Organizational Research Methods
13(3) 515-539

DOI: 10.1177/1094428109333339
http://orm.sagepub.com
SAGE

Customer-Centric Science: Reporting Significant Research Results With Rigor, Relevance, and Practical Impact in Mind

Herman Aguinis[1], Steve Werner[2], JéAnna Lanza Abbott[2], Cory Angert[2], Joon Hyung Park[2], and Donna Kohlhausen[2]

Abstract

In response to the ongoing concern regarding a science-practice gap, we propose a customer-centric approach to reporting significant research results that involves a sequence of three interdependent steps. The first step involves setting an alpha level (i.e., a priori Type I error rate) that considers the relative seriousness of falsely rejecting a null hypothesis of no effect or relationship (i.e., Type I error) relative to not detecting an existing effect or relationship (i.e., Type II error) and reporting the actual observed *p* value (i.e., probability that the data would be obtained if the null hypothesis is true). The second step involves reporting estimates of the size of the effect or relationship, which indicate the extent to which an outcome is explained or predicted. The third step includes reporting results of a qualitative study to gather evidence regarding the practical significance of the effect or relationship. Our proposal to report research results with rigor, relevance, and practical impact involves important changes in how we report research results with the goal to bridge the science-practice gap.

Keywords

statistical significance, practical significance, hypothesis testing

> Have a conviction in the importance of what you are doing, and work more than ever to establish its significance for others—both within and especially beyond the Academy
>
> Hambrick (1994, p. 16)

[1] Indiana University
[2] University of Houston

Corresponding Author:
Herman Aguinis, Department of Management and Entrepreneurship, Kelley School of Business, Indiana University, 1309 E. 10th Street, Bloomington, IN 47405-1701
Email: haguinis@indiana.edu

It has been 15 years since Hambrick (1994), in his 1993 Academy of Management Presidential address, lamented that we, scholars in the organizational sciences, have a "minimalist ethos: ... minimal visibility, minimal impact." He noted that "Each August, we come to talk to each other [at the Academy of Management's annual meetings]; during the rest of the year we read each others' papers in our journals and write our own papers so that we may, in turn, have an audience the following August: an incestuous, closed loop" (p. 13). He went on to write that "if we believe in the significance of advanced thinking and research on management, then it is time we showed it. We must recognize that our responsibility is not to ourselves, but rather to the institutions around the world that are in dire need of improved management, as well as those individuals who seek to be the most effective managers they possibly can be. It is time for us to break out of a closed loop. It is time for us to matter" (1994, p. 13).

Has anything changed in the last 15 years regarding this closed loop and the impact of our research on management practice and society at large? Consider the following evidence. First, managers still do not read our journals. For example, Rynes, Colbert, and Brown (2002) assessed how frequently human resource managers read various publications using a scale with the following anchors: 1 (*never*), 2 (*rarely*), 3 (*sometimes*), 4 (*usually*), and 5 (*always*). The *Academy of Management Journal* received a mean rating of 1.1. The *Academy of Management Executive* (now *Academy of Management Perspectives*), which should be more attractive and appealing to practitioners, also received an appalling low mean rating of 1.1. Second, scholars still write for a scholarly audience and practitioners still write for a practitioner audience (Cascio & Aguinis, 2008). For example, about 94% of articles published in *HRMagazine*, which is a practitioner publication, are written by practitioner authors (Rynes et al., 2002). Conversely, about 90% of articles published in scholarly journals such as the *Journal of Applied Psychology* and *Personnel Psychology* between 2003 and 2007 were written by academics (Cascio & Aguinis, 2008). Third, special issues of our journals continue to be published regularly lamenting the science-practice divide (e.g., Special research forum, 2001) and journal editors continue to write editorials stating that much work is needed so that scholars and practitioners engage in meaningful dialogues (e.g., Hillman, 2009), indicating that the gap still exists. Fourth, several independent recent assessments have reached the conclusion that the research-practice gap has not been bridged and that there is still a great divide between the science and practice worlds (Baldridge, Floyd, & Markóczy, 2004; Bartunek, 2007; Cascio & Aguinis, 2008; Latham, 2007; Rynes, Giluk, & Brown, 2007; Shapiro, Kirkman, & Courtney, 2007; Vermeulen, 2007). In sum, unfortunately, Hambrick's (1994) advice noted in this article's opening quote regarding the need to establish the significance of our work seems as timely and current today as it was 15 years ago.

