2004) as they are all biased toward zero, potentially creating a floor effect. Complicating matters, each predictor and the criterion likely have unique amounts of measurement error (e.g., Roth, Switzer, Van Iddekinge, & Oh, 2011), leading to each zero-order validity coefficient being differentially affected. Even more troubling, the relationships among the predictors are also affected by measurement error variance, creating patterns of interrelationships between measures that can be drastically different from their true underlying construct-level relationships. Taken together, the proportionate amount of variance explained in the dependent variable estimated by the DA weights and the corresponding rank order of weights is likely inaccurate in many instances. The exact severity of the inaccuracy as well as its impact on substantive inferences depend on the specific sets of relationships under investigation as well as the measurement error associated with each instrument.

For a full empirical demonstration of the joint impact of sampling error variance and measurement unreliability on the magnitude and ordering of general dominance weights as well as the effects on determining whether weights significantly differ from zero and from one another, refer to the Monte Carlo simulation presented in the online supplemental materials.

Reporting and Interpreting Relative Importance Results

Due to the prevalence of sampling error variance and measurement unreliability in organizational studies, it would be quite easy to apply DA to data and conclude that (a) there is a grossly incorrect rank ordering of weights in the sample compared with the population weights, (b) predictors are important in the sample yet are trivial in the population, and/or (c) predictors are trivial in the sample yet are important in the population. Any of these situations create the potential for poor user decision-making due to biased results, undermining the benefits that researchers have hoped to achieve by conducting DA and interpreting the resulting weights and ranks. Essentially, this problem is a modern version of the need to avoid overinterpreting regression weights when sample sizes are lower, where Schmidt (1971) showed that unit weights often suffice. The analog for DA is that one should not take DA weights at face value; instead, it might be more reasonable to distinguish more versus less important predictors in general rather than by their specific rank order, and in some cases, when predictors are all similarly important and/or statistical artifacts are strong, the results might suggest not making any judgments about relative importance. Thus, conclusions regarding relative importance in future research should be more informed by taking inaccuracy of DA results into account.

It seems obvious that statistical output should be interpreted carefully and findings should not be overgeneralized; however, with respect to relative importance, our literature review revealed that currently this is not typically being done in the organizational literature. Given the pervasive influences of statistical artifacts on the accuracy of relative importance output in most organizational research, it is highly recommended that users of DA or other approaches to relative importance compute and report 95% confidence intervals for each weight, indicating the level of inaccuracy expected in the estimates. Likewise, rather than simply reporting and interpreting the rank order of weights that are generated from the DA without any qualification (as tends to be the case), users should report the average rank order obtained across bootstrapped replications, along with associated 95% confidence intervals (following our demonstration).

A User-Friendly Tool

To aid researchers and practitioners in better reporting and interpreting relative importance statistics, a user-friendly tool (domWeightTool.R) in the R programming language was built and is available as a supplemental file. This way, researchers and practitioners can use it to help generate and interpret output from either general DA or RWA, thus improving the accuracy (and qualifications) of their resulting inferences. Output from the tool equips researchers with information on the stability of rank ordering and relative weight magnitudes, as well as the frequency with which each predictor was found to significantly differ from zero as well as from one another. This information will allow users to follow the recommendations of this article and make better decisions regarding predictor importance.

The tool is built as a user-defined function in R and requires the packages *yhat*, *boot*, and *MASS* be installed to run.[4] Once the packages are installed, the R file should be downloaded, opened, and executed by highlighting the entire script and selecting "run." Because the tool is a function, executing the script will appear to do nothing. However, doing so initiates the function *DW.accuracy()*, which takes up to nine inputs from the user to test a variety of different conditions depending on the user's data. Each input is described in detail in Appendix A.

