Reporting and Interpreting Research in PSPB: Practices, Principles, and Pragmatics

Deborah A. Kashy, M. Brent Donnellan, Robert A. Ackerman and Daniel W. Russell

Pers Soc Psychol Bull 2009; 35; 1131 originally published online May 19, 2009;
DOI: 10.1177/0146167208331253

Reporting and Interpreting Research in *PSPB:* Practices, Principles, and Pragmatics

Deborah A. Kashy
M. Brent Donnellan
Robert A. Ackerman
Michigan State University

Daniel W. Russell
Iowa State University

This article is designed to provide psychologists who publish articles in Personality and Social Psychology Bulletin *(*PSPB*) with a set of basic issues to consider when reporting their analyses and results. We first assessed the current reporting practices of social and personality psychologists by conducting an analysis of* PSPB *articles published in the first half of 2007. We evaluated the completeness of these reports with respect to the level of detail in both the Method and Results sections. We then used this information to develop recommendations that we hope will enhance the reporting of quantitative research in social and personality psychology. These suggestions emphasize ways to increase transparency in research reports. Transparency facilitates replication and a critical evaluation of research, thereby promoting scientific progress.*

Keywords: *data analysis; social psychology; personality psychology; research methods; research reporting*

Personality and Social Psychology Bulletin (*PSPB*) editors, reviewers, authors, and readers are all stakeholders in the enterprise of psychological research. Although people in these different roles approach articles at different levels, they share a common interest in understanding and evaluating the Method and Results sections of empirical articles. There are ongoing efforts to improve the statistical, methodological, and reporting practices of scientific psychology (e.g., Cumming et al., 2007; Kline, 2004; Wilkinson & Task Force on Statistical Inference, 1999); however, it is unclear how well these efforts have been implemented in substantive journals such as *PSPB*. Our objectives for this article are to describe current reporting practices in *PSPB* and to provide a set of principles and guidelines for authors to consider when reporting research findings.

To understand current reporting practices, we read and evaluated all of the studies reported in articles published in *PSPB* in the first half of 2007. We coded each research report with respect to the adequacy of the Method and Results sections. In this article we describe our findings and provide recommendations that we hope will foster more complete reporting of the kinds of analyses that are typically published in *PSPB*.

The principles and guidelines we put forth are derived from a relatively simple underlying philosophy—the importance of transparency in science. In our minds, transparency is not valuable because it helps detect dishonest or fraudulent reporting. Indeed, we believe that the vast majority of researchers are honest about the science they conduct. Rather, we suggest that transparency is a cardinal virtue because it facilitates a critical reading of research reports, allows fellow scientists to repeat the procedures and analyses in future research,

Authors' Note: We thank Todd Abraham, Rick Hoyle, David Kenny, Norb Kerr, Rich Lucas, Fred Oswald, Radmila Prislin, Rick Robins, Jessica Tracy, Gregory D. Webster, Steve West, and Keith Widaman for their helpful feedback on previous versions of this article. An appendix with example write-ups demonstrating many of our recommendations created by Gregory D. Webster is available at http://www.fun-research.netfirms.com/pspb/appendix.pdf. Correspondence should be addressed to Deborah A. Kashy, Department of Psychology, Michigan State University, East Lansing, MI 48824; e-mail: kashyd@msu.edu.

PSPB, Vol. 35 No. 9, September 2009 1131-1142
DOI: 10.1177/0146167208331253

and provides readers with enough context to thoroughly understand and evaluate the results.

METHOD AND RESULTS REPORTING IN *PSPB:* A 6-MONTH SURVEY

Before we began writing this article, we read the Method and Results sections for every empirical study published in *PSPB* during a 6-month period in 2007. We read 63 articles that reported a total of 155 studies for which the average sample size was 128.14 participants (SD = 96.02) with a range from 16 to 481 participants.[1] We then met and discussed the strengths and weaknesses of how each study was reported. Following these discussions, we developed a coding scheme that allowed us to systematically document the strengths and limitations of the reporting practices in each of the 155 studies. We then reread the *PSPB* issues such that two of us independently coded each study. After coding the studies individually, we met and resolved any discrepancies. For studies that employed more than one data-analytic approach (e.g., both regression and analysis of variance [ANOVA]), we coded the degree to which each analysis included statistical information relevant to that approach. In the sections below, we summarize our findings and provide our recommendations for reporting research in *PSPB*.

METHOD SECTION REPORTING

Participants

PSPB authors are generally clear when laying out the specific methods used in their research. However, descriptions of samples were sparse in many studies. For example, a sizable minority of studies (n = 34; 22%) did not include the number of women and men in the sample. Likewise, the ethnic composition of the sample was not reported in 78% of the studies, and 54% of the studies did not report age information. Thus, it was often the case that studies reported in *PSPB* omitted basic demographic information.

Many theories in social and personality psychology are intended to be universal and therefore do not specify the demographic characteristics that may moderate effects. As a result, researchers may not consider demographic details relevant. However, we think that it is important for every study to provide some basic information concerning sample characteristics because these variables may affect the generalizability of the results. In fact, future meta-analysts may try to use differences in sample characteristics to explain variation in effect sizes across studies. To be sure, the important demographic variables that authors choose to report may depend on the substantive research area, but we suggest routinely reporting such variables as the age, gender, and racial or ethnic composition of samples.

We also noted the potential for discrepancies between the characteristics of the sample as recruited versus the characteristics of the sample as analyzed. For example, researchers sometimes discarded participants who were suspicious of cover stories or who dropped out of longitudinal research. We suggest that it is important to provide descriptive information for the sample on which the critical analyses are based. Moreover, if differences between the initial sample and the analyzed sample are detected, we suggest that these differences be reported.