Several excellent suggestions have been offered to narrow the science-practice divide (for reviews, see Cascio & Aguinis, 2008; Rynes, 2007). For example, Shapiro et al. (2007) suggested sabbaticals for academics in business practice, either as "translators" of research results or as researchers on a set of practitioner-oriented research issues. Baldridge et al. (2004) suggested that academics study questions that challenge not only existing theories but also existing management practices. Anderson (2007) suggested that academics become more strategically involved in senior managerial decision making by serving on boards of directors. Tushman and O'Reilly (2007) advocated greater participation by academics in executive-education contexts as a means to develop relationships with practitioners.

Given that suggestions such as the ones described above have been offered on an ongoing basis for more than 20 years and the divide still exists, it seems painfully obvious that unless there is a real and noticeable change, the science-practice gap will continue to exist. In this article, we propose a change in how we academics report the results of our research with the goal of addressing the monumental challenge of establishing the significance of our work. In addition to bridging the science-practice divide, our approach also bridges the quantitative and qualitative academic traditions and epistemological approaches (Willig & Stainton-Rogers, 2008). We readily recognize that this change will demand effort on the part of authors as well as journal editors and reviewers. But, unless a real and substantive change takes place in how we report research results, the 2013 presidential address of the Academy of Management is likely to raise similar concerns in terms of the science-practice gap and the lack of impact of our research as its 1993 predecessor (i.e., Hambrick, 1994).

Our proposed approach, which we label *customer-centric science*, involves reporting significant research results in a way that is rigorous and relevant and therefore meets the needs of both academics and practitioners (i.e., the customers of our research). Thus, our proposal is consistent with the goal of the entire management field to create theory and research with the goal of improving organizational practices (e.g., Koontz, 1961, Smiddy, 1962). We adopt Gelade's (2006) definition of practitioners, namely, those who make recommendations about the management or development of people in organizational settings, or advise those who do. So, employees and their managers are included in our definition of practitioners. In addition, our proposed approach involves reporting significant results with impact in mind, which also means that research results can be described in such a way as to be relevant for broader stakeholders (e.g., policy makers) and society at large. Our customer-centric approach to reporting research results involves the following interdependent and sequential steps:

1. Reporting statistical significance of results using a precise probability value between 0 and 1.00 (i.e., $0 < p < 1.0$) and an a priori significance level (i.e., α) that is set based on a specific research context and anticipated outcomes and not on arbitrary or conventional values (i.e., $\alpha = .05$ or $\alpha = .01$). As we will demonstrate later in this article, this type of reporting is different from the common practice of reporting whether a result has reached an arbitrary probability value or not, without reporting the actual probability value obtained (e.g., $p < .01$ or $p < .05$). As we will also demonstrate later in this article, current practices in the field of management, compared to other business fields, are particularly troublesome. This first step is required to understand whether a particular result is robust or can be ruled out as a consequence of chance alone.
2. Reporting effect size estimates that describe the strength of the relationship and/or effect found. As we will demonstrate later in this article, this type of reporting is now seen in some journals but has not yet been adopted as a matter of policy in the field of management and other business fields including marketing, management information systems, accounting, and finance. Moreover, we will discuss how the common practice of using one-size-fits-all cutoffs for determining what effect size is small (i.e., unimportant) or large (i.e., important) is misguided. This second step is required to understand the extent to which antecedent variables explain or predict outcome variables.
3. Reporting the practical significance of the results via a qualitative study that describes the importance of the results for specific stakeholder groups in specific contexts. As we will describe later in this article, this type of reporting is quite different from reporting an estimate of effect size, which is often confused as an estimate of practical significance. This step involves determining the meaning of a study's results from the perspective of practitioners (i.e., using their own language and context). Although we do not advocate that this third step be mandatory in every study, we suggest it is useful in enhancing our understanding regarding the extent to which results actually matter (i.e., their impact for various stakeholder groups in various contexts), particularly for new areas of research.

The first two steps are quantitative in nature and the third step is qualitative. Although the first two steps are not likely to have direct appeal to a practitioner audience, they are logically necessary before the third step can be accomplished. In other words, we first need to know whether the obtained results can be explained by chance alone (i.e., statistical significance) and the size of the obtained effect (i.e., effect size computation) before we can understand the practical meaning and significance of our results (i.e., practical significance). Again, the theme that underlies our customer-centric proposal is that academics need to report results in a way that is most useful and makes sense to both the academics and practitioners who are the consumers of our research results. Next, we provide a description of each of the three steps including problems in how research results are currently reported and their associated solutions.