The four most important inputs are the user-provided data (required), the sample size (optional/required), and the reliability of the predictors and criterion (both optional). Separate from executing the function, it is necessary for users to have their data saved as an object in R. There are no limits to the number of predictors[5] that can be considered, but only a single criterion can be entered at a time. Users can enter data in two ways. First, raw data in the form of a traditional data table with observations as rows and variables as columns, which must be saved in R as a "data.frame" object. The data.frame can only contain the predictors and single criterion of interest. If different models are to be considered, each must be saved as unique data.frame objects. It is best if each data.frame is formatted such that the first set of columns contain the predictors and the final column contains the criterion. Second, a correlation matrix and the number of observations to be considered. Like the data.frame, the correlation matrix should include only relevant predictors and a single criterion. The correlation matrix can be entered either as a matrix or as a vector, as the function will reshape the vector into a square matrix. If the user provides a complete dataset, the tool will bootstrap from the user's data to estimate standard errors for all

[4] As such, prior to downloading and running the function, users must install each of these three packages in R (using the command *install.packages()*).

[5] Of course, including more predictors increases simulation time, challenges current computer processing resources (especially as the number of predictors approaches 15–20), and still requires a positive definite correlation matrix.

Table 1
Uncorrected and Corrected Correlation Matrices

Variables	1	2	3	4	5
1. Cognitive ability	—	.39	.04	.45	.49
2. Structured interview	.31	—	.17	.21	.48
3. Conscientiousness	.03	.13	—	.64	.21
4. Biodata	.37	.16	.51	—	.36
5. Job performance	.40	.37	.17	.29	—

Note. The uncorrected matrix is in the lower triangle and the corrected matrix is in the upper triangle. To make the corrections we used alpha reliability estimates of .85 for cognitive ability, .75 for structured interviews, .80 for conscientiousness, .80 for biodata, and .80 for job performance.

relationships, thereby not assuming a multivariate normal distribution. However, if only a correlation matrix is provided then all data are assumed to be multivariate normal.

In addition to the data, sample size represents the number of observations in the user's data (assuming no missing values). This input is required if the user only provides a correlation matrix as it is necessary to properly simulate data. Alternatively, this may be left blank if the user provides a complete dataset. Predictor reliability is a vector the length of the number of predictor variables in which each value in the vector is a scalar from 0 to 1 that represents a reliability estimate (e.g., coefficient alpha). Finally, criterion reliability similarly represents a reliability estimate as a single scalar between 0 and 1 for the dependent variable. If the user wants to have the sample correlation matrix considered, the user may leave the inputs for both predictor and criterion reliability blank as their default values are 1 (perfect reliability). Users may be interested in considering the sample correlation matrix because they are interested in either (a) representing the observed correlations only (e.g., when no reliability correction is available) or (b) representing latent correlations only (i.e., when no reliability correction is needed). The additional inputs to the function include whether to consider general dominance weights or relative importance weights. We recommend saving the output of the function to an object to aid in extracting relevant pieces of information.

The output of the function provides six pieces of information designed to aid users in their interpretation of DA. The first two objects in the output are called avg.weights and avg.ranks. They contain the means, standard deviations, and 95% confidence intervals for the relative importance weights and ranks, respectively, associated with each predictor and avg.weights also contains the overall model R^2. The third output object, diff.zero, provides the proportion of simulation runs where each focal predictor is found to be statistically significantly different from zero. The fourth output object, sig.diffs, provides the proportion of times each pair of predictors is found to be significantly different from one another. The fifth and sixth output objects, raw.weights and raw.ranks, provide the raw weights (along with model R^2) and ranks, respectively, from each individual simulation run so that any additional statistics or visualization can be computed. Each of these pieces of output directly relates to the aforementioned recommendations for additional detail to report in technical reports and publications, so that all decisions made using DA can be better qualified and interpreted.

