Nonlinear Transformations in Organizational Research: Possible Problems and Potential Solutions

Article *in* Organizational Research Methods · October 2019

DOI: 10.1177/1094428118775205

3 authors:

Thomas E. Becker
University of South Florida

Melissa Robertson
Purdue University West Lafayette

Robert J Vandenberg
University of Georgia

Article

Organizational Research Methods
2019, Vol. 22(4) 831-866

DOI: 10.1177/1094428118775205
journals.sagepub.com/home/orm
SAGE

Nonlinear Transformations in Organizational Research: Possible Problems and Potential Solutions

Thomas E. Becker[1], Melissa M. Robertson[2], and Robert J. Vandenberg[3]

Abstract

We examined the use of nonlinear transformation of variables in a random sample of 323 articles published in six top journals during 2012-2017. Coding categories included the number of transformed variables, the type of transformation, the kinds of variables transformed, reasons provided for transforming variables, how transformed results were reported, and pre- and posttransformation analysis of variables. Common problems include insufficient justification for transforming variables, overreliance on log transformations, failure to report important information on the effects of transformation, and incomplete reporting and discussion of transformed results. Perhaps most importantly, there was frequent misalignment between statements of hypotheses, typically stated in terms of nontransformed variables, and the transformed data used to test them. We discuss the implications of these problems for science and practice, offer recommendations for addressing the issues, and provide illustrations of how to implement the recommendations.

Keywords

nonlinear transformations, log transformations, transforming variables, quantitative methods

It is common for organizational researchers to use nonlinear transformations (NLTs) as part of their analysis strategy. A NLT involves inconstant changes in the units or scale of a variable, as when a researcher takes the natural logarithm of scores on a positively skewed variable with the intent of creating a new distribution that is more normal and, thus, more appropriate for use in statistical testing (Cohen, Cohen, West, & Aiken, 2003). In addition to attempting to normalize a variable's

[1]University of South Florida Sarasota-Manatee, Sarasota, FL, USA
[2]Department of Psychology, University of Georgia, Athens, GA, USA
[3]Department of Management, University of Georgia, Athens, GA, USA

Corresponding Author:
Thomas E. Becker, University of South Florida Sarasota-Manatee, College of Business, Sarasota, FL 34243, USA.
Email: teb1@sar.usf.edu

distribution, other reasons for NLTs include reducing heteroscedasticity, linearizing relationships among variables by changing nonlinear relationships into linear ones, turning multiplicative relationships into additive ones, and in other ways promoting analytical convenience (Box & Cox, 1964; Emerson, 1983; Tukey, 1957; Yeo & Johnson, 2000). To accomplish these ends, researchers may consider a variety of NLTs, including log transformations, square roots, reciprocal roots, exponential powers, and dropping or replacing outliers (de Vaus, 2002; Tukey, 1977; Velleman & Hoaglin, 1981).

Although NLTs may at times be appropriate, we argue that the practice is fraught with three fundamental problems. First, NLTs can impugn the construct validity of scores and result in improper tests of hypotheses. Part of this problem is that although it is essential to consider the conceptual basis of NLTs (Cohen et al., 2003), NLTs are often applied with little concern for substantive theoretical implications (Russell & Dean, 2000). The disregard for theory raises concerns for construct validity and hypothesis testing because performing a NLT creates a new variable, the scores on which may not operationalize the latent variable intended by the researcher (Lo & Andrews, 2015). For example, a researcher might hypothesize that firm size predicts organizational sales. If the distribution of sales is platykurtic (flat), the researcher may attempt to normalize it by taking the square root of the original values. Consequently, the variable included in analyses is not sales but, rather, the square root of sales. Thus, the researcher cannot legitimately claim to be testing the original hypothesis. Gullikson (2006) emphasized this point in discussing how taking the natural log of an independent variable implies a relationship of diminishing returns between the independent and dependent variables. Such a relationship is not implied with the original distribution of raw scores, and unless the researcher has theoretical grounds for positing diminishing returns, the results may not be meaningful because the transformed scores do not represent the original variable. In short, unless hypotheses are stated in terms of the transformed variables, analyses using the scores do not accurately test the hypotheses.

Second, NLTs may result in new problems with data that were not present with the original scores. For example, in their Monte Carlo simulation, Russell and Dean (2000) demonstrated that log transformations can severely decrease estimates of true moderator effects, substantially increasing Type II error (incorrectly concluding that an effect is nonsignificant). In addition, the exclusion or replacement of outliers—a form of NLT in that it can change the scale of measurement and shapes of distributions—is often questionable because there are substantial inconsistencies and lack of transparency in how outliers are defined, identified, and addressed (Aguinis, Gottfredson, & Joo, 2013; Aguinis & Joo, 2015). Furthermore, there are a variety of nonnormal distributions varying in shape, mass (the likelihood of a discrete variable taking on a given discrete value), and density (the likelihood of a continuous variable taking on a given value or range of values; Joo, Aguinis, & Bradley, 2017). By failing to consider the characteristics of the original distribution, researchers' normalization attempts may fail, or even increase nonnormality. These unintended effects occur because NLTs can make a distribution more normal in one sense (e.g., shape) while making it less normal in another (e.g., mass or density) (Cohen et al., 2003). This problem exists even for a given characteristic of distributions. For instance, regarding shape, a NLT could reduce positive skew while simultaneously making the distribution more leptokurtic (peaked). All of these issues hinder the interpretation of results by introducing substantial uncertainty into the nature and meaning of transformed scores.

The third problem is that, even when results are properly understood, findings using NLTs may not generalize to other samples or situations. This potential lack of external validity occurs because NLTs can create distributions that do not exist in the real world. For example, an original distribution of job performance might be negatively skewed while the $[\text{performance}]^2$ distribution is more "normal." However, in real-world settings such as performance appraisals, performance would not be measured or interpreted this way (i.e., squared) and probably should not be. For instance, the performance of two employees could be similar with, for instance, one employee receiving a 4 and the other receiving a 5 on a 5-point scale. After transformation, it may appear that the divergence is far greater than it is (e.g., a 16 versus 25 on a 25-point scale), thus exaggerating the difference in performance of the two employees. In addition, NLTs can create relationships where none exist, and mask relationships that do exist (Cohen et al., 2003). For example, in validating a selection test, a researcher might find that there is no relationship between test scores and performance. After transformation, there could be a significant correlation between the two variables. However, it is not clear that the transformed relationship is a more accurate estimate of the test's predictive validity. Because the distributions of original and transformed scores may be very different and can suggest conflicting inferences, conclusions from research involving NLTs must be cautiously stated.

In addition to concerns for construct validity, interpretation, and generalizability, there are other difficulties with using NLTs. One is that NLTs can mask problems with the data or obscure interesting features of distributions. For example, if performance ratings are contaminated by halo error, taking the natural log of performance scores could make the halo effect less visible by appearing to reduce correlations among performance dimensions. Nevertheless, the error in ratings would be just as present. Also, nonlinear distributions may reflect important substantive information about a construct in general or in a particular population. For instance, if a researcher finds that turnover was normally distributed for men but highly skewed for women, the difference could be caused by personal, social, or organizational factors worth studying. A final problem is that NLTs may promote questionable research practices because unethical researchers could experiment with different transformations until they find one that produces statistically significant results in the "expected" direction. These findings could be due to chance, but this would go undetected by readers and reviewers if the researchers did not report family-wise error, the different transformations that were tried, or the reasons for trying them.

In sum, the intention to use NLTs to produce more appropriate statistical estimates is laudable, and NLTs may under certain circumstances serve the intended purpose. However, our reading of top journals in management led us to believe there are problems in how NLTs are selected, reported, interpreted, and discussed. Reducing these problems would benefit the field by encouraging greater confidence in the construct validity of measures, the interpretation of results, and implications for future research and practice. Hence, our goals are to study the use of NLTs in recent organizational research, to assess how effectively issues of transformation have been addressed, and to offer recommendations for improving the application of NLTs.

Method

Selection of Articles

The data for this study were collected from a stratified random sample of research articles published in the *Academy of Management Journal* (*AMJ*), *Administrative Science Quarterly* (*ASQ*), *Journal of Applied Psychology* (*JAP*), *Journal of Management* (*JOM*), *Strategic Management Journal* (*SMJ*), and *Personnel Psychology* (*Personnel Psych*) during the period 2012 through the spring of 2017. Articles were chosen using a random numbers generator (https://www.random.org) with the restriction that, for 2012-2016, 10 articles be selected from each of the six journals. For the spring of 2017, four articles were chosen from each journal, bringing the total to 54 articles per journal and 324 total. We chose the period of study to include recent literature within the analysis, with a sample size large enough for the limited quantitative analyses we planned, but not so large as to make the qualitative analysis infeasible.

Table 1. Interrater Agreement Regarding Nonlinear Transformations in 30 Articles.

Coding Issue	T
1. Did the article include one or more nonlinear transformations?	.96
2. How many nonlinear transformations were reported in the article?	.78
3. What kind of nonlinear transformations were used?	.82
4. What was the basis (reason) provided for using nonlinear transformations?	.79
5. Were citations provided to justify the use of nonlinear transformations?	1.00
6. Did the researchers report prescreening the data for violations of statistical assumptions?	1.00
7. After performing nonlinear transformation, did the researchers report evaluating data for violations of statistical assumptions?	1.00
8. Were nonlinearly transformed variables used to test hypotheses about the untransformed variables?	.71
9. Was there confusion in the results section regarding labeling or description of, or reference to, nonlinearly transformed variables?	.82
10. Was there confusion in the discussion section regarding identification or discussion of nonlinearly transformed variables?	.96

Note: T is the Tinsley and Weiss (1975) index of rater agreement. Questions 1, 5, 6, 7, 8, 9, and 10 were dichotomous (answered yes or no), Question 2 was continuous (number of transformations), and Questions 3 and 4 were categorical (providing several types of transformations and reasons for performing transformations).

Analyzing and Coding Articles

Primary coding. The major aim of coding was to identify categories and subcategories representing the proper and potentially improper use of NLTs. The categories and subcategories were developed based on recommendations from prior literature, particularly Cohen et al.'s (2003) discussion of NLTs (see chap. 6, pp. 221-249). In addition, we read several empirical articles published prior to 2012 in the same six journals identified above to obtain a sense of the practices undertaken by researchers and the kinds of information that should be coded. Based on the prior recommendations and our observations of current practices, we created four categories pertinent to NLTs: justification, reporting, alignment, and discussion, each having several subcategories. The categories and subcategories are included in Table 2. A copy of the coding specifications is included in Online Appendix 1 and a spreadsheet containing coding for a sample of articles is provided in Online Appendix 2.

Other coded variables. For descriptive purposes, we collected information on several other variables. Specifically, we coded data source (whether the data came from archival sources, surveys, interviews, direct observation, or other sources), level of analysis (individual, dyad, team, organization, or other), number and level of measurement of independent, dependent, control, and transformed variables, types of analyses, type of NLT, and names of the transformed variables.

Consistency check. Throughout the coding process, the authors met periodically to discuss coding progress and issues that emerged during coding. Where disagreement existed, we discussed the nature and likely causes and developed decision rules to avoid similar disagreements in future coding. Once coding was completed, the authors independently coded 30 randomly selected articles as a check on the consistency of coding. We were more concerned with the level of agreement among raters rather than the general correlation across ratings. Therefore, we assessed interrater agreement with the Tinsley and Weiss (1975) index, the formula for which is $T = (N_1 - NP) / (N - NP)$, where N_1 is the number of agreements between raters, N is the number of cases rated, and P is the probability that raters would agree by chance. Table 1 reports the questions

Table 2. Percentage of Problematic NLT Practices in Each Subcategory.

Category/Subcategory	Percentage
Justification for using NLTs	
Rationale for transformation	
No rationale provided	50.6
Correct for skewness	26.9
Normalize the distribution	8.1
Reduce the effect of outliers	6.9
Reduce heteroscedasticity	1.3
Other	6.2
Provision of citations	
No citations provided	80.3
At least one citation provided	19.7
Prescreening data	
No mention of prescreening	71.3
Examination of distribution	28.7
Reporting results for NLTs	
Effect of transformation on original distribution	
At least one effect reported	3.1
No effect reported	96.9
Correlation between raw and transformed scores	
Reported	0.8
Not reported	99.2
Reliability for transformed measures	
Reliability reported for transformed scores	0.0
No reliability reported	100.0
Correlations between transformed variables and other variables	
Reported	85.7
Not reported	9.8
Reported for some but not others	4.5
Results for transformed variables reported in main tables/figures	
Reported	82.3
Not reported	10.0
Reported for some but not others	5.4
Reported in tables/figures but not described in methods section	2.3
Effect of the NLT on results	
Reported	4.8
Not reported	95.2
Confusion in results section	
Nontransformed labels used to refer to transformed variables	44.6
NLTs not labeled as such in tables/figures	36.2
Descriptive statistics given for nontransformed data	8.5
No confusion	10.7
Alignment of hypotheses and statistical tests	
Aligned	5.9
Misaligned	91.2
No hypotheses stated	2.9
Discussing results for transformed variables	
Comparison of results for transformed and nontransformed variables	
Yes	3.1
No	96.9

(continued)

Table 2. (continued)

Category/Subcategory	Percentage
Confusion in discussion section	
No, transformed variables discussed as such	2.6
No, no discussion for transformed variables	34.4
Yes, results for transformed variables discussed as if they were for nontransformed variables	33.8
Yes, in the abstract	11.0
Yes, in both the abstract and discussion	18.2

addressed and the values of interrater agreement. Mean agreement across questions and raters was .88 and all values exceeded .70, indicating that independent coders agreed to a reasonable extent on how authors treated NLTs in the articles.