Step 1: Reporting Statistical Significance

Problems inherent in the conventional way statistical significance has been used and reported for hypothesis testing have been discussed not only in management (e.g., Cashen & Geiger, 2004; Cortina & Folger, 1998; Sauley & Bedeian, 1989) but also in several other related fields including finance (e.g., Gómez-Bezares & Gómez-Bezares, 2006), psychology (e.g., Shrout, 1997), marketing (e.g., Iacobucci, 2005; Sawyer & Peter, 1983), economics (e.g., McCloskey, 1985), sociology (e.g., Henkel, 1976), education (e.g., Fan, 2001), and public administration (e.g., Cascio & Aguinis, 2001). According to Hunter (1997), problems related to the current conventions of statistical significance are serious enough to be labeled a "disaster" and are responsible for delaying progress by decades in some research areas. Although various solutions to these problems have been proposed, few changes have occurred in the usage of statistical significance in many fields (Thompson, 1999a, 1999b; Vacha-Haase, Nilsson, Reetz, Lance, & Thompson, 2000).

Problem #1: Conventional α Values are Arbitrary and Misleading

The current convention in hypothesis testing is that researchers choose an a priori Type I error rate (α: probability that the null hypothesis is rejected when it is actually true), compute the appropriate test statistic for the data in hand (e.g., t, F), and then obtain a probability value (p) associated with the test statistic that provides an estimate of the likelihood of obtaining these data if the null hypothesis is true. If $p < \alpha$, then the null hypothesis, usually stating that there is no effect or relationship, is rejected.

The current zeitgeist in the field of management is to use α values of .10 (i.e., results are said to be "marginally significant"), .05 (i.e., results are "significant"), .01 (i.e., results are "highly significant"), or .001 (i.e., results are "very highly significant"; Aguinis & Harden, 2009; Krueger, 1998). For example, if $\alpha = .05$, one would consider an obtained p value lower than .05 to be a significant result because, if the null were true, the probability of obtaining these data is less than 5% and, hence, the null is considered to be false.

How pervasive are these conventional values for α in management and related fields? To answer this question, we conducted a content analysis of each of the articles published from January 2002 to December 2006 that reported a test of statistical significance in 13 prestigious journals in 5 business disciplines:

- Accounting: *The Accounting Review, Journal of Accounting and Economics*, and *Journal of Accounting Research*
- Finance: *Journal of Finance* and *Journal of Financial Economics*
- Management: *Administrative Science Quarterly, Academy of Management Journal*, and *Strategic Management Journal*
- Marketing: *Journal of Consumer Research, Journal of Marketing*, and *Journal of Marketing Research*

Table 1. Use of Type I Error Rates (α) in 13 Business Journals from 2002–2006

Disciplines and Journals	Percentage Using a Conventional α	Percentage Using $\alpha = .10$	Percentage Using $\alpha = .05$	Percentage Using $\alpha = .01$	Percentage Using $\alpha = .001$
Accounting					
The Accounting Review	70	41	63	61	22
Journal of Accounting and Economics	86	56	79	78	5
Journal of Accounting Research	79	48s	68	72	43
Mean	76	46	68	68	16
Finance					
Journal of Finance	76	50	69	62	4
Journal of Financial Economics	78	58	73	70	0
Mean	77	54	71	66	3
Management					
Administrative Science Quarterly	100	43	98	93	50
Academy of Management Journal	99	51	96	93	50
Strategic Management Journal	99	66	94	92	48
Mean	99	57	96	93	49
Marketing					
Journal of Consumer Research	94	21	80	42	19
Journal of Marketing	92	42	82	73	30
Journal of Marketing Research	95	38	83	67	17
Mean	93	35	82	63	24
Management information systems					
Information Systems Research	98	22	87	81	35
Management Information Systems Quarterly	91	36	71	76	38
Mean	95	29	79	79	37
Grand mean	87	49	80	75	25

Note: Conventional $\alpha = .10$, .05, .01, or .001.