Reporting Results From Relative Importance Analyses: An Example

To demonstrate the utility of the current tool, and to aid researchers and practitioners in using its functions and interpreting its output, an example using the completely uncorrected correlation matrix from the Roth et al. (2011) meta-analysis that served as the basis for the Monte Carlo simulation is provided (see Table 1, lower triangle). For the purposes of this example, we wanted to demonstrate the ability to evaluate relative importance using the corrected and uncorrected versions of a correlation matrix. To do so, we corrected all of the observed correlations by their respective reliability estimates (Table 1, upper triangle). We used alpha reliability estimates of .85 for cognitive ability (e.g., Condon & Revelle, 2014), .75 for structured interviews (e.g., Levashina, Hartwell, Morgeson, & Campion, 2014), .80 for conscientiousness

Table 2
General Dominance Analysis Weight and Rank Values for the Uncorrected and Corrected Correlation Matrices

	Uncorrected matrix				Corrected matrix			
		95% CI				95% CI		
Predictors	*M* (*SD*)	*LL*	*UL*	Sig	*M* (*SD*)	*LL*	*UL*	Sig
Weights								
Cognitive ability	.11 (.03)	.06	.18	98%	.14 (.03)	—	—	100%
Structured interview	.10 (.03)	.05	.15	92%	.14 (.03)	—	—	98%
Conscientiousness	.02 (.01)	.00	.05	7%	.02 (.01)	—	—	25%
Biodata	.04 (.02)	.01	.08	42%	.06 (.02)	—	—	86%
R^2	.27 (.04)	.20	.34		.34 (.05)	—	—	
Ranks								
Cognitive ability	1.49 (.56)	1	2		1.53 (.52)	—	—	
Structured interview	1.60 (.62)	1	3		1.50 (.54)	—	—	
Conscientiousness	3.94 (.46)	3	5		3.99 (.17)	—	—	
Biodata	3.06 (.49)	2	4		2.99 (.22)	—	—	

Note. CI = confidence interval; sig = proportion of runs that the predictor was found to be significantly different from zero (i.e., the spurious predictor); LL = lower limit; UL = upper limit. The spurious predictor that was used to test for significant differences from zero was excluded from the table. Mean represents the average value across all simulated runs. All values were based on 100 simulated runs.

(e.g., Johnson, 2014), .80 for biodata (e.g., Rogelberg, 2006), and .80 for job performance (e.g., Ones, Viswesvaran, & Schmidt, 2008) to make the corrections. Starting with the uncorrected correlation matrix (named corMatrix), we estimated general dominance weights for a sample size of 250 and opted to test for whether the weights significantly differed from zero. Appendix B provides a description of the process of executing the example with code.

When possible, it is helpful to examine the relative importance output from both an uncorrected and corrected[6] correlation matrix. Output from the uncorrected matrix should be used to understand the impact of sampling error variance by examining the stability of weight and rank estimates. Differences between the corrected and uncorrected matrix can be used to estimate the impact of measurement error variance. Depending on the goal of the user, it may be more conceptually meaningful to correct for only criterion unreliability or for unreliability in both the predictors and criterion. In accordance with our recommendations, we examine the rank and weight output from the general DA for an uncorrected matrix and a fully corrected matrix. Tables 2 and 3 present the results. Consistent with simulated results, the uncorrected weights are smaller than the corrected weights due to attenuation from measurement error variance. Because the same sample size was used in both calculations, the standard deviations of the weight estimates were approximately equal regardless of correction.

Examining the output, the uncorrected matrix makes it appear as though cognitive ability is more dominant, especially with respect to the structured interview, compared to the corrected matrix. Using the uncorrected matrix, the difference in average weights is larger (.01 compared to .00 in the corrected matrix), the difference in average ranks is larger (.11 compared to .03), the difference in the proportion of times each is significantly different from zero is larger (6% compared to 2%), and they are significantly different from one another in 11% of all cases. Alternatively, using the corrected matrix, cognitive ability and the structured interview appear approximately equal with near equivalent weight, rank, and significance values along with being statistically distinct in only 4% of cases. This example highlights the dangers of ignoring statistical artifacts when interpreting relative importance indices. Raw data can easily make relationships appear systematically larger and more impactful or smaller and less impactful than they would be if all information was considered. As such, it is highly recommended that researchers and practitioners use the provided tool, generate all potential output, and present tables like those in this example when calculating relative importance so that results can be properly interpreted and inferences can be adequately validated. Applied researchers appear to be perpetually interested in understanding predictor variable importance in linear regression models. Thus, we hope that the current tool will allow researchers to communicate their understanding, not only clearly but also carefully, given that dominance weights will always have some instability in their magnitude and rank ordering across predictors.