Results

Of the 324 articles in our data set, 123 (37.96%) included at least one NLT. Across these 123 articles measuring 2,021 variables, 314 (15.54%) variables were nonlinearly transformed, including 66 of 536 (12.31%) independent variables, 41 of 240 (17.08%) dependent variables, and 207 of 1,245 (16.63%) control variables. Thus, 65.92% (207/314) of NLTs were performed on control variables, 21.02% were conducted on independent variables, and 13.06% were performed on dependent variables. In 76% of cases, the original variables on which the NLTs were conducted were measured at the ratio level of measurement; in 17.8% they were measured at the interval level, in 4.7% at the ordinal level, and in 1.6% at the nominal level of measurement. Log transformations comprised 88.8% of NLTs, and winsorizing 4.2%. (Winsorizing involves setting presumably spurious outliers to a specified percentile of the data. For instance, winsorization might involve setting all data below the 5th percentile to the 5th percentile, and data above the 95th percentile to the 95th percentile). Square roots comprised 2.8% of NLTs, and dropping outliers and other forms of NLTs 4.2%. Size (e.g., of firm), counts (e.g., number of employees), age (e.g., of firms, employees), and sales were among the most commonly transformed variables.

Table 2 reports the findings for the coded NLT categories—justification, reporting, alignment, and discussion—and subcategories. There are several encouraging results, including that most authors report correlations between their transformed variables and others, and include the corresponding results in their primary tables and figures. However, the preponderance of poor practices far outweigh effective practices in our sample. As described below, there are many problems in how NLTs are justified, reported, interpreted, and discussed.

Justifying the Use of NLTs

Rationale for transformations. Researchers have several options when confronted with skewed or other nonnormal characteristics of a distribution, only one of which is to nonlinearly transform their data. Other options adopted by some authors in our sample are semiparametric models (e.g., Almandoz & Tilcsik, 2016), Bayesian approaches (Goranova, Abouk, Nystrom, & Soofi, 2017; Grand, 2017; Woo, Chae, Jebb, & Kim, 2016), median regression (Miller, Minichilli, & Corbetta, 2013), and OLS regression with a log-link function (McDonnell & Werner, 2016). For certain distributions of dependent variables, Poisson, negative binomial, and logit regression are options. Given the variety of alternatives, one aspect of justifying the use of NLTs is explaining why NLTs were used instead of other means of addressing nonnormality.

As reported in Table 2, over half the authors in our sample who used NLTs provided no rationale for doing so. Of those who did offer a rationale, none explained why they preferred using NLTs to other options. Instead, many authors simply said NLTs were performed to correct for skewness. For example, Mannor, Wowak, Bartkus, and Gomez-Mejia (2016) explained that they performed a natural log transformation on one of their independent variables, executive job anxiety, "to account for positive skew" (p. 1977); Tatarynowicz, Sytch, and Gulati (2016) entered two control variables into their model "as logged terms due to their skewed distributions over firms" (p. 60); and Merluzzi and Phillips (2016) took the log of a dependent variable, "due to the skewed distribution" (p. 101). Goranova et al. (2017) said they took a log of one independent variable to correct for both skewness and kurtosis. Other authors were less specific, saying only that they used NLTs to make distributions more normal. For example, Porter, Woo, and Campion (2016) stated that "age, organizational tenure, and firm size variables were transformed using either the square root or inverse transformation to ensure that the variable distributions approximated normality" (p. 648), and Kauppila (2016) "transformed the measure of employee age in years to its natural logarithm to ensure a normal distribution" (p. 373).

Other authors intended for NLTs to reduce the effect of outliers. For instance, Zhelyazkov and Gulati (2016), logged a control variable, age of investment fund, to reduce the influence of outliers (p. 286), and Shen and Benson (2016) reported that "since the distribution of employee numbers [a control variable] was skewed, we used a log transformed firm size to reduce the effect of outliers" (p. 1733). A variety of other rationales were provided, including reducing heteroscedasticity (Bermiss & Greenbaum, 2016; Liu, Gong, & Liu, 2014), turning a relative level of a variable into an absolute level (Nandkumar & Srikanth, 2016), accommodating an "uneven distribution" (Bermiss & Greenbaum, 2016, p. 269), reducing or managing "over dispersion" (Ge, Huang, & Png, 2016), and yielding "better-behaved" distributions (Zhelyazkov & Gulati, 2016).

Provision of citations. Instead of, or in addition to, providing a logical rationale for NLTs, researchers could offer citations to prior work that explains why transformation of a given variable or type of distribution is justifiable. For example, Vergne (2012) cites Greene's (2008) chapter on econometric analysis to support the log transformation for a positively skewed distribution of a dependent variable, media disapproval of firms. Another example is Ferris, Spence, Brown, and Heller's (2012) citation of Cohen et al.'s (2003) statistics text to support their square-root transformation of their dependent variable, number of deviant acts at work. Minbashian and Luppino (2014) justified their use of NLTs by saying, "In line with previous research (e.g., del Corral & Prieto-Rodriguez, 2010), the rankings were transformed using the natural logarithmic function to take into account that differences in player quality are larger at higher rankings" (pp. 905-906). These examples are exceptions in that, as reported in Table 2, in 4 out of 5 cases no citations were offered.

Prescreening data. Another means of justifying NLTs is examining distributions of data prior to conducting primary analyses. Accordingly, Cohen et al. (2003) advise that prescreening should be a requirement before undertaking NLTs. Prescreening involves examining untransformed and transformed data, considering relevant alternatives to NLTs, and identifying specific objectives for transformations. As shown in Table 2, in our articles prescreening was reported in less than 30% of cases. Authors who apparently prescreened their data seldom explained their methods but, rather, simply said that they used NLTs because their data were skewed or were in some other way nonnormal. These statements give the impression that perhaps the researchers conducted prescreening, though this was not explicitly stated.

Reporting Results for NLTs

Effects on original distributions. In addition to describing how the need for a NLT was determined, one might expect that the effects of the NLT on the original distribution would be reported (Cohen et al., 2003). One way to report the effects would be to apply the techniques mentioned for prescreening to compare the before-NLT and after-NLT distributions. This approach would reveal the change in number and extent of outliers, levels of skewness and kurtosis, and pattern of residuals produced by transforming scores on a given variable. An excellent example of this approach is Diestel, Wegge, and Schmidt (2014), who also clearly justify the need for transformation:

> Because the distribution of absence frequency deviated from the thresholds that are commonly seen as critical for unbiased parameter estimations (Hammer & Landau, 1981), all individual raw scores were subjected to a square root transformation (Clegg, 1983). After this transformation, the skewness and kurtosis of the pre- and post-questionnaire absence frequencies were either within or below the ranges deemed acceptable by Hammer and Landau (1981; see also Steel, 2003). (p. 363)

Another simple way to report the effect of a NLT would be to correlate the original and transformed scores: The weaker the correlation, the greater the impact the NLT had on the original distribution. As reported in Table 2, both prescreening and reporting correlations between original and transformed scores were extremely rare. In nearly all cases, researchers failed to report the effects of transformation.

Reliability. The reliability of measures is a central consideration in the interpretation of findings because low reliability can impugn the construct validity of measures and the accuracy of results. This is just as true for NLT scores as for nontransformed ones. Because NLTs change the variance of the untransformed variable, reliability estimates may differ between the untransformed and transformed variables. Reliability can be evaluated in a number of ways, including test-retest and parallel-methods approaches, interrater correlations and agreement, and indicators of internal consistency such as Cronbach's alpha. No authors in our sample reported the reliability of NLT variables.

Effects on results. Because NLT distributions differ from their nontransformed counterparts, the results involving NLTs can differ from those produced by nontransformed data. As shown in Table 2, it was very rare for authors to report the effect of NLTs on results (less than 5% did so), but there were exceptions. For example, Nguyen, Groth, and Johnson (2016) reported,

> Absenteeism findings are based on untransformed data. We repeated the analyses using log transformation, but it had little effect on the results, probably because the positive skews and variances for both transformed and untransformed data were similar. Given that we found a similar pattern of results, we retained the raw scores for ease of interpretation. (p. 638)

In several articles, comparisons between results of transformed and untransformed data were framed as checks on the robustness of findings. For example, in a footnote Pohler and Schmidt (2016) report,

> To check the robustness of the findings, we replicated all of the analyses described above in OLS regression by entering the natural log transformed turnover rate variables as the criteria.

> The pattern of results remained the same except for two findings.... These results do not substantively alter the interpretation of the main findings. (p. 412)

Another example is Zhelyazkov and Gulati (2016), who report, "The results are robust whether we use the social overlap measures (which implicitly standardize for the centrality of the two VCs in the dyad) or simpler measures, such as logged count of ties to abandoned/non-abandoned syndication partners of the alter" (p. 291). Finally, Pontikes and Barnett (2017) used a log for a key IV in their primary analyses and then conducted reanalysis with untransformed data: "Results are similar in magnitude if we use a one-year window ($p < .15$) or if we do not take the natural log ($p < .05$)" (pp. 166-167).

Again, we note that the above examples are infrequent relative to the total pool of NLT applications in our sample. Although Cohen et al. (2003) advocated comparing findings from analyses using both transformed and nontransformed data, in over 95% of cases authors in our sample did not do so. Not reporting these comparisons makes it virtually impossible for readers to evaluate the effects NLTs had on the construct validity of the original measures, the veracity of results, and the validity of the conceptual inferences from those results.

Alignment With Hypotheses

In the context of NLTs, we deemed misalignment to have occurred when the hypotheses were given in terms of the original, nontransformed variable and results in terms of the NLT variable. Alignment was moot in the few cases of inductive research in our sample because no hypotheses were offered (e.g., Ge et al., 2016), and when authors, such as Nguyen et al. (2016), reported similar results for both transformed and nontransformed data. However, as reported in Table 2, when NLTs were conducted on independent or dependent variables, misalignment of the tests and corresponding hypotheses occurred in over 90% of cases.

Note that the misalignment at issue here is different than that involving hypotheses that do not include control variables with statistical tests that do. This is an important issue but is beyond the scope of our discussion. We note, however, that authors in our sample rarely treated control variables in a manner consistent with published recommendations (Becker et al., 2016).

Discussion of NLTs

Given that most authors did not conduct analyses on the effects of NLTs on their results, it is not surprising that nearly 97% did not address these effects in the discussion section. Regarding terminology, one potentially confusing situation is when authors use nontransformed labels (e.g., number of employees) to refer to NLT variables (e.g., square root of the number of employees). This is problematic because calling one variable by the name of another variable can hamper the interpretation of research findings. As shown in Table 2, in our sample this situation happened in the results section in 44.6% of cases and in the discussion section in 33.8% of cases (in some cases it also occurred in the abstract).

A similar problem in the results section was that, in over one third of cases, transformed variables were not identified as such in tables or figures. Examples are a 2015 *ASQ* article in which several logged variables were included in tables but not at all discussed in the text, and a 2012 article in the *JOM* in which the authors nonlinearly transformed their primary dependent variable, deviance, but did not mention the transformed nature of the variable elsewhere in tables or text. Furthermore, in a few cases, NLT variables were used in the primary analyses while descriptive statistics were provided for nontransformed data. The relevance of these statistics was not explained.

Some Interesting Illustrations

We offer the following as examples of the variety of issues with NLTs in our sample, some of which are not entirely captured by our coding methodology. We intend these examples to be illustrative rather than accusatory, and as such we do not identity authors by name.