- Management information systems: *Information Systems Research* and *Management Information Systems Quarterly*

We chose to content analyze each of the above journals because they are often classified as being premier journals in their respective disciplines (Trieschmann, Dennis, Northcraft, & Niemi, 2000; Werner & Brouthers, 2002). Four coders were involved in the content analysis of the articles regarding number of statistical tests reported in the tables, number of tests for which authors used conventional α values of .10, .05, .01, or .001; other cutoffs being used; and number of tests for which authors reported the actual observed p value. Following recommendations by Duriau, Reger, and Pfarrer (2007), each article was coded by more than one coder; specifically, each was coded independently by two of the four coders. The coders agreed on 65,869 of 68,652 values or 95.9%. When there was a disagreement, the two coders discussed the differences and came to a consensus as to the correct value. Table 1 shows results regarding the use of conventional values for α.

Results displayed in Table 1 indicate that 99% of articles in the field of management use conventional values for α. Management information systems and marketing are close given that 95% and

93% of articles, respectively, use conventional values. Relatively fewer finance (77%) and accounting (76%) articles use conventional values. The table also shows the percentage of articles that reported a cutoff of a specific value at least once. In terms of specific values, 96% of management articles used $\alpha = .05$, whereas this value is 82% for marketing, 79% for management information systems, 71% for finance, and 68% for accounting. Based on the results shown in Table 1, the conclusion is that the vast majority of articles published in some of the top journals in business use conventional values for α, but management is clearly the discipline that has embraced these conventional values most strongly, particularly the .05 value.

The α values that are so frequently chosen by business researchers are arbitrary. This conclusion becomes obvious after a close examination of the origin of these conventional cutoff scores (Little, 2001). For instance, the $\alpha = .05$ value originated with Fisher (1925, p. 47) who wrote that "The value for which $p = .05$ or 1 in 20 is 1.96 or nearly 2; it is convenient to take this point as the limit in judging whether a deviation is to be considered significant or not. Deviations exceeding twice the standard deviation are thus formally regarded as significant." According to Gigerenzer (1998, p. 200), one of the major reasons why Fisher chose the 5% cutoff was that "he had no table for other significance levels, partly because his professional enemy, Karl Pearson, refused to let him reprint the tables Pearson had." As noted sarcastically by Rosnow and Rosenthal (1989, p. 1277), "Surely, God loves the .06 nearly as much as the .05."

In addition to being arbitrary, the conventional values for α are misleading. For example, when $\alpha = .05$, obtaining a $p = .0499$ would lead to the conclusion that the null hypothesis should be rejected (i.e., there is an effect or relationship), whereas obtaining a $p = .0501$ would lead to the conclusion that the null hypothesis should be retained (i.e., there is no effect or relationship). In other words, the conclusion of whether or not a variable is related to another, or whether a manipulation has an effect or not, can be completely changed by a miniscule difference in the observed p value. Huberty (1993) refers to this approach as being "mindless" and based on "no strong logical reason." Using conventional cutoffs provide the illusion of an objective value against which to evaluate the veracity of the null hypothesis rather than using a researcher's subjective judgment and, therefore, "makes life tidier" (Hubbard & Ryan, 2000). This convention, too, can lead to fundamental changes in a reader's interpretation of findings as well as in the application and future development of those findings.

A second reason why conventional cutoffs are misleading is that they are not based on a consideration of the relative seriousness of making at Type I versus a Type II error. As noted earlier, Type I error is the probability of rejecting the null hypothesis erroneously. In other words, concluding that there is an effect or relationship when in fact this is not true. The other side of the coin of a Type I error is a Type II error (β): the probability of retaining the null hypothesis erroneously. In other words, concluding the effect or relationship does not exist when in fact it does. Type I error is inversely related to Type II error. So, for the sake of the illustration, if we decided that we absolutely do not want to make a Type I error because we do not want to claim a finding that is not true and therefore set α at .000001, if we make an error, most likely we will end up making a Type II error. That is, we will not reject the null hypothesis even if the null hypothesis is false. Alternatively, if we decide that we can relax the Type I error rate because we want to make sure we detect an effect if it exists and therefore set α at .80, if we make an error most likely, we will end up making a Type I error (i.e., there is an 80% chance we will "find" an effect or relationship that does not exist). How many of the authors of the 96% of articles published in *Academy of Management Journal, Administrative Science Quarterly*, and *Strategic Management Journal* actually made a conscious decision that a 5% chance of finding a "false" effect was an acceptable risk? Moreover, how many of these authors actually computed what was the probability of not detecting an effect (Type II error rate) in their studies and how many of them were comfortable with this level of risk? Sedlmeier and Gigerenzer (1989) calculated that when researchers set a Type I error rate at .05, they usually have a Type II error of about

60%. This led Hunter (1997) to conclude that the use of significance testing is a disaster because the actual error rate in the social sciences is 60% and not 5%. How many of these authors were aware of their Type II error rates?