Table 3
Significant Differences Between Predictors for the Uncorrected and Corrected Correlation Matrices

Predictors	1	2	3	4
1. Cognitive ability	—	4%	92%	70%
2. Structured interview	11%	—	90%	59%
3. Conscientiousness	78%	65%	—	43%
4. Biodata	50%	35%	15%	—

Note. The values represent the percentage of times that the two predictors were found to be significantly different from one another. Values on the lower triangle are from the uncorrected matrix and values on the upper triangle are from the corrected matrix. All values were based on 100 simulated runs.

[6] Because correction for unreliability inflates the impact of sampling error variance and no known correction for this upward bias on dominance weight estimates is currently available, 95% confidence intervals are only displayed for the uncorrected matrix.

References

Arthur, W., Jr., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology, 93,* 435–442. http://dx.doi.org/10.1037/0021-9010.93.2.435

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods, 8,* 129–148. http://dx.doi.org/10.1037/1082-989X.8.2.129

Azen, R., & Budescu, D. V. (2006). Comparing predictors in multivariate regression models: An extension of dominance analysis. *Journal of Educational and Behavioral Statistics, 31,* 157–180. http://dx.doi.org/10.3102/10769986031002157

Azen, R., & Traxel, N. (2009). Using dominance analysis to determine predictor importance in logistic regression. *Journal of Educational and Behavioral Statistics, 34,* 319–347. http://dx.doi.org/10.3102/1076998609332754

Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology, 74,* 478–494. http://dx.doi.org/10.1037/0021-9010.74.3.478

Braun, M. T., & Oswald, F. L. (2011). Exploratory regression analysis: A tool for selecting models and determining predictor importance. *Behavior Research Methods, 43,* 331–339. http://dx.doi.org/10.3758/s13428-010-0046-8

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin, 114,* 542–551. http://dx.doi.org/10.1037/0033-2909.114.3.542

Budescu, D. V., & Azen, R. (2004). Beyond global measures of relative importance: Some insights from dominance analysis. *Organizational Research Methods, 7,* 341–350. http://dx.doi.org/10.1177/1094428104267049

Burns, G. N., Morris, M. B., & Wright, C. P. (2014). Conceptual and statistical interactions: An illustration with the AB5C and CWBs. *Journal of Business and Psychology, 29,* 47–60. http://dx.doi.org/10.1007/s10869-013-9287-8

Chamberlin, M., Newton, D. W., & LePine, J. A. (2017). A meta-analysis of voice and its promotive and prohibitive forms: Identification of key associations, distinctions, and future research directions. *Personnel Psychology, 70,* 11–71. http://dx.doi.org/10.1111/peps.12185

Chevan, A., & Sutherland, M. (1991). Hierarchical partitioning. *The American Statistician, 45,* 90–96.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence, 43,* 52–64. http://dx.doi.org/10.1016/j.intell.2014.01.004

Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin, 69,* 161–182. http://dx.doi.org/10.1037/h0025471

Feldman, B. (2005). *Relative importance and value* (Version 1.1). Unpublished manuscript, Kellstadt Graduate School of Business, DePaul University, Chicago, IL. http://dx.doi.org/10.2139/ssrn.2255827

Genizi, A. (1993). Decomposition of R^2 in multiple regression with correlated regressors. *Statistica Sinica, 3,* 407–420.