Procedures and Materials

As we noted, most *PSPB* reports had very clear procedure sections. However, a number of studies contained Materials or Measures sections that lacked important details. For example, approximately 15% of the studies (n = 20) used relatively unknown measures but did not provide information concerning item content. Sample items provide readers with a better sense of the operationalization of the construct being measured. We suggest that researchers who use scales that are not well known include at least two example items for each scale. By well-known scales, we mean measures that roughly 50% of the readers of *PSPB* would recognize. If authors are unsure whether a measure is well known, then we recommend that they include sample items. In addition, approximately 16% of the studies did not include scale endpoints, and 66% did not explicitly state how scale scores were computed (e.g., by averaging over items versus summing over items). This information facilitates an accurate interpretation of a variety of statistics including means and unstandardized regression coefficients.

We also coded whether researchers provided complete information concerning the reliability of their measures. Of the 128 studies that employed scale scores, 86 (67%) reported reliability coefficients for each of their measures, 14 (11%) presented partial reliability information, and 28 (22%) presented virtually no reliability information. Reliability coefficients are important because they index measurement error, and studies that use measures with lower reliabilities are less powerful than are studies that use measures with higher reliabilities.

Recommendations. Based on these findings, we recommend that authors do the following:

- Report demographic characteristics of their samples including gender, age, ethnicity, and other factors that may be relevant to the research context. Inclusion of this information is specified in the American Psychological Association (APA) *Publication Manual* (2001, pp. 18-19).

- Provide enough detail so that readers can understand the manipulations used in the study.
- Describe the scale content, identify scale endpoints, and provide explicit information concerning how the summary scores are computed. We also suggest that authors report descriptive statistics for all variables and note when the distributions substantially violate standard assumptions of normality.
- Report appropriate reliability coefficients for the data used in the study. Reporting alpha coefficients from previous studies or scale manuals is not a substitute for reporting sample-specific values. Reliability is a property that applies to the observed scores in a given study, rather than an immutable property of the instrument.

RESULTS SECTION REPORTING

Data analytic methods have become increasingly complex in journals such as *PSPB* (Sherman, Buddie, Dragan, End, & Finney, 1999) and the *Journal of Personality and Social Psychology* (Reis & Stiller, 1992; West, Newsom, & Fenaughty, 1992). Specifically, although ANOVA remains the most common analytic technique used in social and personality psychology, the use of multiple regression has grown over time. In addition, other multivariate analytic methods such as factor analysis, path analysis, structural equation modeling (SEM), and multilevel modeling (MLM) have gained in popularity.

Consistent with previous findings, our review of articles indicated that ANOVA and multiple regression are the two primary data analytic methods used by *PSPB* authors. Of the 155 coded studies, 80 (52%) used ANOVA (or t tests) and 64 (41%) used multiple regression. Of the less frequently applied data analytic techniques, 7 (5%) studies reported exploratory factor analyses (EFA), 14 (9%) used MLM, and 17 (11%) used SEM or confirmatory factor analysis (CFA). Only 17 studies (11%) used some other analytic technique such as logistic regression, log-linear analysis, or meta-analysis.

In this section, we describe our findings for each of the commonly used methods and then provide suggestions for reporting each method. We reiterate that our recommendations are intended to provide authors with a set of issues to consider when describing their analyses and results. At the end of this section, we consider other methodological issues, including the use of median splits, the extreme groups approach, and methods for testing mediation.

Reporting ANOVA

ANOVA techniques were used in more than half of the studies we coded, and there was considerable variability in how these analyses were described. The most comprehensive articles introduced their analyses with a summary statement that described the experimental design, specified the independent variables along with their levels, and noted which factors were between-subjects and which were within-subjects. These articles also tended to present a thorough reporting of descriptive information including means, standard deviations, and correlations, and they also provided a thorough description of their follow-up analyses (e.g., type of post-hoc test used, whether the global error term was used for follow-up tests).

The majority of articles reporting ANOVA results were not as comprehensive. The most common omission was an absence of relevant standard deviations, as 31% of the studies did not present standard deviations along with each reported mean. In a few cases, no estimates of variability were given anywhere in the article (i.e., no standard deviations in the text or tables, no error bars on graphs, no mean squared errors [MSEs]). The absence of this critical information makes it very difficult for readers to get a sense of the size of reported effects.

When factorial designs were employed, we coded whether researchers provided cell means, standard deviations, and *n*s for their highest-order interactions. These statistics are particularly valuable because they allow readers to compute relevant statistics (e.g., means and standard deviations) for all lower-order effects. We found that 22% of the studies did not present these cell means, 41% did not report the cell standard deviations, and 77% did not include the cell *n*s. In addition, although graphs of means were presented in 36 (45%) of the studies, only 12 graphs included error bars of some sort, and only 5 included labels describing what the error bars signified (e.g., 95% confidence intervals or standard errors). Finally, of the 47 studies that employed ANOVA and reported more than one outcome measure, only 13 studies (28%) provided the correlations among the outcome measures.

Consistent with discussions concerning statistical reform (Wilkinson & Task Force on Statistical Inference, 1999), measures of effect size were presented in close to half of the studies that reported ANOVAs (n = 38; 48%). These effect sizes were primarily η^2 coefficients. One issue is that these coefficients are difficult to interpret in designs that include more than two conditions (Rosenthal & Rosnow, 2008, p. 416; Tabachnick & Fidell, 2001, pp. 52-53). Often, using some form of Cohen's d to characterize planned contrasts between two conditions is an intuitive and effective way to illustrate the magnitude of an effect. A broader issue, and contrary to the specific recommendations by Wilkinson and Task Force on Statistical Inference (1999), is that very few authors explicitly discussed the magnitude of the effects they observed. Indeed, Wilkinson et al. noted that "reporting and interpreting effect sizes in the context of previously reported effects is essential to good research" (p. 599).