- A 2016 article in *JAP* wherein the authors ran analyses using the untransformed dependent variable, but then plotted the relationships in figures using the log-transformed dependent variable. Authors of a 2016 article in *Personnel Psych* did just the opposite, converting NLT turnover data (logarithmic scores from their negative binomial analysis) to untransformed scores to plot turnover rates. Both cases seem questionable because a variable and its log can have different relationships with other variables.
- A 2017 article in *SMJ* in which authors calculated logs of two independent and two control variables, and included these in Bayesian analyses. This seems odd given that one advantage of a Bayesian approach is that it does not assume or require normal distributions (Kruschke, Aguinis, & Joo, 2012).
- A 2015 article in *ASQ* in which the authors log transformed 13 of their 26 variables, but only mentioned one logged variable in the correlation table. The rest of the variables were not identified as transformed in any tables.
- A 2012 article in the *JOM* in which the authors nonlinearly transformed their primary dependent variable, but did not mention the transformation in the tables or text. Rather, the variable was discussed in text as if it were untransformed, and the only way a reader would know that the variable was transformed would be to examine the footnotes. This would be a challenge given that footnotes in the journal are collected at the end of the manuscript rather than at the bottom of the referencing page.
- A number of articles conducted additional transformations on NLT variables. Sometimes authors used linear transformations of the NLT variable, and in several cases NLT variables were used to create other variables, including interaction terms, difference scores, and aggregated level-2 variables in multilevel modeling. One 2014 *JAP* article used a log transformation on a predictor variable, then subtracted the transformed scores from 10. Then the authors centered the linearly transformed, logged predictor, squared it, and included it in an interaction term. Construct validity and generalizability issues associated with NLTs are compounded in such cases.

Ancillary Analyses

In analyses not reported in full here, we found that more articles in *AMJ*, *ASQ*, *and SMJ* reported NLTs than did articles in *JAP*, *JOM*, and *Personnel Psych*. Furthermore, the number of NLTs varied, with *ASQ* having significantly more NLTs per article than *JAP*, *JOM*, and *PP*, and *SMJ* having significantly more than *JOM* and *Personnel Psych*. No significant journal differences were found for data prescreening, the extent or nature of reporting, misalignment between hypotheses and analyses, or discussion of NLTs. There were also no journal differences regarding the rationale provided for using NLTs, with the exception of citations. The use of citations for justifying NLTs varied by journal, with marginal significance ($p < .10$). Specifically, the use of citations for this purpose occurred more often in *AMJ*, *JAP*, and *JOM* (42.86%, 50%, and 67.67%, respectively) than in *ASQ*, *SMJ*, and *Personnel Psych* (3.70%, 18.75%, and 25%, respectively). We suspect that the journal differences reflect the widespread use of archival data in the more macro journals (especially *SMJ and ASQ*), a proposition supported by our findings that NLTs are more likely to appear in archival studies than in studies using data from other sources, and that archival studies are more often reported in macro journals. Archival data often include size variables, counts, sales, and other commonly transformed variables, and such data may tend to be less normally distributed than variables typically found in more micro journals.

In addition, NLTs were most often performed on ratio level variables but, oddly, were also performed on interval, ordinal, and even nominal level data. NLTs often changed the level of measurement from ratio or interval to ordinal, as equal intervals between values in the original scale were not retained after transformation. This may make NLTs of continuous variables less appropriate than the original variables for statistical tests, despite nonnormality. Details on the ancillary analyses are available upon request from the first author.

Discussion and Recommendations

Our results suggest that there are serious problems with how many organizational science researchers choose, analyze, report, interpret, and discuss findings involving NLTs. The solutions to these problems are tied to the issues we earlier discussed regarding construct validity, interpretation, and generalizability. For instance, carefully choosing and evaluating NLTs is likely to improve construct validity by increasing the use of theoretically relevant transformations, and ensuring transformed variables are correctly analyzed. Properly reporting and discussing the corresponding results would improve the interpretation of findings, including the extent to which they may generalize to other populations and settings.

We acknowledge that we have engaged in some of the same practices as the authors in our study and, thus, have contributed to the problems we have identified. It is our hope that we can now offer recommendations, summarized in Table 3, to help avoid or reduce these problems in the future. The following is a deeper discussion of the issues and recommendations for more appropriately using NLTs.

Choosing NLTs

Our first four recommendations pertain to prescreening data as a precursor to deciding whether to use NLTs and, if relevant, which ones to use. Because some of the corresponding methods may not be widely known to or used by researchers, Appendix A provides specific examples to illustrate how the recommendations can be put into practice. The corresponding SPSS and Mplus syntax is provided in Appendix B.

> *Recommendation 1:* Prior to analysis, carefully examine the untransformed data for violations of statistical assumptions.

Prior to deciding whether to conduct NLTs, researchers should determine whether their data violate statistical assumptions and, if so, whether NLTs are the best way to address the problem. Because NLTs are ostensibly used to overcome violations of assumptions regarding normality, we focus our discussion on evaluating assumptions of multivariate normality, nonlinearity in relationships, and the distributions of residuals. However, we note that researchers should evaluate the extent to which their data violate all relevant statistical assumptions, including those unrelated to normality (e.g., reliability, independence of observations).

Our results indicate that researchers rarely engage in appropriate prescreening of their data for violations, or do not adequately report this information in their results. A prudent first step for testing multivariate normality (a key assumption of maximum likelihood estimation, MANOVA, and other multivariate statistics) and identifying potential outliers is to examine univariate normality for each variable by computing skewness and kurtosis and visualizing distributions using graphs (e.g., box-and-whisker plots, stem-and-leaf plots, histograms; Burdenski, 2000). For smaller samples, Kernel

Table 3. Recommendations for Using NLTs in Organizational Research.

Choosing NLTs

1. Prior to analysis, carefully examine the untransformed data for violations of statistical assumptions regarding normality.
2. When faced with nonnormality, consider all relevant alternatives rather than assuming NLTs are the best option.
3. Offer a clear, specific objective for nonlinearly transforming variables and, where feasible, provide relevant citations.
4. Inspect the shape of the variables' distributions to select the type of NLT to use, and consult the ladder of powers and other available techniques before making a final decision.

Analyzing

5. Examine the transformed data for violations of statistical assumptions regarding normality. One cannot assume that NLTs automatically produced normality.
6. Analyze the effect that the NLTs had on the original distribution, including measures of central tendency, dispersion, skewness, kurtosis, and outliers.
7. Wherever possible, calculate the reliability of both the original and nonlinearly transformed measures.
8. Evaluate the effects that NLTs had on results by comparing findings with and without the NLTs, and summarize the findings in the results section. Include both the original and transformed variables in the correlation matrix.
9. Transform the findings back into the variables' original units of analysis.

Reporting

10. Consistently label and report the results for NLTs in tables and figures.
11. In the title, abstract, and text, consistently identify and label transformed variables in a manner that makes clear that they were transformed, and how they were transformed.

Interpreting and discussing

12. Ensure that the interpretation and discussion of hypothesis tests reflect the NLTs.
13. Discuss results tentatively, remembering that the use of NLTs often changes the analyses from confirmatory to exploratory.
14. Be cautious in generalizing results. Findings based on NLTs may not extend to settings in which variables are not or cannot be nonlinearly transformed.

density plots may be a useful diagnostic for evaluating the normality of a distribution. Kernel density estimation is a nonparametric way to estimate the probability density function of a random variable, and serves as a data smoothing tool where inferences about the distribution of a population are made based on a sample (see Cohen et al., 2003; Howell, 2013). In the context of deciding whether NLTs are needed, Kernel density estimation would help researchers determine if their data are nonnormal in their populations of interest, or if the appearance of nonnormality is due to chance fluctuations in their samples. In our experience, however, plots and graphs of the data are at least as useful as Kernel density estimation, and more convenient because Kernel density programs are not readily available.

If univariate normality is established, researchers using multivariate statistics can proceed to test bivariate normality (normality of the joint distribution of each pair of dependent variables) and multivariate normality (normality of the joint distribution of all dependent variables, conditional on independent variables). Testing for bivariate normality involves examining scatterplots of the relationships between each pair of the dependent variables and noting elliptical patterns for each combination (Burdenski, 2000). Multivariate normality can be tested using the Mahalanobis Distance, the distance between two multivariate populations, and Mardia's (1985) statistic. We note, though, that Mardia's statistic may be overly sensitive to minor violations of multivariate normality in large samples and have inadequate power in small samples (Kline, 2016).

To test the potential for nonlinearity in the relationships between independent and dependent variables, researchers should use bivariate scatterplots of the variables, and of the residuals against each continuous predictor variable and against the predicted dependent variable (Cohen et al., 2003).

Linear and loess (locally weighted smoothing) fit lines can be superimposed on these scatterplots to allow researchers to better visualize potential nonlinearity. Researchers who do not hypothesize nonlinear relationships can test the robustness of their findings by including nonlinear terms in post hoc analyses—for example, including quadratic terms when testing moderation (Edwards, 2008a).

During prescreening, researchers should test for violations of assumptions regarding homoscedasticity and the normality of residuals by examining the scatterplots of residuals against each predictor variable and against the predicted criterion variable, with interpretation aided by loess fit lines. Cohen et al. (2003) also describe a modified Levene test as a means of evaluating whether the residuals corresponding to low values of a variable are equal to the residuals corresponding to high values of the variable. We recommend the score test to evaluate heteroscedasticity. The score test (Breusch & Pagan, 1979; Rosopa, Schaffer, & Schroeder, 2013) involves regressing the squared residuals from the original regression on the predictors believed to be the cause of the heteroscedasticity. The regression sum of squares (SSR) is obtained from this analysis and compared with the sum of squares error (SSE) from the original regression in the ratio $\chi^2(df) = (\text{SSR}/2) \div (\text{SSE}/2)^2$, where $df =$ number of predictors. To test for normality of residuals, researchers can use a normal Q-Q (quantile-quantile) plot to compare theoretical (population) and empirical (sample) distributions of residuals to assess the normality of the sample distribution (Cohen et al., 2003). For a more thorough assessment, these plots can be compared with other methods for examining the distribution of residuals (e.g., histograms, the Shapiro-Wilk [1965] test). Illustrations of prescreening techniques are provided in Appendix A.

Recommendation 2: When faced with nonnormality, consider all relevant alternatives rather than assuming NLTs are the best option.

If prescreening suggests nonlinearity, there are alternative analytical procedures for addressing it other than automatically turning to NLTs. For nonnormally distributed dependent variables, alternative analyses include Poisson regression, logistic regression, negative binomial models, zero-inflated models, censored models (e.g., Tobit), sample selection (Heckman) models, and two-part regression. Although the choice of model depends on the distribution of the dependent variable, we agree with Cohen et al. (2003) that when a choice is available to use NLTs or employ a form of the generalized linear model (GLM), use of the GLM is preferable. Instead of transforming individual scores, GLM transforms means via a link function, a formula that relates the mean of responses to the linear predictors in the model. NLTs and GLM can lead to different results in that, for instance, the mean of log-transformed responses is not the same as the logarithm of the mean response (Lindsey & Jones, 1998). Where feasible, the use of a GLM to transform means often allows the results to be more easily interpreted than does NLTs because mean parameters remain on the same scale as the measured responses.

If prescreening indicates that the form of the relationship is misspecified (e.g., should include a quadratic term), researchers should respecify the relationship by including relevant terms in the appropriate equation. This decision is critical as both parameter estimates and standard errors will be biased if the form of the relationship is misspecified (Cohen et al., 2003). If heteroscedasticity is present, researchers can test the model using a robust estimator (e.g., a sandwich estimator; White, 1980), or can explicitly model the distribution of residuals using a nonlinear function (Muthén, Muthén, & Asparouhov, 2017). For regression analyses, another option is to use random coefficient estimation to allow the slope to vary across individuals. This approach explicitly allows heteroscedastic residual variance and permits researchers to add predictors to account for variance in the random slope.

Skewed distributions or outliers can result in nonnormal sampling distributions for parameter estimates. Researchers faced with nonnormal data can test for this possibility by comparing results to those using nonsymmetric confidence intervals produced by bootstrapping or Bayesian estimation (Muthén et al., 2017). Other options include median regression (to address outliers), partial least square approaches, and semiparametric models. For researchers interested in linearizing nonlinear relationships among variables in regression analyses, weighted least squares, polynomial, nonlinear least squares, and nonparametric regressions are alternative estimation procedures. Each of these methods has its strengths and limitations and it is incumbent on the researcher to consider these in his or her specific case to determine which are most appropriate for a given set of data.

It is worth noting that most statistical techniques used by organizational researchers, including multiple regression, do not make explicit assumptions about the distributions of predictor variables. For example, it is common practice to include dichotomous predictors (e.g., gender) in regression models without transforming them. However, nonnormally distributed predictors may result in nonnormality in the sampling distributions of parameter estimates. Thus, if nonlinearity in the distributions of continuous predictors is present, the robustness of results should be evaluated using the appropriate procedures described above (e.g., by reanalyzing data using a robust estimator or bootstrapped standard errors). Appendix A provides comparative results across several alternative methods.