Green, B. F. (1977). Parameter sensitivity in multivariate methods. *Multivariate Behavioral Research, 12,* 263–287. http://dx.doi.org/10.1207/s15327906mbr1203_1

Grömping, U. (2007). Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician, 61,* 139–147. http://dx.doi.org/10.1198/000313007X188252

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin, 57,* 116–131. http://dx.doi.org/10.1037/h0047807

Hoffman, P. J. (1962). Assessment of the independent contributions of predictors. *Psychological Bulletin, 59,* 77–80. http://dx.doi.org/10.1037/h0044373

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage. http://dx.doi.org/10.4135/9781412985031

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research, 35,* 1–19. http://dx.doi.org/10.1207/S15327906MBR3501_1

Johnson, J. W. (2004). Factors affecting relative weights: The influence of sampling and measurement error. *Organizational Research Methods, 7,* 283–299. http://dx.doi.org/10.1177/1094428104266018

Johnson, J. A. (2014). Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51,* 78–89. http://dx.doi.org/10.1016/j.jrp.2014.05.003

Joseph, D. L., Jin, J., Newman, D. A., & O'Boyle, E. H. (2015). Why does self-reported emotional intelligence predict job performance? A meta-analytic investigation of mixed EI. *Journal of Applied Psychology, 100,* 298–342. http://dx.doi.org/10.1037/a0037681

Kruskal, W. (1987). Relative importance by averaging over orderings. *The American Statistician, 41,* 6–10.

LaHuis, D. M., & Avis, J. M. (2007). Using multilevel random coefficient modeling to investigate rater effects in performance ratings. *Organizational Research Methods, 10,* 97–107. http://dx.doi.org/10.1177/1094428106289394

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria. What did they really say? *Organizational Research Methods, 9,* 202–220. http://dx.doi.org/10.1177/1094428105284919

Le, H., Schmidt, F. L., & Putka, D. J. (2009). The multifaceted nature of measurement artifacts and its implications for estimating construct-level relationships. *Organizational Research Methods, 12,* 165–200. http://dx.doi.org/10.1177/1094428107302900

LeBreton, J. M., Ployhart, R. E., & Ladd, R. T. (2004). A Monte Carlo comparison of relative importance methodologies. *Organizational Research Methods, 7,* 258–282. http://dx.doi.org/10.1177/1094428104266017

LeBreton, J. M., & Tonidandel, S. (2008). Multivariate relative importance: Extending relative weight analysis to multivariate criterion spaces. *Journal of Applied Psychology, 93,* 329–345. http://dx.doi.org/10.1037/0021-9010.93.2.329

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology, 67,* 241–293. http://dx.doi.org/10.1111/peps.12052

Lindeman, R. H., Merenda, P. F., & Gold, R. Z. (1980). *Introduction to bivariate and multivariate analysis*. Glenview, IL: Scott Foresman.

Luo, W., & Azen, R. (2013). Determining predictor importance in hierarchical linear modeling using dominance analysis. *Journal of Educational and Behavioral Statistics, 38,* 3–31. http://dx.doi.org/10.3102/1076998612458319

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114,* 185–199. http://dx.doi.org/10.1037/0033-2909.114.1.185

Madden, J. M., & Bottenberg, R. A. (1963). Use of an all possible combination solution of certain multiple regression problems. *Journal of Applied Psychology, 47,* 365–366. http://dx.doi.org/10.1037/h0040365

Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Murphy, K. R., & DeShon, R. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology, 53,* 873–900. http://dx.doi.org/10.1111/j.1744-6570.2000.tb02421.x

Nimon, K. F., & Oswald, F. L. (2013). Understanding the results of multiple linear regression: Beyond standardized coefficients. *Organizational Research Methods, 16,* 650–674. http://dx.doi.org/10.1177/1094428113493929

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2008). No new terrain: Reliability and construct validity of job performance ratings. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1,* 174–179. http://dx.doi.org/10.1111/j.1754-9434.2008.00033.x

Ortmann, K. M. (2000). The proportional value of a positive cooperative game. *Mathematical Methods of Operations Research, 51,* 235–248. http://dx.doi.org/10.1007/s001860050086

Oswald, F. L., Ercan, S., McAbee, S. T., Ock, J., & Shaw, A. (2015). Imperfect corrections or correct imperfections? Psychometric corrections in meta-analysis. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 8,* 1–4. http://dx.doi.org/10.1017/iop.2015.17

Pratt, J. W. (1987). Dividing the indivisible: Using simple symmetry to partition variance explained. In T. Pukilla & S. Duntaneu (Eds.), *Proceedings of second Tampere conference in statistics* (pp. 245–260). Finland: University of Tampere.