Recommendations. Consistent with the instructions given in the APA *Publication Manual* (2001, pp. 20-26), we suggest that authors do the following:

- Explicitly state their design and identify the levels of each factor, the specific values for each level, and the outcome measure(s).
- Provide cell means, standard deviations, and ns in tables for the highest order interactions that are interpreted. Means and standard deviations for lower-order effects can be presented as needed. Means should always be accompanied by their corresponding standard deviations.
- If bar graphs are used to display means, include error bars signifying the 95% confidence interval for the mean (Cumming & Finch, 2005). It is important that the note accompanying the figure identify what the error bars represent (i.e., confidence intervals versus standard errors). Cumming and Finch (2005) provided a useful discussion of confidence intervals and calculating error bars for both within-subjects and between-subjects designs.
- If multiple outcome variables are assessed, include a table of correlations among the outcome variables.[2] These correlations help readers evaluate how much redundancy is present in the dependent measures.
- Report the degrees of freedom for each *F* test and provide the MSEs.
- If the analyses require follow-up tests, describe and justify the strategy used for follow-up or post-hoc tests.

Reporting Regression Analyses

Multiple regression is a very flexible and commonly used tool in social and personality psychology, as more than 40% of the studies in *PSPB* reported regression results. In addition to its traditional uses (e.g., estimating partial relationships and estimating proportions of variance explained), regression is now frequently used to estimate and test interactions involving continuous variables. Our review of *PSPB* articles indicates that researchers understand how interactions can be tested and interpreted using multiple regression, following procedures described by Aiken and West (1991; Cohen, Cohen, West, & Aiken, 2003). Nonetheless, we encountered several problematic reporting practices in our review.

One common issue is that important descriptive data were sometimes omitted. For example, only 36 (56%) of the 64 studies reporting regression analyses presented zero-order correlations among the predictors, and only 34 (53%) presented zero-order correlations between the predictor variables and the criterion variable(s). These correlations are essential because multiple regression is used to obtain estimates of partial associations between predictor and outcome variables. Partial associations can diverge substantially from their corresponding zero-order values if the predictor variables are correlated, and this in turn can present substantial interpretational challenges (Lynam, Hoyle, & Newman, 2006).

For instance, in a regression in which both attachment anxiety and neuroticism are used to predict relationship quality (e.g., Noftle & Shaver, 2006), the regression coefficient for attachment anxiety captures the relation between this aspect of attachment and relationship quality, holding neuroticism constant. In other words, the attachment anxiety slope estimates the association between relationship quality and that part of attachment anxiety that is unrelated to neuroticism. The problem here is that giving a psychological interpretation to *that part of attachment anxiety that is unrelated to neuroticism* can be difficult because the two variables are often highly correlated (e.g., r =.52 for Study 2; Noftle & Shaver, 2006). Therefore, it is important that results from multiple regression are interpreted within the context of the zero-order effects (e.g., Courville & Thompson, 2001). Our concern is that multicollinearity is often considered by researchers as something that is primarily a statistical concern rather than a conceptual issue that affects the interpretation of regression results.

A second point is that many researchers did not specify their coding scheme when using categorical predictor variables. For example, of the 42 studies using dummy coding, 18 (43%) did not report the specific values assigned to the dummy-coded groups. It is essential to know that "men were coded as 1 and women as 0," for example, when interpreting the correlation or regression slope estimating the association between gender and another variable (e.g., self-esteem).

A third concern is that some authors reported only a subset of the results from their final regression model. Of the studies using multiple regression, 38% (n = 24) did not include all of the final model coefficients, and in some cases we were not even able to determine what predictors were included in that model. Most often the unreported coefficients were for variables that the authors deemed to be uninteresting control factors (e.g., gender). Nonetheless, because regression coefficients estimate partial relations, their size and interpretation may change as a function of the other variables included in the model. For example, if one of these unreported control variables is correlated with both the outcome and the focal predictor, then the slope for the focal variable is conditioned on the effects of the other variable. On a related note, a number of researchers included interactions among the predictors but only reported the interaction slopes without providing information concerning the first-order effects (sometimes referred to as main effects, but see Aiken & West, 1991, pp. 38-39).[3]

Last, a more complex and very common problem was the inconsistency with which hierarchical regression analyses were reported. We believe that the most useful application of hierarchical regression is to determine whether the inclusion of a set of predictors results in a statistically significant increase in explained variance over and above an initial set of predictors (i.e., incremental validity). In other words, when researchers want to know whether a set of predictors accounts for a significant amount of additional variance over and above other predictors, they use hierarchical regression and test the R^2_{change} for statistical significance.

In cases in which researchers use hierarchical regression for testing incremental validity, we suggest that the only coefficients that should be interpreted are those from the final model that includes all of the relevant predictors. This point comes directly from one of the key assumptions in regression—that the structural model is valid and, therefore, all relevant variables are included in the analysis. In a hierarchical regression, if the last (or latter) step includes important predictors, the implication is that the model is improperly specified in earlier steps. This means that the regression coefficients and their standard errors from the earlier steps may be biased. In such a case, we do not think there is much point in interpreting coefficients from earlier steps when the model is clearly known to be inadequate.