> *Recommendation 3:* Offer a clear, specific objective for nonlinearly transforming variables and, where feasible, provide relevant citations.

If, after following Recommendations 1 and 2, a researcher decides to further consider NLTs, he or she needs to specify and justify the purpose of using NLTs. Cohen et al. (2003) proposed that there are three basic goals for carrying out NLTs:

1. Simplifying relationships among variables. NLTs always change the form of relationship between two variables, and researchers sometimes want to "linearize" the relationship for ease of communication and interpretation. This goal is justifiable when the NLT produces a more conceptually meaningful unit of analysis, as when a log of a variable is taken to test a hypothesized relationship of diminishing returns.
2. Eliminating heteroscedasticity. NLTs change the variance of residuals around best fit functions (as in a regression line) and, if properly chosen and implemented, can make the conditional variances of residuals for different distributions more equal.
3. Normalizing residuals. NLTs change the distribution of both the original variable and the residuals and, hence, can under certain conditions help to satisfy the requirements of statistical tests.

Researchers using NLTs should specify which one or more of these goals they are seeking to accomplish, noting that NLTs will simultaneously affect the relationship between variables and the distributions of residuals, even if this is not what is intended. Cohen et al. (2003) further specify when NLTs may be justifiable:

(a) When strong theory dictates the use for NLTs for estimating critical model parameters
(b) When an equation using NLTs provides a better explanation of the phenomenon (e.g., log dollars to reflect the utility of money)
(c) When NLTs produce substantially improved overall model fit
(d) When NLTs do not introduce new difficulties into the model

We consider the first two justifications most persuasive. NLTs should be used when there is a theoretically sound reason for preferring a transformed distribution over a nontransformed one.

Recommendation 4: Carefully examine the shape of the variables' distributions to select the type of NLT to use, and consult the ladder of powers and other available techniques before making a final decision.

Assuming a defensible goal and convincing justification for NLTs, researchers next need to choose which NLTs to use. Which transformation should be chosen depends upon the particular shape of a distribution. To provide guidance on which NLTs to choose, Tukey (1977) introduced the ladder of powers to explain the effects of various NLTs on the shape of a distribution (see also Mosteller & Tukey, 1977, and Velleman & Hoaglin, 1981). For example, reexpressing each value in a set of scores to a power less than one pulls in a stretched-out upper tail while spreading out a bunched in lower tail, thus reducing positive skew. The further from one the reexpression, the greater the effect of the NLT: taking the square root of scores has a small effect, logging scores a somewhat larger effect, and so on for more extreme NLTs. Reexpressing values to a power greater than one reduces asymmetry to the low side, reducing negative skew. Again, the further from one the reexpression, the greater the effect the NLT: squaring scores has a modest effect, cubing scores a greater effect, and so on. Other techniques for determining which NLT to use include bulge rules, loess lines, and the Box-Cox and Box-Tidwell procedures (Cohen et al., 2003; Kline, 2016). Appendix A illustrates some of these procedures.

Given the many ways that distributions depart from normality, it is surprising that the great majority of NLTs in our sample involved log transformations. This finding suggests that either numerous distributions are nonnormal in precisely the same way (i.e., moderately and positively skewed), that many researchers are aware of the difficulties of interpretation accompanying NLTs other than logs, or that most authors are simply following convention in using the log as their NLT of choice. We find the last explanation most plausible because it seems unlikely that many different variables in different populations and contexts have identical shapes. Furthermore, few of our authors identified a moderate positive skew as the reason for logging scores, and none identified interpretation problems as reasons for ruling out other NLTs. These points, along with our finding that most authors offered scant rationale for conducting NLTs, lead us to emphasize the importance of stating a definite goal for using NLTs and a clear justification for the specific choice of NLT.

Analyzing and Reporting

Our results suggest that it is very rare for organizational science researchers to analyze and report the effect of NLTs on original distributions or on their results, or the correlation between raw and transformed scores. In no case did researchers report the reliability of their NLT data and, in a majority of cases, researchers reported their results in potentially confusing ways, that is, by speaking of their transformed variables as if they were not transformed or failing to label variables as transformed in tables and figures.

Recommendation 5: Examine the transformed data for violations of statistical assumptions, and analyze the effect that the NLTs had on the original distribution.

It is important to assess the effect of NLTs on a distribution of scores because there is no guarantee that a given NLT will produce a normal distribution, or even a distribution that is more normal than the original one. As Cohen et al. (2003) noted, NLTs may "improve [analyses] with respect to one goal while degrading it with respect to others" (pp. 221-222). For example, reducing skew can increase kurtosis (the sharpness of the peak of the distribution) and vice versa, and a given NLT can create outliers or exacerbate their effects compared to the original scores. It is also vital to note that normality is just one of several assumptions of a particular statistical technique and, in meeting normality assumptions using NLTs, researchers may simultaneously violate a different assumption. For example, NLTs may change the level of measurement from ratio or interval to ordinal, as equal intervals between values in the original scale are not always retained after transformation. This may make NLT variables less appropriate than the original variables for statistical tests, despite greater normality.

Therefore, researchers are well-advised to evaluate their posttransformation data as carefully as they do the pretransformation data, using the same techniques to evaluate statistical assumptions, including normality. (Appendix A includes the evaluation of posttransformation effects on distributions and comparisons with their pretransformation counterparts.) If the NLTs served their purpose, the researcher can move to the next stage of analysis. If not, then the researcher needs to consider additional action. Perhaps a less or more extreme type of NLT is needed; for example, a square root or reciprocal root rather than a log transformation. If this approach does not help, then one of the alternatives to conducting NLTs (e.g., Poisson or negative binomial regression, partial least square approaches, Bayesian estimation) is likely preferable. If Recommendation 2 was followed, the researcher has already considered these options, but if NLTs do not serve their intended purpose then the decision needs to be revisited.

> *Recommendation 6:* Wherever possible, calculate the reliability of both the original and nonlinearly transformed measures.

As with other variables, proper methodology requires researchers to evaluate and report the reliability of NLT scores. Cronbach's alpha is typically used for untransformed data from surveys or tests, and it is equally applicable and necessary for transformed data. Test-retest, parallel forms, or split-half reliability can also be assessed and, as with Cronbach's alpha, needs to be calculated separately for the NLT data. For archival data, reliability can pertain to the consistency of scores obtained through different sources, the consistency of information recorded in a particular source, and the consistency of collecting and reporting of retrieved data. More specifically, issues of reliability and accuracy of archival data include differences in sampling across databases (Lara, Osma, & Noguer, 2006), data recording errors and differences in coding policies (Courtenay & Keller, 1994), definitional discrepancies and measurement error (Yang, Vasarhelyi, & Liu, 2003), misclassification errors (Mills, Newberry, & Novack, 2003), and omission, delisting, and survivorship bias (Olbrys & Majewska, 2014). These problems exist even in the most popular databases, such as Compustat. Thus, assessing and reporting the reliability of measures is equally important whether data are archival or from other sources.

Finally, one might think that reliability is most critical for independent and dependent variables and that, because NLTs are most often used for control variables, this issue is not particularly relevant for most researchers using NLTs. However, the quality of measurement of control variables, including their reliability, is as essential to a study's analyses and results as are the metrics for other variables (Atinc, Simmering, & Kroll, 2012; Becker et al., 2016; Edwards, 2008b).

> *Recommendation 7:* Evaluate the effects that NLTs had on results by comparing findings with and without the NLTs, and summarize the findings in the results section. Include both the original and transformed variables in the correlation matrix.

If NLT variables are reliable and have the desired effect on distributions of interest, then a reasonable next step is to evaluate the effect the transformations have on the findings. A straightforward means of accomplishing this is to analyze the data both with and without transforming the variables. If analyses using NLTs and those using untransformed scores yield essentially the same results then, for ease of interpretation, researchers can present the untransformed findings and simply report that using NLTs produced the same or similar results. However, as Cohen et al. (2003) note, when transformed and nontransformed results differ, "the researcher is pressed to develop an explanation of why the results in the transformed metric are more appropriate" (p. 249). Assuming the researcher has followed the prior recommendations, this should not be difficult to do: The results in the transformed metric are more appropriate because the original data clearly violated assumptions of normality, other alternatives for addressing the problem were ruled out, and the use of NLTs met a specific objective. Furthermore, the particular NLTs were carefully selected and satisfied statistical assumptions, and the corresponding scores were as reliable, if not more so, than the scores in the original metric. Without this kind of explanation, readers and reviewers can be justifiably skeptical about the superiority of the results in the transformed metric.

> *Recommendation 8:* Transform the findings back into the variables' original units of analysis.

The final set of needed analyses is transforming findings based on the NLTs back into the original metrics. The method of retranslation depends on which variables are transformed and the particular NLT that is used. For example, if a dependent variable is log-transformed in a regression analysis, the interpretation of a standardized regression coefficient for a given independent variable (β) is that it estimates the predicted change in the log of the dependent variable associated with a one-unit increase in the independent variable, holding other independent variables constant. An exponential equation can be applied to express the β in terms of the percentage change in the original units of the dependent variable, thereby providing a meaningful measure of effect size. Appendix A includes an illustration of retranslation.

In our judgment, translation to the original metric should be a standard procedure whenever NLTs are used. Without this process, interpreting effect sizes and drawing practical conclusions is very difficult. We note that retranslations other than logs of dependent variables can be difficult to perform and challenging to interpret. This is additional reason to consider all options to addressing nonnormality before deciding upon NLTs.

> *Recommendation 9:* Consistently identify and label NLTs in a manner that makes clear that they were transformed, and how they were transformed.

In addition to the reporting issues discussed above, we found other potentially confusing practices in our sample. As shown in Table 2, these practices included using labels for untransformed variables to refer to transformed variables, and failing to label NLTs in tables, figures, and even in abstracts and article titles. Given that NLTs create a different set of scores for an entirely different variable, we propose that much greater caution be taken in reporting NLT results.

Interpreting and Discussing NLT Results

Serious misinterpretations can occur when hypotheses are stated in terms of an original, nontransformed variable while results are provided in terms of the NLT variable. In such instances, the construct validity of the transformed scores will likely be low and, therefore, tests involving the transformed variable will be improper tests of the hypotheses. This misalignment problem is moot with exploratory or inductive research or when authors find similar results for both transformed and nontransformed data. However, in our sample, when NLTs were conducted on independent or dependent variables, misalignment of the tests and corresponding hypotheses occurred in the great majority of cases (see Table 2).

> *Recommendation 10:* Ensure that the interpretation and discussion of hypothesis tests reflect the NLTs. Either state hypotheses in terms of NLTs or translate NLT results back into the original units of analysis.

There are two ways to reduce NLT misalignment. The first is to state the hypotheses in terms of transformed variables (prior to conducting analyses). This approach is reasonable when the NLT produces a more conceptually meaningful unit of analysis, as when a log of a variable is taken to test a hypothesized relationship of diminishing returns. The second way to decrease (but not eliminate) misalignment is, per Recommendation 8, to translate results back into the original units of analysis. We earlier discussed ways to accomplish retranslation, and simply note here that correspondence between hypotheses and analyses is crucial for interpreting hypothesis tests.

Recommendation 11: Discuss results tentatively, remembering that the use of NLTs changes the analyses from confirmatory to exploratory.

Finally, even under the best of conditions there is reason for researchers to be cautious when interpreting results that involve NLTs. Consider the following example:

1. Bob hypothesizes that affective commitment (AC) is positively associated with job performance (JP), controlling for tenure with the company.
2. He collects data on the three variables and runs a regression analysis to estimate the relationship between AC and JP. Call this β_1.
3. Based on analyses he did during prescreening, Bob decides to take the log of JP. He then runs a regression of JP on AC and tenure, producing a NLT-based regression coefficient representing the relationship between AC and log(JP). Call this β_2.
4. Following our recommendations, Bob interprets β_2 in terms of the percentage change in JP (in original units) corresponding to a one-unit increase in AC.

It is critical to note that the interpretation, in terms of percentage change, based on β_2 is highly unlikely to be the same as a corresponding interpretation of β_1. The first coefficient is based on JP while the second, though retranslated, result is based on log(JP). More intuitively, if the two results were the same there would be no reason to perform the NLT in the first place. This has two implications. First, although a retranslation of results may be more interpretable than the nontranslated version, this does not mean the test based on NLTs was a test of the hypothesis. In this example, Bob's hypothesis implies β_1 as the test, not β_2 or any interpretation of β_2. Hence, retranslation, while helpful in some ways, does not solve the problem of misalignment. The second and related point is that β_2 and any translation of the coefficient should be considered exploratory because Bob's hypothesis was stated in the original units of analysis. Thus, like other exploratory findings, tentative interpretation and cross-validation of results are called for. In sum, the use of NLTs often changes a research effort from one of deduction (theory testing) to one of induction (theory building).