In short, although a conventional cutoff aids researchers in simplifying the decision of whether or not a null hypothesis should be rejected, these values are arbitrary and are based on the untenable assumption that the relative seriousness of making a Type I and Type II error is identical across research domains, outcomes, and contexts.

Solution for Problem #1: Use α Values Based on the Relative Seriousness of Type I Versus Type II Error

Our customer-centric approach to science implies that the customer, in this case most likely a researcher, should not use a conventional value for α. A researcher should not make a decision about whether a null hypothesis should be rejected using a Type I error of, say, 5%. As a customer, a researcher may have a preference for a Type I error rate of, say, 12%. In other words, a researcher may believe that it is less dangerous to find a ''false'' effect than to miss an effect that exists and, if found, may make a difference for society. The challenge is how to find the right balance between risking a Type I in relation to a Type II error. In a customer-centric approach to science, researchers choose a rationale for this balance based on their assessment of the relative seriousness of making these two types of errors.

Consider the relative seriousness of making a Type I relative to a Type II error in the following two actual research situations. First, Zhang, Bartol, Smith, Pfarrar, and Khanin (2008) explored a relationship between certain managerial incentives and CEO earnings manipulation behaviors. In this case, a Type II error would be quite serious. Concluding that there was no such relationship, when in fact there is one, may lead some firms to introduce such incentives, given that they do not believe this would lead to the negative consequences. Furthermore, the finding would justify the incentive use to those that are already using them. The use of these incentives would then lead to greater earnings manipulation behaviors, resulting in some firms experiencing severe declines in their stock price, exposure to reputational damage, top management turnover, possible bankruptcies, and loss of general investor confidence (Zhang et al., 2008). These costs are very high to the firms, investors, and society as a whole. On the other hand, a Type I error, where certain incentives are believed to lead to earnings manipulations, when in fact they do not, would cause some firms to not use those incentives, which may or may not have some negative consequences, depending on the other effects of the incentive. In this case, a Type II error would be more serious than a Type I error.

In a second example also based on an actual research situation, Miller, Fern, and Cardinal (2007) examined the relationship between the use of interdivisional knowledge and invention impact. In this case, a Type I error would be more severe. Concluding that there is such a relationship, when in fact there is not one, would lead to some firms investing a great deal of money, time, and other resources into facilitating knowledge transfer across divisions for no resultant gain. On the other hand, a Type II error would lead to some opportunity costs for firms because they did not make investments in interdivisional knowledge transfers that could have lead to greater invention impact. In this case, a Type I error would be more serious than a Type II error.

Murphy and Myors (1998) suggested a useful way to weigh the pros and cons of increasing the Type I error rate for a specific research situation. The appropriate balance between Type I and Type II error rates can be achieved by using a preset Type I error rate that takes into account the desired relative seriousness (DRS) of making a Type I versus a Type II error. Instead of increasing α to any arbitrary value, researchers can make a more informed decision regarding the specific value to give to α. Once the DRS of making a Type I versus a Type II error has been established, we can proceed to compute the Type I error rate.

Note that deciding on an appropriate DRS value could be seen as an arbitrary process. Thus, this decision should be well thought out and argued and not determined by fiat. Consider the following illustration described by Aguinis (2004, pp. 86-87). A researcher is interested in testing the hypothesis that the effectiveness of a training program for unemployed individuals varies by region such that the training program is more effective in regions where the unemployment rate is higher than 6%. Assume this researcher decides that the probability of making a Type II error (i.e., β, incorrectly concluding that unemployment rate in a region is not a moderator) should not be greater than .15. The researcher also decides that the seriousness of making a Type I error (i.e., incorrectly concluding that percentage of unemployment in a region is a moderator) is twice as serious as making a Type II error (i.e., DRS = 2). Assume the researcher makes the decision that DRS = 2 because a Type I error means that different versions of the training program would be needlessly developed for various regions, and this would represent a waste of the limited resources available. The desired preset Type I error can be computed as follows (Murphy & Myors, 1998):

$$\alpha_{\text{desired}} = \left[\frac{p(H_1)\beta}{1 - p(H_1)}\right]\left(\frac{1}{\text{DRS}}\right) \tag{1}$$

where $p\,(H_1)$ is the estimated probability that the alternative hypothesis is true (i.e., there is a moderating effect), β is the Type II error rate, and DRS is a judgment of the seriousness of a Type I error vis-à-vis the seriousness of a Type II error.