Of the studies that used regression analyses, 27 (42%) specifically stated that they used a hierarchical regression approach, and 16 of these (59%) presented the results from the final regression model. In 14 of the 27 studies, authors discussed regression coefficients but it was unclear which models generated those coefficients. This was an especially common problem when researchers used hierarchical models that included interactions among the predictors. The typical approach was to include the first-order effects in Step 1 and then include the interactions in Step 2. We often found it impossible to tell whether the interpreted first-order effect coefficients were from the first step of the regression or from the final model that included the interaction(s). In other words, we noted a considerable amount of ambiguity in the reporting of analyses that used hierarchical regression techniques.[4]

Recommendations. More extensive details on conducting, reporting, and interpreting regression analyses are available in West, Aiken, Wu, and Taylor (2007). On the basis of our review of *PSPB* articles, we recommend that authors do the following:

- Provide a clear statement regarding which variables are included in an analysis and why.
- Justify the ordering of variables if hierarchical regression is used.
- Specify how categorical variables are coded.
- Examine data sets for outliers or extreme cases.
- Include a table of means, standard deviations, and correlations for all variables included in regression analyses. This is consistent with the APA *Publication Manual* (2001, p. 23).
- Report regression coefficients (either unstandardized or standardized or preferably both) along with *t* values or standard errors for all predictors included in the model. This is also consistent with the APA *Publication Manual* (pp. 160-163). A related stylistic suggestion: Use lower-case *b*s to indicate unstandardized coefficients because upper-case *B*s can easily be confused with symbols for standardized coefficients (i.e., βs).
- When hierarchical analyses are used, note when coefficients that are discussed are not from the final model. We also suggest that authors provide all of the parameter estimates from the model that includes all relevant predictors (e.g., control variables, all first-order effects involved in higher order interactions, and all interaction terms) regardless of whether they report coefficients from earlier steps.
- When using moderated multiple regression, indicate whether continuous variables were centered or standardized before computing interaction terms.
- When following up statistically significant interactions with tests of simple slopes, indicate what values are used to define high and low (e.g., ±1 standard deviation) levels. When predictor variables are measured in interpretable units, use actual values of conceptual interest (e.g., 20 years of age versus 40 years of age) rather than the ±1 standard deviation option that is commonly used when plotting interactions (West et al., 2007).
- If standardized regression coefficients for interactions in moderated regression are reported, use correct procedures to obtain these coefficients (see Aiken & West, 1991, pp. 43-44). The issue is that the standardized value of a product of two variables is not the same as the product of two standardized variables.
- Avoid using regression equations to predict beyond the range of values in the data set.

Reporting Scale Development and Exploratory Factor Analyses

Articles that are exclusively concerned with scale development are not typically published in *PSPB*. Nonetheless, several of the published studies reported the development and application of new psychological measures. A total of seven studies reported using EFA. One issue we encountered was that authors of five of these studies stated that they conducted EFAs using principal components analysis (PCA). However, there are important conceptual and analytic differences between PCA and EFA (Russell, 2002; Widaman, 2007). The goal of EFA is to identify a smaller set of latent variables that best explain or give rise to the correlations between observed variables. The goal of PCA is data reduction—that is, to reduce a large number of variables to a smaller set of components that account for a large amount of the observed variance. Although these two approaches often yield similar results, differing assumptions underlie each method (Widaman, 2007). EFA embodies the idea that latent variables create variation on observed variables that also contain measurement error; PCA, on the other hand, is not based on an underlying measurement model. PCA is a perfectly appropriate data-reduction technique, but it is not synonymous with EFA. In most scale development contexts, EFA is the more appropriate technique, especially when authors use CFA in a second study to verify the structure of their measure. In short, EFA and CFA share a common measurement philosophy that is not embodied in PCA.

Recommendations. An extensive discussion of EFA was presented by Russell (2002) in *PSPB* and we summarize his main recommendations here. Specifically, we suggest that authors do the following:

- Justify the use of EFA or PCA.
- Use care when applying EFA techniques to a data set that includes two or more distinct groups. If substantial mean differences exist between the groups, these differences can confound the results.
- If EFA is used, report the initial eigenvalues, explain what criteria is used in deciding on the number of factors to extract, and then justify the decision. Exclusive reliance on the K1 rule (i.e., extract all factors with eigenvalues greater than 1.0) is not optimal (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999, p. 278) and is based on faulty logic (Cliff, 1988). Hoyle and Duvall (2004) provided an overview of procedures for determining the number of factors. One alternative to the K1 rule is to use a scree-plot analysis to identify a break between larger and smaller eigenvalues. A more sophisticated approach is to conduct a parallel analysis (see Russell, 2002, p. 1633), and O'Connor (2000) has created easy-to-run SPSS and SAS macros for this purpose. Finally, Hoyle and Duvall (2004) presented a tutorial on using maximum likelihood strategies for determining the number of factors.
- Justify the rotation procedure used. Orthogonal rotations (e.g., varimax) require factors to be independent, whereas oblique rotations (e.g., promax or oblimin) allow factors to be correlated. We recommend that researchers use oblique rotations because this method will uncover independent factors if they exist. That is, if factors are truly uncorrelated, then this will be apparent in the matrix that displays the correlations between latent variables. These correlations between the factors also should be reported.
- When oblique rotations are used, refer specifically to pattern and structure coefficients as opposed to the generic term *factor loading*.[5]
- Consider reporting all coefficients and not just coefficients larger than some arbitrary cutoff (e.g., 0.30).
- Given the complexities of factor analysis, it is a good idea to extract several solutions with different numbers of factors and evaluate each for interpretability and consistency with theoretical expectations. Researchers do not need to present all such analyses, but they should engage in this process, report taking this approach, and justify their preferred solution.

SEM and CFA

SEM is a powerful tool for analyzing data, and one that was used with some regularity in the studies we reviewed ($n = 17$; 11%; note that CFA is a special application of SEM and so it is considered here). One critical element we looked for was whether authors presented the variance-covariance matrix (or the correlation matrix with standard deviations) that was used to generate their results, as suggested by the APA *Publication Manual* (2001, p. 23). This information is important because researchers can use it to specify and test alternative or competing models with the same dataset. Some authors did not report a table of correlations ($n = 6$; 35%), whereas others reported correlations but did not report standard deviations ($n = 8$; 47%). In addition, some authors used item parcels in their models ($n = 5$; 29%) but did not include those item parcels in their correlation matrix.[6] All in all, we found only two studies that actually reported the exact matrix required for re-estimating the published model. Moreover, using the covariance matrices provided, we were able to replicate closely the results from only one of these two published models.[7]

Recommendations. McDonald and Ho (2002) have presented recommendations regarding the reporting of structural equation models, and our recommendations are generally consistent with theirs. Specifically, we suggest that authors do the following:

- Describe and justify the a priori model, as SEM is best used to evaluate the degree to which a theoretically derived model fits observed data. In many cases, a figure depicting the model helps structure this discussion.
- Report means, standard deviations, and correlations for all variables used in the analyses. If observed variables deviate substantially from the normal distribution, report skewness and kurtosis statistics with the descriptive data.
- Report item parcels as variables in the correlation/covariance matrix included in the article. The use of item parceling should be justified in the text.
- Most readers will assume that maximum likelihood techniques were used for estimation. However, other

techniques are more appropriate if multivariate normality is violated. If maximum likelihood is not used, we suggest that authors report which estimation method was employed.