Putka, D. J., & Hoffman, B. J. (2014). "The" reliability of job performance ratings equals 0.52. In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends* (pp. 247–275). New York, NY: Routledge.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rogelberg, S. G. (2006). *Encyclopedia of industrial and organizational psychology*. Thousand Oaks, CA: Sage Publications.

Roth, P. L., Switzer, F. S., III, Van Iddekinge, C. V., & Oh, I.-S. (2011). Toward better meta-analytic matrices: How input values can affect research conclusions in human resource management simulations. *Personnel Psychology, 64,* 899–935. http://dx.doi.org/10.1111/j.1744-6570.2011.01231.x

Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement, 31,* 699–714. http://dx.doi.org/10.1177/001316447103100310

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1,* 199–223. http://dx.doi.org/10.1037/1082-989X.1.2.199

Shockley, K. M., & Allen, T. D. (2015). Deciding between work and family: An episodic approach. *Personnel Psychology, 68,* 283–318. http://dx.doi.org/10.1111/peps.12077

Stelzl, I. (1986). Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research, 21,* 309–331. http://dx.doi.org/10.1207/s15327906mbr2103_3

Thomas, D. R., Zumbo, B. D., Kwan, E., & Schweitzer, L. (2014). On Johnson's (2000) relative weights method for assessing variable importance: A reanalysis. *Multivariate Behavioral Research, 49,* 329–338. http://dx.doi.org/10.1080/00273171.2014.905766

Tonidandel, S., & LeBreton, J. M. (2011). Relative importance analysis: A useful supplement to regression analysis. *Journal of Business and Psychology, 26,* 1–9. http://dx.doi.org/10.1007/s10869-010-9204-3

Tonidandel, S., & LeBreton, J. M. (2015). RWA Web: A free, comprehensive, web-based, and user-friendly tool for relative weight analyses. *Journal of Business and Psychology, 30,* 207–216. http://dx.doi.org/10.1007/s10869-014-9351-z

Tonidandel, S., Lebreton, J. M., & Johnson, J. W. (2009). Determining the statistical significance of relative weights. *Psychological Methods, 14,* 387–399. http://dx.doi.org/10.1037/a0017735

Trattner, M. H., & O'Leary, B. S. (1980). Sample sizes for specified power in testing for differential validity. *Journal of Applied Psychology, 65,* 127–134. http://dx.doi.org/10.1037/0021-9010.65.2.127

Vandenberg, R. J., & Grelle, D. M. (2009). Alternative model specifications in structural equation modeling: Facts, fictions and truth. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 165–192). New York, NY: Routledge.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81,* 557–574. http://dx.doi.org/10.1037/0021-9010.81.5.557

Wilcox, R. R., & Tian, T. (2008). Comparing dependent correlations. *The Journal of General Psychology, 135,* 105–112. http://dx.doi.org/10.3200/GENP.135.1.105-112

Appendix A

DW.accuracy Function Documentation

Description

Calculate the accuracy of general dominance analysis or relative weight analysis weights and ranks using Monte Carlo simulation.