For this example, assume that based on a strong theory-based rationale and previous experience with similar training programs, the researcher estimates that the probability that the moderator hypothesis is correct is $p\,(H_1) = .6$. Solving Equation 1 yields the following:

$$\alpha_{\text{desired}} = \left[\frac{(.6)(.15)}{1 - .6}\right]\left(\frac{1}{2}\right) = .11. \tag{2}$$

Thus, in this particular example, using a nominal Type I error rate of .11 would yield the desired level of balance between Type I and Type II statistical errors.

Implementing this procedure for choosing the specific a priori Type I error rate provides a more informed and better justification than using any arbitrary value such as .05 or .01 without carefully considering the trade-offs and consequences of this choice. Moreover, in our customer-centric approach, each researcher (including those who are reading a published article or report) can set his or her own α level before making a decision of whether a particular null hypothesis should be rejected.

Problem #2: Reporting Crippled p Values

A second problem related to the reporting of statistical significance is that researchers often note whether an observed *p* value is below a conventional value instead of reporting the actual observed *p* value. In essence, researchers are crippling a continuous scale ranging from 0 to 1.0 and dichotomizing a continuous variable (Aguinis, Pierce, & Culpepper, **IN PRESS**). For example, if an author reports whether $p < .05$, then a significance test with a *p* value of .000001 is reported exactly the same as a significance test with a *p* value of .049. The problem with this practice is that valuable information is lost (Cohen, 1983). Knowing that the probability that the observed value would have occurred if the null hypothesis were true is 4.9% substantially differs from that probability being 1 in a million. Thus, by only reporting *p* values relative to cutoffs, important information, which can fundamentally change a reader's (i.e., customer) interpretation of a study's findings is lost.

How frequently do researchers in management report observed *p* values in reference to certain cutoffs instead of reporting actual *p* values? A second facet of our content analysis described in the previous section provides an answer to this question. Table 2 shows results of our analysis of the same 13 business journals. In this analysis, we used a value of 1 if an actual *p* value was reported in any given article in any of the tables and 0 if all *p* values were reported in reference to cutoffs. Also, we calculated the percentage of actual *p* values reported by dividing the actual number of *p* values reported by the total number of statistical significance tests reported in all the tables in each article.

Results displayed in Table 2 indicate that a miniscule percentage of articles in the three management journals we examined (i.e., 1.2%) reported actual *p* values rather than observed *p* values in relationship to a cutoff at least once. In contrast, this figure was 5.1% in management information systems, 6.1% in marketing, 37.7% in accounting, and 48.9% in finance. So, overall, the majority of articles in each of the five disciplines report observed *p* values in relationship to cutoffs and not the actual *p* values. However, the field of management is clearly ahead of all the others in terms of a preference for using cutoffs given that about only 1% of all articles reported actual observed *p* values even once. Consistent with these results, Table 2 also shows that the percentage of actual *p* values reported relative to the total number of tests of statistical significance reported is also smaller for management (i.e., .8%) than for all of the other disciplines (i.e., management information systems: 5.5%, marketing: 7.2%, finance: 23.3%, and accounting: 24.1%).

The irony of this situation is that Fisher, who is believed to be the originator of the $\alpha = .05$ cutoff (Fisher, 1925), toward the end of his career suggested that, in addition to using cutoffs, authors should also report actual *p* values (Fisher, 1959). This position was echoed by practically every other eminent statistician of the time (Gigerenzer, 1998). However, as indicated by our content analysis, this recommendation has been ignored in management and other business fields as well.