- Report and interpret several omnibus fit statistics. At a minimum, report the chi-square and degrees of freedom, along with the root mean square error of approximation and imation and comparative fit index CFI or the Tucker-Lewis Index. Hu and Bentler (1999) suggested that a root mean square error of approximation less than .06, a CFI of .95 or greater, and a Tucker-Lewis Index of .95 or greater indicates that the model provides a reasonable fit to the data. Several studies we reviewed suggested thresholds for the CFI that were lower (i.e., .90). To be sure, model fit continues to be a contentious issue (e.g., see Barrett, 2007, along with commentaries; Marsh, Hau, & Wen, 2004), and we anticipate that work in this area will continue.
- One reasonable strategy for evaluating sources of model misfit is to evaluate the adequacy of the underlying measurement model before estimating the theoretical model (see Anderson & Gerbing, 1988). In other words, authors may first conduct a CFA on all measures simultaneously before moving to the second step of testing their preferred structural model.
- Report results for the structural equation model in either a table or a figure. Coefficients and standard errors (or z values) should be reported for all estimated paths—even paths that are not statistically significant.
- Describe and evaluate alternative models (MacCallum, Wegener, Unchino, & Fabrigar, 1993), both those models that are theoretically interesting and those models that are statistically equivalent. The fit of conceptually interesting alternative models is useful for judging the fit of the preferred model.
- Report and justify any post-hoc modifications to models that were made to achieve satisfactory model fit. MacCallum (1986) noted that modifications designed to improve the fit of a particular model to a given data set often do not replicate. Researchers (and readers) should, therefore, view post-hoc modifications with a fair bit of skepticism.

Reporting MLM Results

MLM is one of the newest additions to the toolbox of social and personality psychologists, and this approach was used in 14 of the 155 studies (9%) we reviewed. Many of the principles described for reporting the results of ordinary regression analyses also apply to MLM. As we noted for regression, it is very important that researchers provide sufficient descriptive information concerning their data. However, such descriptive data can be considerably more complex given the nested nature of the observations.

Recommendations. We recommend that authors do the following (with the caveat that standards for MLM reporting are evolving):

- Clearly describe the research design and data collection strategy so that readers can understand the nested structure of the data (e.g., individuals within groups or time points within individuals).
- Describe how categorical variables are coded.
- Indicate how variables are centered (i.e., grand mean centering versus group mean centering) and justify this decision. Several sources provide discussions of the relative merits of each approach to centering (e.g., Enders & Tofighi, 2007; Kreft, de Leeuw, & Aiken, 1995).
- Provide basic descriptive data including means, standard deviations, and correlations among the variables.
- Report sample sizes at each level of analysis as well as intraclass correlations that estimate the degree of nonindependence of the nested observations. Intraclass correlations can be estimated using models with no predictors, and partial intraclass correlations can be estimated using models that include predictors.
- Specify which variables are included in each estimated model and report all estimated effects. We also suggest that authors identify which effects in the model are treated as fixed and which are treated as random, and report which estimation technique was used (e.g., restricted maximum likelihood (REML) versus maximum likelihood). For models involving variables assessed over time, we suggest that researchers indicate the error structure specification (e.g., AR1, diagonal).
- Appreciate the distinction between MLM as a statistical technique and the HLM statistical program (Raudenbush, Bryk, Cheong, & Congdon, 2004). An overreliance on HLM (the program) notation makes it difficult for non-HLM users to interpret results. Of the 14 studies we reviewed, 5 (36%) presented all of their results using notation associated with the HLM computer program without an explanation of what each symbol represented.
- Briefly walk readers through the interpretation of model estimates because MLM is a relatively new technique.

Categorizing Continuous Predictors and the Extreme Groups Approach

In the course of our review of *PSPB* articles, we noted a few other methodological practices that deserve comment. The first has to do with categorizing continuous variables (e.g., using median splits) for use in ANOVA. This practice seems to be on the decline as most authors are now comfortable with the multiple regression approach to estimating interactions with continuous predictor variables. However, MacCallum, Zhang, Preacher, and Rucker (2002) found that close to 16% of articles published in *Journal of Personality and Social Psychology* between January 1998 and December 2000 used dichotomized continuous variables. Our survey of *PSPB* articles showed that this practice may be continuing to decline, as median split (or tertile split) methods were used in just 9 of the 155 studies (6%). MacCallum et al. noted that there are a few occasions in which dichotomization is a legitimate approach; however, none of the 9 *PSPB* studies that used this method met MacCallum et al.'s criteria.

As a number of methodologists (e.g., MacCallum et al., 2002; Maxwell & Delaney, 1993) have discussed, the practice of dichotomizing variables has several negative consequences. First, researchers lose information about individual differences and run the risk of misclassifying participants who score close to the median or cut point. Second, researchers may find spurious interaction effects when two correlated continuous predictors are both dichotomized (Maxwell & Delaney, 1993). In addition, the likelihood of spurious effects increases as the correlation between the two predictors increases (Maxwell & Delaney, 1993). Third, this approach typically reduces statistical power and biases effect size estimates. Finally, because these procedures are typically sample specific (i.e., median values are dependent on the particular sample under consideration), aggregating results across studies that use this practice is difficult. All in all, we believe that dichotomizing continuous variables into discrete categories is usually an inadvisable research practice.