Recommendation 12: Be cautious in generalizing results. Findings based on NLTs may not extend to settings in which variables are not or cannot be nonlinearly transformed.

We have explained the generalizability issue in previous sections but highlight the issue here to further alert researchers to the necessity for caution. Because any NLT creates a new variable that may be substantively different from the untransformed variable, relationships between the transformed variable and other variables cannot be assumed to generalize to the untransformed variables. As shown above, this is true even when NLT results are retranslated back into the original units of analysis. Furthermore, as in the performance appraisal example described in the introduction, NLTs can create distributions that do not exist in the real world. Hence, they should not be presumed to apply to the workplace.

Conclusion

We have documented a variety of common problems in how organizational science researchers choose, analyze, report, and interpret results involving NLTs. Furthermore, we have explained how these problems likely hamper scientific progress by decreasing the construct validity of transformed scores, producing improper tests of hypotheses, and in other ways hindering the interpretation and generalizability of findings. Our goal has not been to scold researchers or to introduce inflexible rules to replace sound professional judgment. Rather, we hope that our recommendations and the illustrations in the appendices will allow researchers to make more sound judgments in the future—and will encourage editors and reviewers to hold authors accountable for the quality of such judgments.

Appendix A. Illustrations of Sound Practice

Carefully Examine Untransformed Data

The following examples use simulated data, included in Online Appendix 3, to illustrate the use of scatterplots and loess lines to evaluate data for violation of statistical assumptions, including normality and heteroscedasticity. The applications of the score test and Q-Q plots are also illustrated.

Suppose we hypothesized that job satisfaction (measured via a 7-point scale) at Time 1 (T1) would have a negative linear association with absenteeism at Time 2 (T2) after controlling for absenteeism at T1. Absenteeism was scored as a count of the number of absences in the prior year. As shown in Table A1 and Figure A1, we found high levels of skewness and kurtosis for absenteeism (only T2 is shown in Figure A1). The histogram shows that absenteeism was positively skewed, with about half of employees experiencing no absences. When absences did occur, they tended to be infrequent, although there were a few employees with more than two absences.

The boxplot in Figure A2 shows that T2 absenteeism scores above 5 were outliers. (A similar finding occurred for T1 absenteeism.) These outliers were not deleted because they were accurate observed values of absenteeism.

Based on the preceding analyses, these data appear to violate univariate normality. Although OLS regression makes no assumption about the distribution of the variables, nonnormality may have potential implications for the form of the association and the distribution of the residuals. Figure A3 is the scatterplot of T1 job satisfaction and T2 absenteeism, with linear and loess lines superimposed. This analysis suggests that the association was approximately linear in form, although the loess line (the lower one) suggests the potential for curvilinearity. (The scatterplot of the T1 and T2 absenteeism scores indicated that the association between these two variables was approximately linear.)

We next ran OLS regression without transforming the variables or considering possible outliers or curvilinearity. This allowed us to save the residuals, predicted values, Cook's distance values, and influence statistics that were used to test relevant statistical assumptions. Job satisfaction and T1 absenteeism accounted for 3.6% of the variance in T2 absenteeism and, after controlling for

Table A1. Statistics.

	T1Absence	T2Absence	T1JobSat
N	498	498	498
M	1.2430	1.1968	4.6949
SD	1.96529	1.84332	1.12668
Skewness	2.543	2.881	–0.255
SE of skewness	0.109	0.109	0.109
Kurtosis	7.041	9.485	–0.186
SE of kurtosis	0.218	0.218	0.218
Min	0.00	0.00	1.11
Max	10.00	10.00	6.98

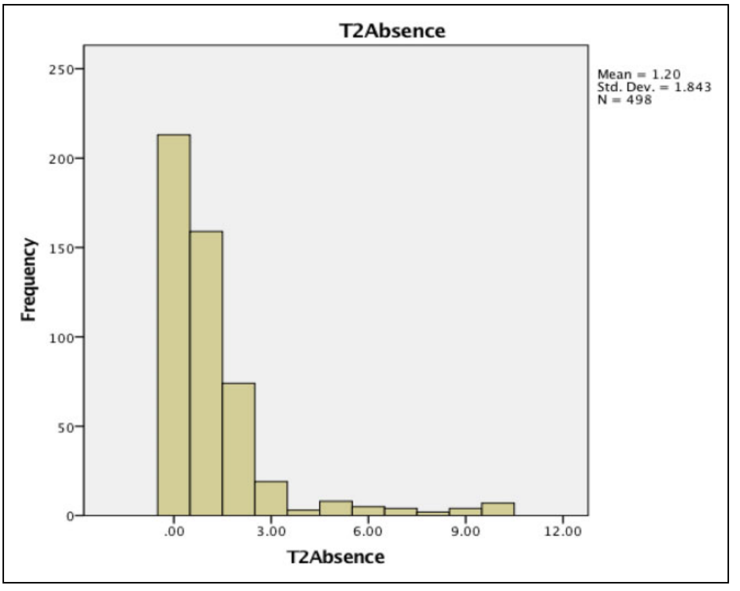

Figure A1. Frequency distribution for Time 2 absenteeism.

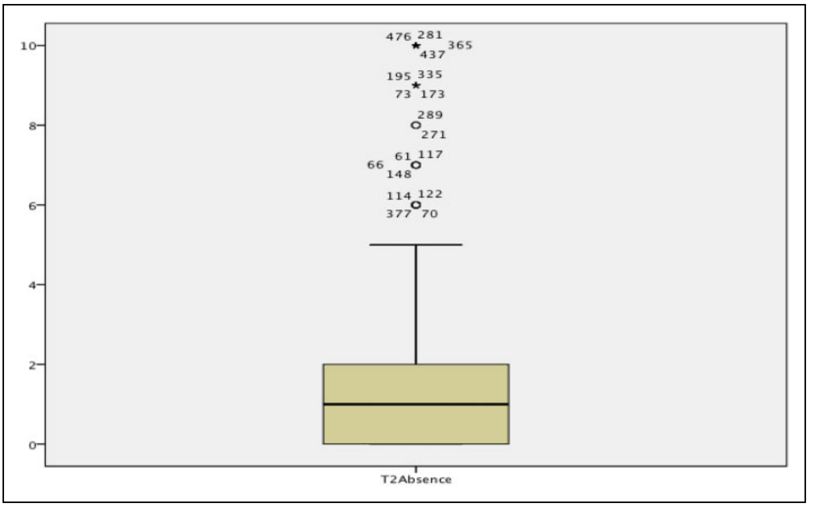

Figure A2. Boxplot for Time 2 absenteeism.

absenteeism at T1, for every unit increase in job satisfaction, absenteeism at T2 decreased by .29. Figure A4 is helpful for visualizing potential heteroscedasticity, in this case for T1 absenteeism. The plot indicates that the variance of residuals was larger at lower levels of absenteeism that at higher levels of absenteeism.

To statistically test whether variance in the residuals changed as a function of the predictors, T1 absenteeism and job satisfaction, we used the score test. We found that heteroscedasticity did not appear to be a function of the predictors, $\chi^2(2) = 0.0002$, *ns*. However, this test is biased if residuals are not normally distributed. Therefore, we examined the relevant Q-Q plot to investigate normality. (The Q-Q plot tests the unstandardized residual against the expected normal value and also indicates

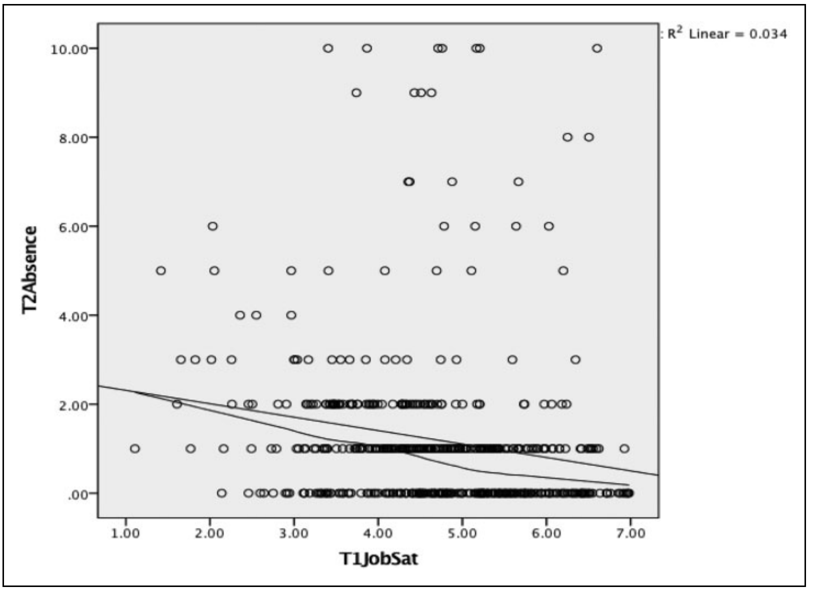

Figure A3. Scatterplot of Time 1 job satisfaction and Time 2 absenteeism scores.

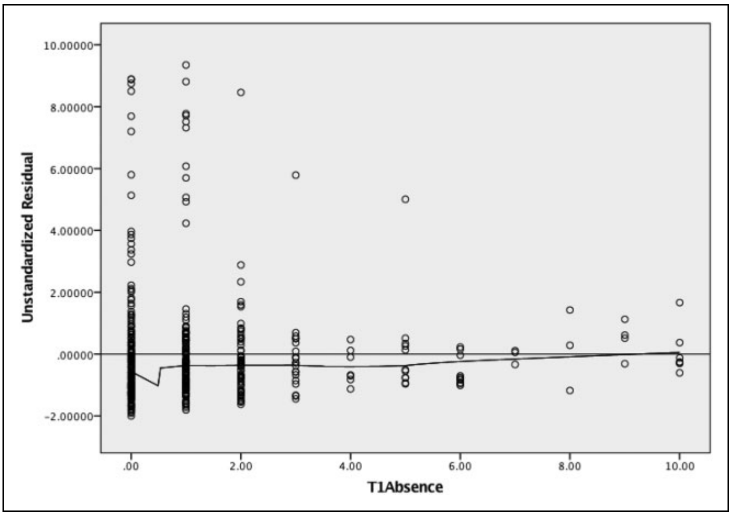

Figure A4. Scatterplot of Time 1 absenteeism and unstandardized residuals.

the potential for curvilinearity.) Figure A5 indicates potential curvilinearity in the residuals, so the score test may have been biased. In addition, the scatterplot of the predicted T2 absenteeism values against the residuals (Figure A6) indicates the potential for curvilinearity.

Finally, in analyses not reported in full here, we followed Aguinis et al.'s (2013) recommendations regarding outliers in the predictors. Specifically, we examined Cook's D, DFFITS, and DFBETAS to determine the extent to which a data point affects the regression coefficients. There were no cases of influential prediction outliers based on the cutoffs proposed by Cohen et al. (2003).

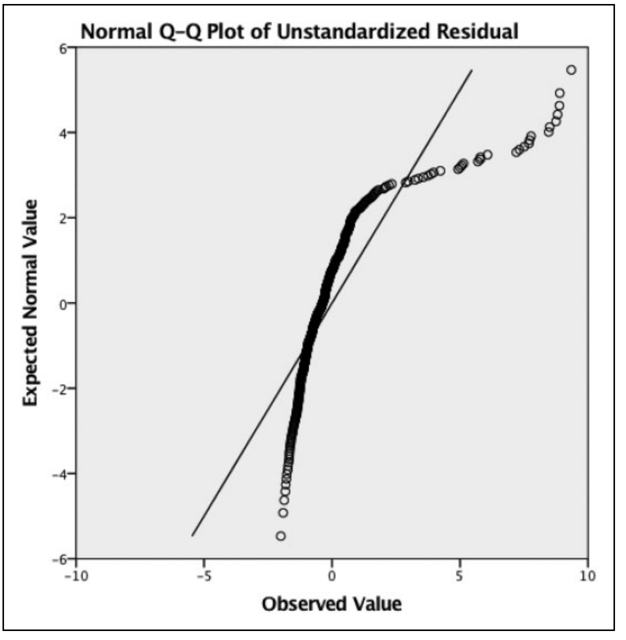

Figure A5. Normal Q-Q plot of unstandardized residuals.