Usage

DW.accuracy(userDat,n.obs,iv.relia,dv.relia,iv.names,whichCor, spurIV,epsilon,n.sims)

Arguments

userDat	Either a data.frame object containing observations on the relevant predictors and a single dependent variable or a correlation matrix from observed data. If entering a correlation matrix, it may be in the form of a matrix object or a vector. This is required for the function to run.
n.obs	The number of observations in the observed dataset. This is required for the function to run if users enter a correlation matrix. It is irrelevant if users enter a data.frame.
iv.relia	A vector containing the reliability estimates for each independent variable (predictor). Default value is a vector of all 1's, indicating perfect reliability (or that reliability is not being modeled).
dv.relia	A value representing the reliability estimate for the dependent variable (criterion). Default value is set to 1, indicating perfect reliability (or that reliability is not being modeled).
iv.names	A vector of characters identifying the variable names to be used for the independent variables. Default values are "X" followed by a number (e.g., ×1, ×2).
whichCor	Variable to indicate which correlation matrix to use. 0 = observed (do not correct for unreliability in any variables), 1 = fully corrected which corrects for unreliability in both the independent and dependent variables, 2 = operational validity (Binning & Barrett, 1989) which only corrects for unreliability in the criterion. This only matters if values are inserted for iv.relia and dv.relia; otherwise, it will provide identical estimates.
spurIV	Logical statement indicating whether the user wishes to test whether each predictor weight is significantly different from zero by inserting a random predictor consistent with the procedure described in Tonidandel et al. (2009). Default value is TRUE.
epsilon	Logical statement indicating whether the user wishes to estimate relative importance statistics from Relative Weight Analysis (Johnson, 2000). Default value is FALSE.
n.sims	The number of simulated runs to execute. default value is 1000.

(*Appendices continue*)

Value

DW.accuracy returns an object of class "list" containing the following elements:

avg.weights	A data frame containing the means, standard deviations, upper and lower critical values (95%) for each predictor's weight, and overall model R^2.
avg.ranks	A data frame containing the means, standard deviations, and upper and lower critical values (95%) for each predictor's rank order.
diff.zero	A numeric vector containing the proportion of simulated runs where each predictor was determined to be significantly different from zero. This will only contain values when spurIV = T, noting that a spurious predictor was inserted into the model.
sig.diffs	A numeric vector containing the proportion of simulated runs where each predictor was found to be significantly different from every other predictor.
raw.weights	A data frame of the raw general dominance weights for each predictor along with the model R^2 value.
raw.ranks	A data frame of the raw general dominance ranks for each predictor.

Appendix B

Example Data Generation

The final function call for estimating relative importance through general dominance weights with the uncorrected correlation values and saving it as an object (uncorOut) is as follows: uncorOut <- *DW.accuracy*(userDat, n.obs = 250, iv.relia = c(.85,.75,.80,.80), dv.relia = .80, iv.names = c("cog", "inter", "consc", "biodata"), whichCor = 0, spurIV = T, epsilon = F, n.sims = 100).[7]

The final function call for estimating relative importance through general dominance weights with the corrected correlation values and saving it as an object (corOut) is as follows: corOut <- *DW.accuracy*(userDat, n.obs = 250, iv.relia = c(.85,.75,.80,.80), dv.relia = .80, iv.names = c("cog", "inter", "consc", "biodata"), whichCor = 1, spurIV = T, epsilon = F, n.sims = 100).

The only difference between the uncorrected and correction function calls is the input provided for the argument *whichCor*. Setting the value to zero chooses the uncorrected matrix while setting the value to one chooses the fully corrected matrix.

To produce the data displayed in Table 2, it is necessary to examine the output from both the corrected and uncorrected calculations. To print out the average weights for both the uncorrected and corrected correlation matrices use the command uncorOut$avg.weights and corOut$avg.weights, respectively. Likewise, to produce the average ranks the commands uncorOut$avg.ranks and corOut$avg.ranks are used. To print out the proportion of times that each predictor significantly differed from zero, the commands are uncorOut$diff.zero and corOut$diff.zero. Finally, to produce the data displayed in Table 3, the commands are uncorOut$sig.diffs and corOut$sig.diffs.

[7] If copying and pasting the function calls into R, the symbols such as quotation marks may need to be deleted and rewritten to get the function to execute due to a symbol translation issue. Also, because the function uses Monte Carlo simulations of a bootstrapping process, it will run for several minutes prior to printing output.

Received October 10, 2017
Revision received August 29, 2018
Accepted August 31, 2018 ■