A related methodological issue is the use of extreme groups, and we found three studies that used this approach. Preacher, Rucker, MacCallum, and Nicewander (2005) provided an important discussion on the costs and benefits of the extreme groups approach. They concluded,

> Given the risks associated with [the extreme groups approach], we suggest that any implementation of its use should be accompanied by careful consideration and clear justifications. . . . We urge reviewers, editors, and consumers to consider the appropriateness of instances of [the extreme groups approach] encountered in the literature. (p. 190)

None of the three studies that used the extreme groups approach justified their usage of this practice, and all violated Preacher et al.'s recommendation to refrain from dichotomizing the data after participants were selected because of their extreme scores. Specifically, Preacher et al. recommended that researchers use regression techniques, treating the selection variable as a continuous predictor. Based on these considerations, we think that it is best for social and personality researchers to avoid using the extreme-groups approach whenever possible. In cases in which researchers are compelled to use this method, it is important that they justify their decision in the text and alert readers to the potential limitations of this approach.

Conducting and Reporting Mediation Analyses

Social and personality psychologists are very interested in identifying the processes that explain why certain independent variables are connected to certain outcome variables (e.g., Spencer, Zanna, & Fong, 2005). Close to half of the articles in our survey (26 of 63) reported at least one study ($n = 38$; 25%) in which tests of mediating processes were conducted. In terms of statistical tests, by far the most common approach involved computing the Sobel test ($n = 29$; 76%), and many of the studies that tested mediation referred only to the classic Baron and Kenny (1986) framework ($n = 15$; 39%). Baron and Kenny is a landmark article but recent treatments describe more powerful approaches to evaluating mediation (e.g., MacKinnon, 2008; MacKinnon, Fairchild, & Fritz, 2007; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Shrout & Bolger, 2002; for discussions of moderated mediation see Edwards & Lambert, 2007; Judd & Kenny, 1981; Muller, Judd, & Yzerbyt, 2005).

One issue is that the Sobel test lacks statistical power, especially with small to moderate sample sizes, which are often defined as $n < 400$ (Dearing & Hamilton, 2006). Indeed, it is interesting to note that only 4 of the 155 studies in our review had samples where $n > 400$. One solution that has been recommended (MacKinnon, 2008; Shrout & Bolger, 2002) is for researchers to use a bootstrap method for testing the statistical significance of mediators. Preacher and Hayes (2004, 2007) described methods and provided macros for implementing these procedures using SPSS and SAS. A number of SEM programs (e.g., AMOS, EQS, LISREL, Mplus) can also be used to conduct bootstrap procedures. Given the sample sizes of the typical studies published in *PSPB*, we suggest that these more powerful approaches to testing mediation be more widely used.

Beyond the statistical issues, we would like to raise a broader issue about evaluating hypotheses concerning mediation. Spencer et al. (2005) made the compelling argument that researchers sometimes evaluate process models where the mediator is conceptually quite similar to the outcome variable. We observed some tests of mediation where the mediator seemed like an alternative measure of the dependent variable. In these cases, evidence of statistical mediation does not represent real added value to a research report. The bottom line is that compelling theoretical models and careful research design, rather than statistical procedures, should drive the evaluation of mediation (Spencer et al., 2005).

More General Considerations: Significance Testing and Replication

We have purposely avoided some of the issues surrounding null hypothesis significance testing (e.g., Fraley & Marks, 2007). Our concern is that these discussions often create reactance and cause readers to stop reading. Our primary goals for this article are pragmatic—we hope to enhance reporting practices in *PSPB*—and thus we made a strategic choice to keep our comments about significance testing brief and place them in the closing sections. Nonetheless, we think there are important philosophical issues at stake.

Misunderstandings of null hypothesis significance testing persist in scientific psychology (Kline, 2004). The *p* value obtained from inferential statistics has a very precise interpretation: It represents the probability of observing a particular test value (or one more extreme) if the null hypothesis was in fact true. The most common misconception is that the *p* value provides the probability that rejecting the null hypothesis is the wrong decision (Kline, 2004). Another misconception is that the *p* value provides information about the probability of replicating a finding. To be clear, the *p* value does neither. We hope that all social and personality psychologists will continue to reflect on the meaning and value of significance tests and take a principled stand on this issue. Practically, we would like to see indications of statistical significance (e.g., ****$p < .0001$) used more sparingly and confidence intervals, effect sizes, and indications of statistical power reported more widely. Likewise, we hope that authors draw distinctions between findings that are theoretically and practically important and findings that are "highly significant" in the statistical sense.

The emphasis on the importance of $p < .05$ was evident in some of the *PSPB* articles we read. One of the more troubling issues was the seemingly arbitrary switching between two-tailed and one-tailed tests in several of the articles. Although researchers often argued that the one-tailed approach was appropriate given their directional hypothesis (but see Harris, 1997, for a counterargument), we found it interesting that one-tailed tests were used primarily when the one-tailed *p* value was larger than .025; this is not a principled stand on the significance testing issue.

All in all, we hope that researchers strive to find replicable effects, the building blocks of a cumulative science. Indeed, Steiger (1990) noted, "*An ounce of replication is worth a ton of inferential statistics*" (p. 176). As we have emphasized throughout, clear and transparent reporting is vital to this aim. Providing enough details in the Method and Results sections allows other researchers to make meaningful attempts to replicate the findings. A useful heuristic is for authors to consider whether the draft of their paper includes enough information so that another researcher could collect similar data and replicate their statistical analyses.