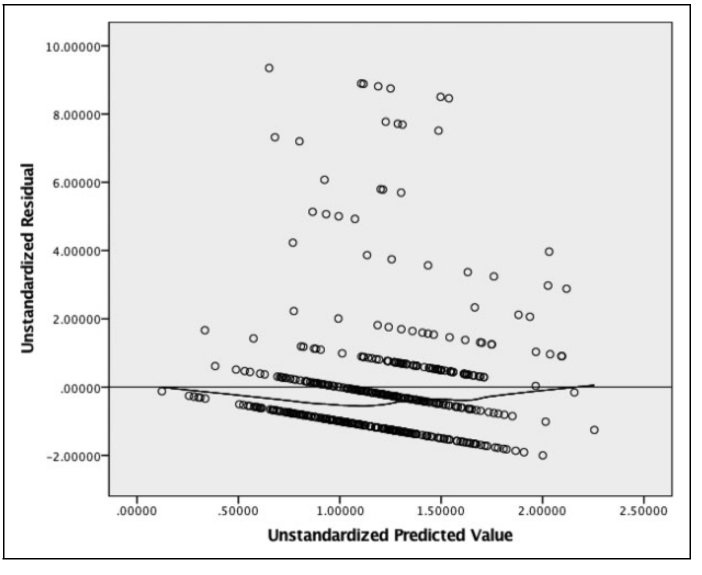

Figure A6. Scatterplot of unstandardized predicted values and unstandardized residuals.

Consider Alternatives to NLTs

The above data visualization indicated a skewed predictor and outcome variable, and the potential for curvilinearity in the association between job satisfaction and T2 absenteeism. Heteroscedasticity and nonnormality of residuals also appeared to be present. Given these conditions, potential alternatives to OLS regression are (a) Poisson regression to better account for the distribution of the

Table A2. Parameter Estimates.

			95% Wald CI		Hypothesis Test				95% Wald CI for Exp(*B*)	
Parameter	*B*	*SE*	Lower	Upper	Wald χ^2	*df*	Sig.	Exp(*B*)	Lower	Upper
(Intercept)	1.315	0.1610	0.999	1.630	66.665	1	.000	3.724	2.716	5.105
T1JobSat	–0.237	0.0355	–0.307	–0.168	44.542	1	.000	0.789	0.736	0.846
T1Absence	–0.053	0.0252	–0.102	–0.003	4.343	1	.037	0.949	0.903	0.997
(Scale)	1[a]									

Note: Dependent variable: T2Absence. Model: (Intercept), T1JobSat, T1Absence.
a. Fixed at the displayed value.

outcome variable, (b) a squared job satisfaction term in post hoc analyses to control for potential curvilinearity in the OLS estimates, and (c) bootstrapped confidence intervals, robust estimators, or Bayesian estimation. These methods may be preferable to NLTs because the original scores and distributions are retained. In the current approach, these methods were treated as checks to the robustness of the findings. Thus, they constituted post hoc analyses because they were considered only after inspecting the data.

Poisson regression. We ran a Poisson regression in SPSS as an alternative to OLS estimation. As shown in Table A2, for every one-unit increase in job satisfaction, the predicted number of absences decreased by 0.79 (e^b). For example, a person with a job satisfaction rating of 5 had an expected number of absences of 1.14; a person with a job satisfaction rating of 1 had an expected number of absences of 2.94. Unlike the use of NLTs, Poisson regression provides results in the original units. Poisson regression assumes that the variance of the residuals around each predicted rate is equal to the predicted rate. If this assumption does not hold, the negative binomial model may be preferable to Poisson regression.

Adding squared terms. Because data screening indicated the potential for nonlinearity in the association between job satisfaction and T2 absenteeism, we tested for curvilinearity by including a squared job satisfaction term in the regression equation. The results indicated that the curvilinear term was nonsignificant. Moreover, the regression coefficient was only minimally affected by the inclusion of the curvilinear term. These results indicate that the prior original estimate of association between job satisfaction and absenteeism was not confounded by curvilinearity.

Using alternative estimation procedures. First, we repeated the regression analyses using a maximum likelihood robust (MLR) estimator in Mplus. This estimator is robust to nonnormality, and adjusts standard errors by using a sandwich estimator. Results showed that the standard errors were slightly smaller when using the MLR estimator compared to OLS (0.072 for job satisfaction and 0.042 for T1 absenteeism). However, the parameter estimates and significance of results did not change.

Second, we repeated the analysis in Mplus using a maximum likelihood estimator with bootstrapped confidence intervals. Bootstrapping offers nonsymmetric confidence intervals and is therefore useful for situations involving parameter estimates with nonnormal sampling distributions. The 95% confidence interval for the regression of T2 absenteeism on job satisfaction was –0.42 to –0.15, and the regression coefficient and significance remained the same as in the OLS analysis.

Finally, we repeated the analysis in Mplus using a Bayesian estimator. Unlike maximum likelihood, Bayesian estimation does not assume multivariate normality or normality of sampling distributions. Using the Bayes estimator, the estimate for the regression of T2 absenteeism on job

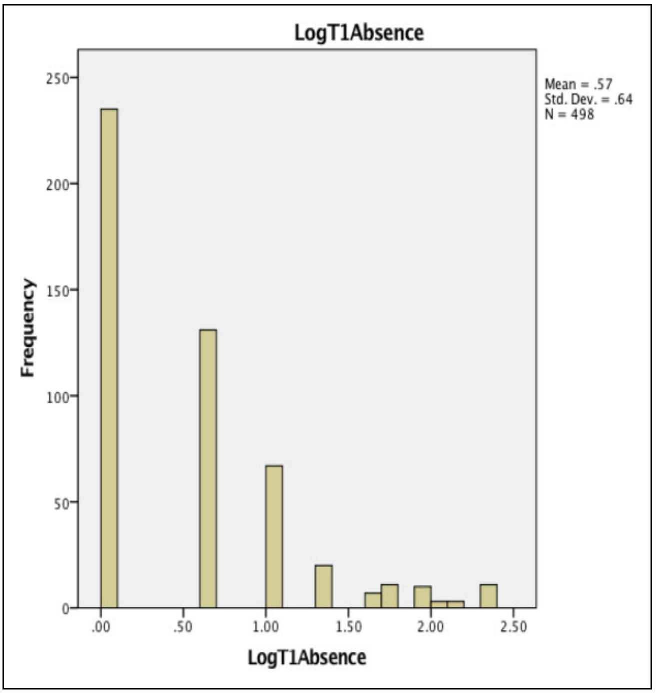

Figure A7. Frequency distribution of log(Time 1 absenteeism + 1).

satisfaction was –0.31 (95% CI [–0.45, –0.21]). Thus, in this case, the interpretation of results remains the same as in OLS estimation.

Consider Which NLTs to Use (if any)

Log transformation. Assume that rather than using the alternative procedures described above, we decided to log-transform the two absenteeism variables so that they better approximated a normal distribution. How would that affect the residuals, results, and interpretation? Because log(0) is undefined, transforming these variables required that we add a constant to all values so that the zero values in the original variables could be included in analyses. The choice of constant has implications for the transformed distribution. For example, adding a constant of 1 produces the distribution in Figure A7, whereas adding a constant of .001 produces the distribution in Figure A8.

Suppose we added a constant of 1 and then took the natural log (as in Figure A7). As shown in Table A3, the skewness and kurtosis were reduced compared to the original distribution. Note, too, that the level of measurement changed from ratio to ordinal because the equal intervals between responses for the original variable were not retained after transformation (e.g., the difference between 0 and 1 is now 0.69, the difference between 1 and 2 is now 0.41).

The correlation matrix of original and transformed variables indicated that the logged variables were strongly related to the original variables. However, whereas the original T1 absenteeism variable was significantly associated with job satisfaction, the log of T1 absenteeism was unrelated to satisfaction. The test-retest reliability also differed somewhat depending on whether the original scores were used ($r = -.07$) or the transformed scores were used ($r = -.03$). In contrast, as shown in the scatterplot in Figure A9, the association between job satisfaction and log T2 absenteeism was stronger than the original association between satisfaction and T2 absenteeism.

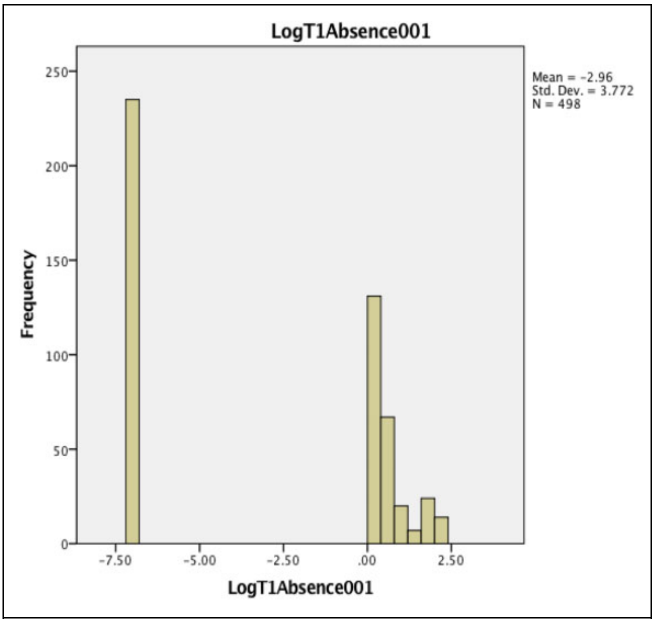

Figure A8. Frequency distribution of log(Time 1 absenteeism $\pm$ 0.001).

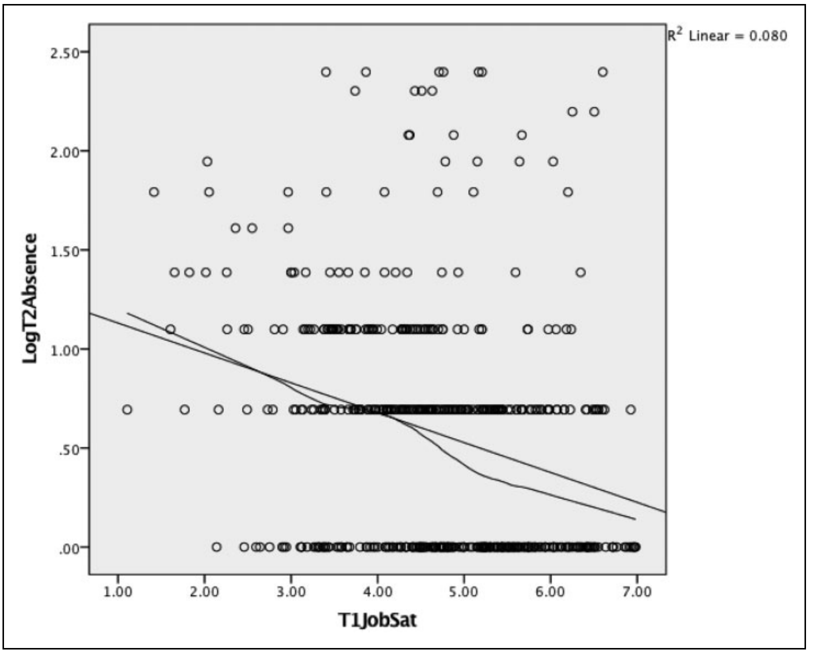

Figure A9. Scatterplot of Time 1 job satisfaction and log(Time 2 absenteeism + 1).

Furthermore, OLS regression demonstrated that the size of the standardized regression coefficient for satisfaction was substantially larger using transformed variables compared to the original variables, and the R-square was more than double the original. However, because the dependent variable was in different units than the original variable, these results could not be directly compared

Table A3. Statistics.

	LogT1Absence	LogT2Absence
M	0.5651	0.5732
SD	0.63983	0.60081
Skewness	0.935	0.906
SE of skewness	0.109	0.109
Kurtosis	0.149	0.478
SE of kurtosis	0.218	0.218
Min	0.00	0.00
Max	2.40	2.40

Table A4. Correlations.

		AntilogY	PRE_1	T2Absence
AntilogY	Pearson correlation	1	.967**	.188**
	Sig. (2-tailed)		.000	.000
	N	498	498	498
PRE_1	Pearson correlation	.967**	1	.191**
	Sig. (2-tailed)	.000		.000
	N	498	498	498
T2Absence	Pearson correlation	.188**	.191**	1
	Sig. (2-tailed)	.000	.000	
	N	498	498	498

**Correlation is significant at the .01 level (2-tailed).

Table A5. Statistics.

	SqrtT1Absence	SqrtT2Absence
Skewness	0.830	0.773
SE of skewness	0.109	0.109
Kurtosis	0.036	0.318
SE of kurtosis	0.218	0.218
Min	0.00	0.00
Max	3.16	3.16

to those from the nontransformed model. The unstandardized regression coefficient for job satisfaction was –0.15 ($SE = 0.02$), indicating that, when controlling for log absenteeism at Time 1, for every one unit increase in job satisfaction, log absenteeism at Time 2 decreased by 0.15. Using the equation $(e^{\beta 2} - 1) * 100$, we find, holding log Time 1 absenteeism constant, Time 2 absenteeism decreases by 13.92% for every one unit increase in job satisfaction.