Space Constraints and Writing an Interesting Article

In discussing these principles and guidelines with our colleagues, we discovered that a common objection was that our recommendations would result in tediously detailed articles that were simply uninteresting. That is absolutely not our intention. To be clear, we are not suggesting that authors produce manuscripts with 25-page Method and Results sections.

We realize that not every statistical analysis needs to be thoroughly reported. For example, researchers often follow up their main regression analyses with analyses that include control variables that might qualify their results. In many cases, a brief summary of the findings of such analyses will suffice. Thus, we are not arguing that every coefficient from every model needs to be included in the text or a table.

Our point is that all of the results from the primary analyses should be clearly reported and that these results should be accompanied with a thorough reporting of descriptive statistics. These details are a large part of what makes the research valuable to the science of social and personality psychology. All in all, we think these details provide the essential foundation for the larger story being told by the research report.

A practical concern is that word limits may constrain the amount of detail authors believe they can include in a report. We do not think that our recommendations will add considerable length to manuscripts. Moreover, we argue that if space is an issue, then authors should consider trimming the Method and Results sections as the option of last resort. We suggest that before authors cut methodological details, they should consider trimming overlapping references (e.g., Adair & Vohra, 2003) or omitting unnecessary details (e.g., participants were greeted, SPSS was used for analyses).

Occasionally, authors may find that the level of detail we are recommending represents a hardship with respect to word limits in *PSPB*. For example, if the study involves a large number of measures, correlation tables that include all variables may be unwieldy. Likewise, detailed descriptions of supplemental analyses may be deemed of relatively little value. In these cases, we suggest that researchers prepare such materials as if they were going to be included in the manuscript. Authors can then report abbreviated versions of this material in the article (e.g., correlation tables involving focal variables) and note that complete versions are available upon request.

The Potential Value of Data Archives

One long-term solution to the space problem in printed journals is to establish Web sites for supplementary materials, as is now current practice for the journal *Science*. We would like to see social and personality psychologists give serious consideration to creating data archives. Such archives might contain deidentified raw data, codebooks, syntax for complicated analyses, summary tables for supplemental analyses, detailed descriptions of experimental procedures and manipulations, questionnaires, and other ancillary materials. This information would facilitate replication and would contribute to a cumulative science. Freese (2007) addressed many of the counterarguments against data archiving and provided a compelling discussion of their merits.

In fact, Wicherts, Borsboom, Kats, and Molenaar (2006) proposed that all authors submit an ASCII file with deidentified data and a codebook for each study when manuscripts are accepted for publication. They proposed that journals publish this information on the Internet as an electronic appendix. We hope that *PSPB* will consider adopting this policy in the future. That said, no such resource is currently available, and so we suggest that individual researchers take proactive steps toward making their research archivable when they prepare to submit their manuscripts for review. At a minimum, this means constructing clean copies of data sets used in final analyses that are accompanied by appropriate documentation (e.g., variable labels, value labels, details of data collection and coding). This material will facilitate data sharing and may even help authors when it becomes necessary to revise papers or revisit analyses.

CONCLUDING THOUGHTS AND SUMMARY RECOMMENDATIONS

Our primary goal for this article was to enhance the reporting of empirical articles submitted to and published in *PSPB*. Our recommendations were based on the value we place on the principle of transparency in the scientific enterprise. At the most basic level, all of our recommendations to authors can be distilled into the following three overarching principles:

- Be clear and precise so that readers can critically evaluate the work and replicate procedures and data analyses.
- Routinely report descriptive data such as means, standard deviations, correlations, and reliability coefficients because this information provides the context for interpreting more complex analyses.
- Write Method and Results sections in a thorough and complete manner so that future researchers can extend the work or include the findings in their meta-analyses.

We have restricted our focus to basic principles as they relate to current practices, and so there are many important methodological and data analytic issues we did not cover. For example, our discussion of measurement emphasized clarity of measure description and measure reliability; we did not discuss modern psychometric techniques such as item response theory (e.g., Embretson & Reise, 2000) or issues of measurement invariance when making group comparisons (e.g., Vandenberg & Lance, 2000). Likewise, we did not discuss techniques for dealing with missing data, which have improved substantially in recent years (McKnight, McKnight, Sidani, & Figueredo, 2007; Schafer & Graham, 2002; Sinharay, Stern, & Russell, 2001; see also West et al., 2007). All in all, we made a conscious decision to avoid proselytizing for new methodological techniques; instead, we focused on ANOVA and multiple regression—the two most commonly used approaches in *PSPB*.

We would like to emphasize, however, that research methods evolve and that investigators need to keep up with these advances. For example, the Sobel test was initially regarded as the preferred method for testing indirect effects, whereas more recent treatments of mediation emphasize bootstrapping approaches (Shrout & Bolger, 2002). We believe that researchers have an obligation to maintain and update their methodological skills. To that end, we have provided citations to a variety of accessible methodological articles that describe in considerably greater detail the issues we have raised. In addition, readers should be aware that our recommendations may be superseded by future developments. *Psychological Methods* is a useful resource for keeping up with developments in methodology.

In sum, we hope that our discussion facilitates transparent reporting, fair-minded criticism, and scientific progress in social and personality psychology. In our view, the Method and Results sections contain the heart of the science that is reported in an empirical article, and so it is vital that these critical sections are thorough and informative. There is a tradition in social and personality psychology that articles need to tell a good story. We agree—and in our view the data and the results are a crucial part of that story.

NOTES

1. The 6-month period actually included a total of 65 articles. One article was excluded because it was not an empirical study, and a second article reported two studies that were so different from the standard *PSPB* empirical publication that they defied coding. Finally, one study from one article was excluded because it presented meta-analytic data. In addition, the sample size estimates reported here represent liberal estimates (i.e., the total number of participants before any were removed due to suspicion, missing data, attrition, etc.).

2. If there are large mean differences between treatment groups, these correlations can be computed as partial correlations, controlling for the effects of the independent variables.