To facilitate the comparison, we took the antilog of each predicted Y score and subtracted 1 from the resulting values to "undo" the transformation so that the predicted Y scores were in the original units. (For information on antilogs, see Cohen et al., 2003, sec. 6.4.3; Bland & Altman, 1996; and http://www.wikihow.com/Do-Antilog.) This allowed us to compare the correlation between the predicted Y and the original scores in the untransformed and transformed models to evaluate the fit of each respective model (Cohen et al., 2003). The results in Table A4 indicate that the correlations between the predicted Y scores from the untransformed and transformed models were very

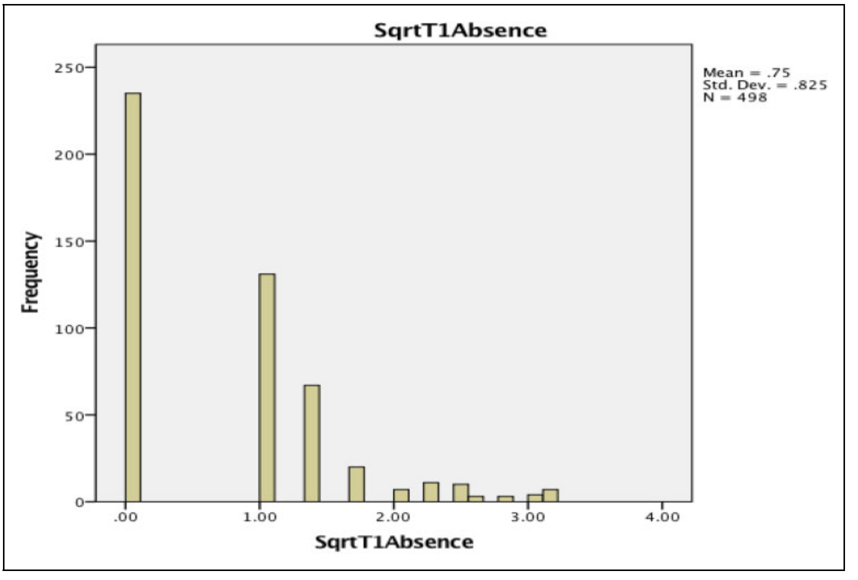

Figure A10. Frequency distribution of square root (Time 1 absenteeism).

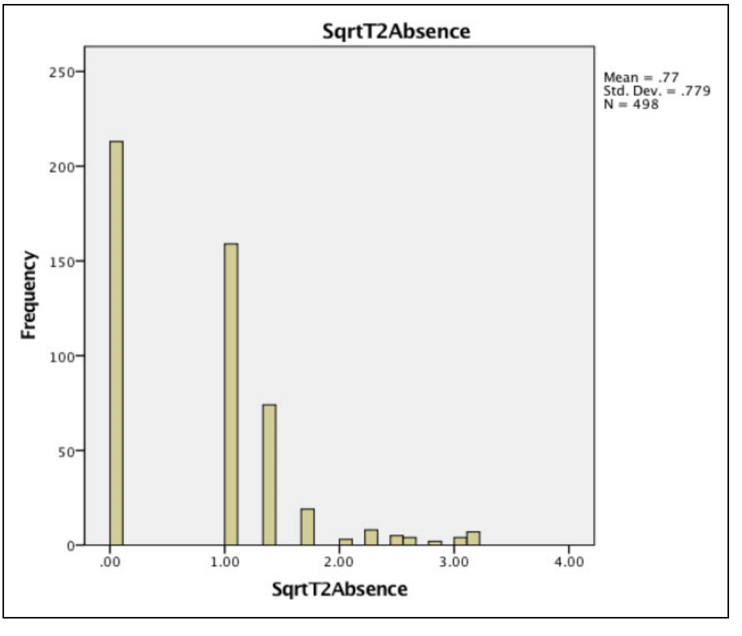

Figure A11. Frequency distribution of square root (Time 2 absenteeism).

similar (.191 versus .188, respectively). Thus, the models appeared to have similar fit to the data, with the results from the untransformed model slightly favored.

As in the original OLS regression, the OLS regression with transformed variables required that we check assumptions regarding the residuals and the form of the X-Y relationships. The results indicated similar patterns of heteroscedasticity in T1 absenteeism compared to the original analyses. Likewise, the plot of the predicted values against the unstandardized residuals and the Q-Q plot were

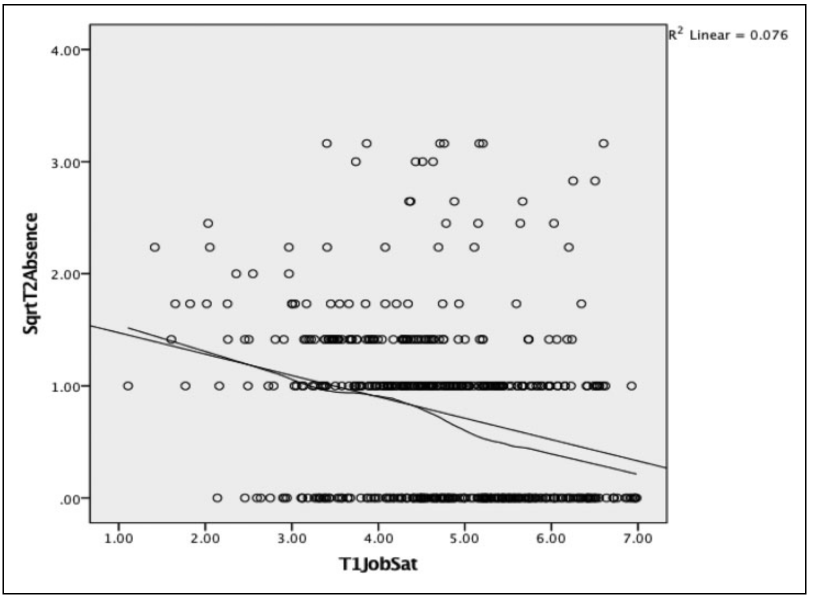

Figure A12. Scatterplot of Time 1 job satisfaction and square root (Time 2 absenteeism).

very similar to those obtained from the original analyses, although the potential for curvilinearity identified in the original analysis was somewhat obscured when transformed data are used. Comparison of the untransformed and transformed results indicated that, in this case, transformation affected the skewness and kurtosis of the variable but did not have a substantial impact on the residuals. This is important to note, as OLS regression makes assumptions about the normality of residuals, not the normality of the variables themselves. Thus, transformation did not make the data any more suitable for OLS regression than the original untransformed data. It did, however, change the meaning of the results so that correlation and regression coefficients could not be directly compared across models.

Square-root transformation. Suppose we decided to use a different transformation than the natural logarithm. In this example, we took the square root of the absenteeism variables. The simple descriptive findings are shown in Table A5.

The regression results indicated that T1 job satisfaction is a significant predictor of the square root of T2 absenteeism, after controlling for the square root of T1 absenteeism. However, these results are difficult to interpret. The chain rule, a mathematical formula used to take the derivative of a composition of functions, can be used to describe the changes in Y as a function of X, but the computations are cumbersome (for us) and not readily available in mainstream statistical software. However, to compare the fit of the square root model with that of the untransformed model, we took the square of each predicted Y to "undo" the transformation. The results indicated that the predicted Y scores from the untransformed model were correlated .191 with the raw T2 absenteeism scores, whereas the predicted Y scores from the transformed model were correlated .187 with the raw T2 absenteeism scores. Thus, as with the log transformation, the original and transformed models appeared to have similar fit to the data, with the results from the untransformed model slightly favored. Also as with the logged analyses, heteroscedasticity appeared to be relatively unaffected by the transformation, and the plot of the predicted values against the unstandardized residuals and the Q-Q plot were very similar to those obtained from the original analyses (although the potential for curvilinearity was somewhat obscured when transformed data were used).

Thus, neither the log nor the square root transformations resolved the issues with heteroscedasticity and nonnormality. This calls into question the need for nonlinear transformation for these data.

Appendix B. Syntax for Analyses Reported in Appendix A

Unless otherwise noted, the syntax is for SPSS.

Table A1 and Figures A1 and A2

```
* Encoding: UTF-8.
*First examine descriptive statistics to visualize potential distribution issues

FREQUENCIES VARIABLES=T1Absence T2Absence T1JobSat
  /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN SKEWNESS SESKEW KURTOSIS
SEKURT
  /HISTOGRAM
  /ORDER=ANALYSIS.

EXAMINE VARIABLES=T1Absence T2Absence T1JobSat
  /COMPARE VARIABLE
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL
  /MISSING=LISTWISE.
```

Figure A3 and Related Plots

```
*Examine scatterplots of each IV-DV and IV-IV combination for potential violations of linearity
*Linear and loess lines can be added to these scatterplots in SPSS chart editor by double-clicking on
the chart and then selecting Elements > Fit Line at Total

GRAPH
  /SCATTERPLOT(BIVAR)=T1JobSat WITH T2Absence
  /MISSING=LISTWISE.

GRAPH
  /SCATTERPLOT(BIVAR)=T1Absence WITH T2Absence
  /MISSING=LISTWISE.

GRAPH
  /SCATTERPLOT(BIVAR)=T1JobSat WITH T1Absence
  /MISSING=LISTWISE.
```

Figure A4, Related Regression, Score Test, and Outlier Analysis

```
*Untransformed regression saving residuals, predicted values, Cook's distance, DFBETA, and
DFFIT
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
```

```
   /NOORIGIN
   /DEPENDENT T2Absence
   /METHOD=ENTER T1Absence T1JobSat
   /SAVE PRED COOK RESID DFBETA DFFIT.
```

*Note the Sum of Squares Residual from this analysis for the score test
*Examine potential outliers based on Cook's Distance, DFBETA, DFFIT as per Aguinis et al. (2013)

```
GRAPH
   /SCATTERPLOT(BIVAR)=T1Absence WITH RES_1
   /MISSING=LISTWISE.

GRAPH
   /SCATTERPLOT(BIVAR)=T1JobSat WITH RES_1
   /MISSING=LISTWISE.
```

*Score Test

```
COMPUTE SqRes = RES_1*RES_1.
EXECUTE.

REGRESSION
   /MISSING LISTWISE
   /STATISTICS COEFF OUTS CI(95) R ANOVA
   /CRITERIA=PIN(.05) POUT(.10)
   /NOORIGIN
   /DEPENDENT SqRes
   /METHOD=ENTER T1Absence T1JobSat.
```

Figure A5

*QQ Plot

```
PPLOT
   /VARIABLES=RES_1
   /NOLOG
   /NOSTANDARDIZE
   /TYPE=Q-Q
   /FRACTION=BLOM
   /TIES=MEAN
   /DIST=NORMAL.
```

Figure A6

*Examine scatterplot of predicted Y values against residuals

```
GRAPH
   /SCATTERPLOT(BIVAR)=PRE_1 WITH RES_1
   /MISSING=LISTWISE.
```

Table A2: Poisson Regression

*Poisson regression

```
GENLIN T2Absence WITH T1Absence T1JobSat
  /MODEL T1Absence T1JobSat INTERCEPT=YES
DISTRIBUTION=POISSON LINK=LOG
  /CRITERIA METHOD=FISHER(1) SCALE=1 COVB=MODEL MAXITERATIONS=100
MAXSTEPHALVING=5
  PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-012 ANALYSISTYPE=3(WALD)
CILEVEL=95 CITYPE=WALD
    LIKELIHOOD=FULL
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION
(EXPONENTIATED).
```

Adding Squared Terms

```
*Testing for curvilinearity

COMPUTE c_T1JobSat = T1JobSat-4.6949.
COMPUTE T1JobSatSq = c_T1JobSat*c_T1JobSat.
EXECUTE.

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT T2Absence
  /METHOD=ENTER T1Absence T1JobSat c_T1JobSatSq.
```

Alternative Estimation Procedures (in Mplus)

```
TITLE: Alternative Estimation Procedures
DATA: FILE IS
VARIABLE: Names are
T1Absence
T1JobSat
T2Absence;
USEVARIABLES are
T1Absence
T1JobSat
T2Absence;

ANALYSIS:
! ESTIMATOR=MLR; ! for maximum likelihood estimation
! ESTIMATOR=BAYES; ! for Bayesian estimation
! BOOTSTRAP=10000; ! for bootstrapped estimates

MODEL:
T2Absence on T1JobSat T1Absence;

OUTPUT: SAMPSTAT STDYX; ! CINT(BCBOOTSTRAP); ! for bootstrapped estimates
```