3. The first-order effects are required to compute the simple slopes for a moderated regression, and so omission of this information can be problematic.

4. Cohen, Cohen, West, and Aiken (2003) also illustrated how hierarchical regression can be used to estimate the size of the causal effect associated with independent variables that are specified in a causal hierarchy (see pp. 158-160). In such an application, authors should justify the underlying causal model and explain why the size of coefficients at particular steps is informative. However, none of the

articles we coded used hierarchical regression for this purpose. At any rate, the point is that authors need to be clear about which step generated the coefficients that are discussed in the text or reported in the tables.

5. When orthogonal rotations are used, these two types of coefficients are the same.

6. Item parcels are used when researchers want to create a latent variable from a single measure of the construct in question. Parceling involves breaking multiple-item scales into two or more subsets of items, which are then used to identify a latent variable (see Little, Cunningham, Shahar, & Widaman, 2002, for a review of the issues in parceling).

7. We e-mailed the corresponding author and were not able to obtain the raw data because the data set was proprietary. The corresponding author could not explain the discrepancies between our replication and their published results. We suggest that authors who are unable to provide data upon request for legitimate reanalyses should state such restrictions in their Author's Note.

REFERENCES

Adair, J. G., & Vohra, N. (2003). The explosion of knowledge, references, and citations: Psychology's unique response to a crisis. *American Psychologist*, *58*, 15-23.

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

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

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*, 411-423.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

Barrett, P. (2007). Structural equation modeling: Adjudging model fit. *Personality and Individual Differences*, *42*, 815-824.

Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin*, *103*, 276-279.

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

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement*, *61*, 229-248.

Cumming, G., Fidler, F., Leonard, M., Kalinowski, P., Christiansen, A., Kleinig, A., et al. (2007). Statistical reform in psychology: Is anything changing? *Psychological Science*, *18*, 230-232.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, *60*, 170-180.

Dearing, E., & Hamilton, L. C. (2006). Contemporary advances and classic advice for analyzing mediating and moderating variables. *Monographs of the Society for Research in Child Development*, *71*(3), 88-104.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytic framework using moderated path analysis. *Psychological Methods*, *12*, 1-22.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*, 121-138.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272-299.

Fraley, R. C., & Marks, M. J. (2007). The null hypothesis significance testing debate and implications for personality research. In R. Robins, R. C Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality* (pp. 149-169). New York: Guilford.

Freese, J. (2007). Replication standards for quantitative social science: Why not sociology? *Sociological Methods & Research*, *36*, 153-172.

Harris, R. J. (1997). Significance tests have their place. *Psychological Science*, *8*, 8-11.

Hoyle, R. H., & Duvall, J. L. (2004). Determining the number of factors in exploratory and confirmatory factor analysis. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 301-315). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, *5*, 602-619.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multilevel Behavioral Research*, *30*, 1-21.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*, 151-173.

Lynam, D. R., Hoyle, R. H., & Newman, J. P. (2006). The perils of partialling: Cautionary tales from aggression and psychopathy. *Assessment*, *13*, 328-341.

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*, 107-120.

MacCallum, R. C., Wegener, D. T., Unchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*, 185-199.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19-40.

MacKinnon, D. P. (2008). *An introduction to statistical mediation analysis*. Mahwah, NJ: Erlbaum.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. (2007). Mediation analysis. *Annual Review of Psychology*, *58*, 593-614.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*, 83-104.

Marsh, H. W., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cut-off values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320-341.

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

McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*, 64-82.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York: Guilford.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, *89*, 852-863.

Noftle, E. E., & Shaver, P. R. (2006). Attachment dimensions and the Big Five personality traits: Associations and comparative ability to predict relationship quality. *Journal of Research in Personality*, *40*, 179–208.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*, 396-402.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*, 717-731.

Preacher, K. J., & Hayes, A. F. (2007). Contemporary approaches to assessing mediation in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research*. Thousand Oaks, CA: Sage.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods*, *10*, 178-192.

Raudenbush, S. W., Bryk, A., Cheong, Y. F., & Congdon, R. (2004). *HLM6: Hierarchical linear and nonlinear modeling*. Chicago: Scientific Software International.

Reis, H. T. & Stiller, J. (1992). Publication trends in *JPSP*: A three-decade review. *Personality and Social Psychology Bulletin*, *18*, 465-472.

Rosenthal, R., & Rosnow, R. L. (2008). *Essentials of behavioral research: Methods and data analysis* (3rd ed.). New York: McGraw-Hill.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in *Personality and Social Psychology Bulletin*. *Personality and Social Psychology Bulletin*, *28*, 1629-1646.

Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147-177.

Sherman, R. C., Buddie, A. M., Dragan, K. L., End, C. M., & Finney, L. J. (1999). Twenty years of *PSPB*: Trends in content, design & analysis. *Personality and Social Psychology Bulletin*, *25*, 177-187.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*, 422-445.

Sinharay, S., Stern, H. S., & Russell, D. W. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, *6*, 317-329.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective in examining psychological process than mediational analyses. *Journal of Personality and Social Psychology*, *89*, 845-851.

Steiger, J. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173-180.

Tabachnick, B. G., & Fidell, L. S. (2001). *Computer-assisted research design and analysis*. Boston: Allyn & Bacon.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4-69.

West, S. G., Aiken, L. S., Wu, W., & Taylor, A. B. (2007). Multiple regression: Applications of the basics and beyond in personality research. In R. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality* (pp. 149-169). New York: Guilford.

West, S. G., Newsom, J. T., & Fenaughty, A. M. (1992). Publication trends in *JPSP*: Stability and change in the topics, methods, and theories across two decades. *Personality and Social Psychology Bulletin*, *18*, 473-484.

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, *61*, 726-728.

Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 177-203). Mahwah, NJ: Erlbaum.

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

Received December 14, 2007
Revision accepted November 26, 2008