Table A3 and Figures A7 and A8 and Related Analyses

```
COMPUTE T1Absence1 = T1Absence+1.
COMPUTE T2Absence1 = T2Absence+1.
COMPUTE LogT1Absence = ln(T1Absence1).
COMPUTE LogT2Absence = ln(T2Absence1).
EXECUTE.

*First examine descriptive statistics to visualize potential distribution issues

FREQUENCIES VARIABLES=LogT1Absence LogT2Absence T1JobSat
  /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN SKEWNESS SESKEW KURTOSIS
SEKURT
  /HISTOGRAM
  /ORDER=ANALYSIS.

EXAMINE VARIABLES=LogT1Absence LogT2Absence T1JobSat
  /COMPARE VARIABLE
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL
  /MISSING=LISTWISE.

*Examine scatterplots of each IV-DV and IV-IV combination for potential violations of linearity
*Linear and loess lines can be added to these scatterplots in SPSS chart editor by double-clicking on
the graph, Elements > Fit Line at Total

GRAPH
  /SCATTERPLOT(BIVAR)=T1JobSat WITH LogT2Absence
  /MISSING=LISTWISE.

GRAPH
  /SCATTERPLOT(BIVAR)=LogT1Absence WITH LogT2Absence
  /MISSING=LISTWISE.

GRAPH
  /SCATTERPLOT(BIVAR)=T1JobSat WITH LogT1Absence
  /MISSING=LISTWISE.

*Examine how transformation affected correlations

CORRELATIONS
  /VARIABLES=T1Absence T2Absence LogT1Absence LogT2Absence T1JobSat
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.

*OLS regression with transformed data

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT LogT2Absence
```

```
/METHOD=ENTER LogT1Absence T1JobSat
/SAVE PRED COOK RESID DFBETA DFFIT.
```

Table A5 and Related Analyses

```
*Compute Antilog of predicted Y

COMPUTE AntiLogY = EXP(PRE_2) – 1.
EXECUTE.

*Correlate predicted Y from each model with original scores

CORRELATIONS
   /VARIABLES=T2Absence PRE_1 AntiLogY
   /PRINT=TWOTAIL NOSIG
   /MISSING=PAIRWISE.
```

Note: The syntax for the square-root transformation (Table A5, Figures A10-A12) is analogous to that for the log analyses.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Melissa M. Robertson and Robert J. Vandenberg's contributions to this article were supported in part by Grant R18DK090672 from the National Institute of Diabetes and Digestive Kidney Diseases. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Diabetes and Digestive Kidney Diseases or the National Institutes of Health.

Supplemental Material

Online Appendices 1 and 2 are available at http://orm.sagepub.com.

References

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*, 270-301.

Aguinis, H., & Joo, H. (2015). Debunking myths and urban legends about how to identify influential outliers. In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends* (pp. 206-223). New York, NY: Routledge.

Almandoz, J., & Tilcsik, A. (2016). When experts become liabilities: Domain experts on boards and organizational failure. *Academy of Management Journal*, *59*, 1124-1149.

Atinc, G., Simmering, M. J., & Kroll, M. J. (2012). Control variable use and reporting in macro and micro management research. *Organizational Research Methods*, *15*, 57-74.

Becker, T. E., Atinc, G., Breaugh, J. A., Carlson, K. D., Edwards, J. R., & Spector, P. E. (2016). Statistical control in correlational studies: 10 essential recommendations for organizational researchers. *Journal of Organizational Behavior*, *37*, 157-167.

Bermiss, Y. S., & Greenbaum, B. E. (2016). Loyal to whom? The effect of relational embeddedness and managers' mobility on market tie dissolution. *Administrative Science Quarterly*, *61*, 254-290.

Bland, J. M., & Altman, D. G. (1996). Transformations, means, and confidence intervals. *British Medical Journal*, *312*, 1079.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*, 211-252.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, *47*, 1287-1294.

Burdenski, T. K. (2000, April). *Evaluating univariate, bivariate, and multivariate normality using graphical procedures*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

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

Courtenay, S. M., & Keller, S. B. (1994). Errors in databases revisited: An examination of the CRSP shares-outstanding data. *Accounting Review*, *69*, 285-291.

de Vaus, D. (2002). *Analyzing social science data: 50 key problems in data analysis*. Thousand Oaks, CA: Sage.

Diestel, S., Wegge, J., & Schmidt, K. H. (2014). The impact of social context on the relationship between individual job satisfaction and absenteeism: The roles of different foci of job satisfaction and work-unit absenteeism. *Academy of Management Journal*, *57*, 353-382.

Edwards, J. R. (2008a). Seven deadly myths of testing moderation in organizational research. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends* (pp. 143-164). New York, NY: Routledge.

Edwards, J. R. (2008b). To prosper, organizational psychology should . . . overcome methodological barriers to progress. *Journal of Organizational Behavior*, *29*, 469-491.

Emerson, J. D. (1983). Mathematical aspects of transformation. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Understanding robust and exploratory data analysis* (pp. 247-282). New York, NY: John Wiley.

Ferris, D. L., Spence, J. R., Brown, D. J., & Heller, D. (2012). Interpersonal injustice and workplace deviance: The role of esteem threat. *Journal of Management*, *38*, 1788-1811.

Ge, C., Huang, K. W., & Png, I. P. (2016). Engineer/scientist careers: Patents, online profiles, and misclassification bias. *Strategic Management Journal*, *37*, 232-253.

Goranova, M., Abouk, R., Nystrom, P. C., & Soofi, E. S. (2017). Corporate governance antecedents to shareholder activism: A zero-inflated process. *Strategic Management Journal*, *38*, 415-435.

Grand, J. A. (2017). Brain drain? An examination of stereotype threat effects training on knowledge acquisition and organizational effectiveness. *Journal of Applied Psychology*, *102*, 115-150.

Greene, W. H. (2008). The econometric approach to efficiency analysis. In H. O. Fried, C. A. K. Lovell, & S. S. Schmidt (Eds.), *The measurement of productive efficiency and productivity growth* (pp. 92-250). New York, NY: Oxford University Press.

Gullikson, A. (2006). *Interpreting OLS regression and transformations*. Retrieved from http://www.uoregon.edu/aarong/teaching/V3212_Outline/node8.html

Howell, D. C. (2013). *Statistical methods for psychology* (8th ed.). Belmont, CA: Wadsworth.

Joo, H., Aguinis, H., & Bradley, K. J. (2017). Not all nonnormal distributions are created equal: Improved theoretical and measurement precision. *Journal of Applied Psychology*, *102*, 1022-1053.

Kauppila, O. P. (2016). When and how does LMX differentiation influence followers' work outcomes? The interactive roles of one's own LMX status and organizational context. *Personnel Psychology*, *69*, 357-393.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford.

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, *15*, 722-752.

Lara, J. M. G., Osma, B. G., & Noguer, B. G. D. A. (2006). Effects of database choice on international accounting research. *Abacus*, *42*, 426-454.

Lindsey, J. K., & Jones, B. (1998). Choosing among generalized linear models applied to medical data. *Statistics in Medicine*, *17*, 59-68.

Liu, W., Gong, Y., & Liu, J. (2014). When do business units benefit more from collective citizenship behavior of management teams? An upper echelons perspective. *Journal of Applied Psychology*, *99*, 523-534.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171.

Mannor, M. J., Wowak, A. J., Bartkus, V. O., & Gomez-Mejia, L. R. (2016). Heavy lies the crown? How job anxiety affects top executive decision making in gain and loss contexts. *Strategic Management Journal*, *37*, 1968-1989.

Mardia, K. V. (1985). Mardia's test of multinormality. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Vol. *5*, pp. 217-221). New York, NY: John Wiley.

McDonnell, M. H., & Werner, T. (2016). Blacklisted businesses: Social activists' challenges and the disruption of corporate political activity. *Administrative Science Quarterly*, *61*, 584-620.

Merluzzi, J., & Phillips, D. J. (2016). The specialist discount: Negative returns for MBAs with focused profiles in investment banking. *Administrative Science Quarterly*, *61*, 87-124.

Miller, D., Minichilli, A., & Corbetta, G. (2013). Is family leadership always beneficial? *Strategic Management Journal*, *34*, 553-571.

Mills, L. F., Newberry, K. J., & Novack, G. F. (2003). How well do Compustat NOL data identify firms with US tax return loss carryovers? *Journal of the American Taxation Association*, *25*, 1-17.

Minbashian, A., & Luppino, D. (2014). Short-term and long-term within-person variability in performance: An integrative model. *Journal of Applied Psychology*, *99*, 898-914.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression*. Reading, MA: Addison-Wesley.

Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2017). *Regression and mediation analysis using Mplus*. Los Angeles, CA: Muthén & Muthén.

Nandkumar, A., & Srikanth, K. (2016). Right person in the right place: How the host country IPR influences the distribution of inventors in offshore R&D projects of multinational enterprises. *Strategic Management Journal*, *37*, 1715-1733.

Nguyen, H., Groth, M., & Johnson, A. (2016). When the going gets tough, the tough keep working: Impact of emotional labor on absenteeism. *Journal of Management*, *42*, 615-643.

Olbrys, J., & Majewska, E. (2014). On some empirical problems in financial databases. *Pensee Journal*, *76*, 1-9.

Pohler, D., & Schmidt, J. A. (2016). Does pay-for-performance strain the employment relationship? The effect of manager bonus eligibility on nonmanagement employee turnover. *Personnel Psychology*, *69*, 395-429.

Pontikes, E. G., & Barnett, W. P. (2017). The non-consensus entrepreneur: Organizational responses to vital events. *Administrative Science Quarterly*, *62*, 140-178.

Porter, C. M., Woo, S. E., & Campion, M. A. (2016). Internal and external networking differentially predict turnover through job embeddedness and job offers. *Personnel Psychology*, *69*, 635-672.

Rosopa, P. J., Schaffer, M. M., & Schroeder, A. M. (2013). Managing heteroscedasticity in general linear models. *Psychological Methods*, *18*, 335-351.

Russell, C. J., & Dean, M. A. (2000). To log or not to log: Bootstrap as an alternative to the parametric estimation of moderation effects in the presence of skewed dependent variables. *Organizational Research Methods*, *3*, 166-185.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591-611.

Shen, J., & Benson, J. (2016). When CSR is a social norm: How socially responsible human resource management affects employee work behavior. *Journal of Management*, *42*, 1723-1746.

Tatarynowicz, A., Sytch, M., & Gulati, R. (2016). Environmental demands and the emergence of social structure: Technological dynamism and interorganizational network forms. *Administrative Science Quarterly*, *61*, 52-86.

Tinsley, H. E., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, *22*, 358-376.

Tukey, J. W. (1957). On the comparative anatomy of transformations. *Annals of Mathematical Statistics*, *28*, 602-632.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Velleman, P. F., & Hoaglin, D. C. (1981). *Applications, basics, and computing of exploratory data analysis*. Boston, MA: Duxbury Press.

Vergne, J. P. (2012). Stigmatized categories and public disapproval of organizations: A mixed-methods study of the global arms industry, 1996-2007. *Academy of Management Journal*, *55*, 1027-1052.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817-838.

Woo, S. E., Chae, M., Jebb, A. T., & Kim, Y. (2016). A closer look at the personality-turnover relationship: Criterion expansion, dark traits, and time. *Journal of Management*, *42*, 357-385.

Yang, D. C., Vasarhelyi, M. A., & Liu, C. (2003). A note on the using of accounting databases. *Industrial Management & Data Systems*, *103*, 204-210.

Yeo, I., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*, 954-959.

Zhelyazkov, P. I., & Gulati, R. (2016). After the break-up: The relational and reputational consequences of withdrawals from venture capital syndicates. *Academy of Management Journal*, *59*, 277-301.

Author Biographies

Thomas E. Becker is a professor of management in the College of Business at the University of South Florida Sarasota-Manatee. He attended the Ohio State University, and his primary research interests are in the areas of employee commitment, motivation, job performance, and research methods. His prior work on the latter includes the proper treatment of control variables and the examination of additive versus multiplicative method effects in organizational research.

Melissa M. Robertson is a post-doctoral research associate in the Industrial-Organizational Psychology program at the University of Georgia. Her primary research interests are in the areas of relational processes, the work-life interface, research methods, and employee well-being.

Robert J. Vandenberg is the Robert O. Arnold Professor of Business and the Chair of the Department of Management in the Terry College of Business at the University of Georgia. He is a former editor of *ORM* and a fellow in and short course instructor for the Consortium for the Advancement of Research Methods and Analysis (CARMA; http://cba.unl.edu/outreach/carma/). His primary research interests are in research methods and statistics, and in the interface between individual health and well-being, and the workplace.