Why is it that authors do not report actual observed *p* values? One reason is that, decades ago, the lack of computers made it practically impossible to obtain actual *p* values. Instead, researchers had no choice but to consult available tables showing values for statistical tests (i.e., *t, F*) that were large enough to reject the null hypothesis at either the .01 or .05 level. It is very likely that implementing this practice over decades resulted in the institutionalization of this convention because several generations of researchers were trained this way. Over time, using the .01 and .05 cutoffs without reporting the actual *p* values became the standard practice and researchers created their own methodological comfort zones. Change is difficult and researchers are not immune to resisting change as it relates to methodological practices (Aguinis, Pierce, Bosco, & Muslin, 2009). This is a likely reason why there is a "scientific community's persistence in the use of particular methods" (Podsakoff & Dalton, 1987, p. 433). A second related reason for not reporting actual *p* values is that, as a consequence of the institutionalization of reporting results in relationship to certain cutoffs only, some journal editors require that authors choose a specific value for α and use it for all tests of significance in a given study. For example, in an annual update sent by the past editor of the *Journal of Applied Psychology* to all editorial board members, he indicated that "authors should stick with the conventional levels of $p < .05$ or .01" so as to avoid statements about results approaching significance or being marginally significant (e.g., S. Zedeck, personal communication, January 18, 2005). A third reason, which is related to why editors and reviewers may require that authors use the conventional cutoffs and not report actual *p* values, is a generalized misunderstanding about significance testing and the meaning of the observed *p* value. As also noted in the yearly update by the past editor of the *Journal of Applied Psychology*, "statements such as the 'results are marginally significant' or that the 'results approach significance' are not appropriate" (S. Zedeck, personal communication, January 18, 2005). Many researchers have noted that significance testing is abused and misused (e.g., Cohen, 1994; Schmidt, 1996) and, therefore, there seems to be a "human factors problem" (Tryon, 2001, p. 371). Significance testing allows researchers to infer whether the data obtained are unlikely assuming a true null hypothesis. On the other hand, significance testing is used incorrectly when conclusions are made regarding the magnitude of the effect and, for example, a

Table 2. Statistical Significance Test Reports in 13 Business Journals From 2002–2006

Disciplines and Journals	Number of Articles 2002–2006	Number of Statistical Significance Tests Reported	Number of Statistical Significance Tests Reporting p Values in Reference to a Cutoff	Percentage of Articles Using Actual p Values at Least Once	Percentage of Actual p Values Relative to the Total Number of Tests Reported in Each Article
Accounting					
The Accounting Review	185	127.4	89.6	37.8	29.7
Journal of Accounting and Economics	100	176.9	147.5	29.4	16.6
Journal of Accounting Research	103	189.9	144.3	45.6	24.0
Mean	388	156.7	119.0	37.7	24.1
Finance					
Journal of Finance	270	209.2	157.7	51.5	24.6
Journal of Financial Economics	237	211.1	165.3	45.8	21.7
Mean	507	210.1	161.2	48.9	23.3
Management					
Administrative Science Quarterly	56	232.9	232.9	0.0	0.0
Academy of Management Journal	253	158.4	157.3	1.1	0.7
Strategic Management Journal	265	138.1	136.5	1.6	1.2
Mean	574	156.3	155.1	1.2	0.8
Marketing					
Journal of Consumer Research	80	45.2	41.5	3.7	8.2
Journal of Marketing	142	106.6	98.1	8.5	8.0
Journal of Marketing Research	87	87.5	82.9	4.6	5.3
Mean	309	85.3	79.2	6.1	7.2
Management information systems					
Information Systems Research	54	74.6	68.6	6.0	8.0
Management Information Systems Quarterly	55	111.2	106.9	4.3	3.9
Mean	109	93.0	87.9	5.1	5.5
Total *N*	1,887				
Grand mean		155.6	133.0	22.6	14.5

statistically significant result at the .01 level is interpreted as a larger effect than a result at the .05 level (Aguinis, 2004).

Solution for Problem #2: Report Precise p Values

Our customer-centric approach to science suggests that researchers should report the actual observed p values so that the customers in this case (i.e., readers of their article) have this information available. Because actual p values are now available in most statistical software packages, there is no impediment to including them in research reports. Thus, authors should report the precise p value obtained (e.g., $p = .052$) and readers can decide for themselves whether the findings should be considered "statistically significant." Reporting should include a maximum of two or three decimal places because using more decimal places is unnecessary (Bedeian, Sturman, & Streiner, **IN PRESS**; Cohen, 1990).

By reporting actual p values, valuable information is not lost by dichotomizing continuous variables (Aguinis et al., **IN PRESS**). Readers know whether the p value is .049 or .000001, and not just that it is less than .05. Second, the arbitrary cutoffs no longer exist; thus, .049 and .051 are treated as what they are and are not arbitrarily separated as diametrically opposed findings. Reporting actual p values helps solve an additional problem of statistical significance mentioned earlier: frequently occurring Type II errors. Many results are interpreted as being nonsignificant because they do not reach a conventional level for α. For example, if $\alpha = .05$, a researcher may conclude that the finding is statistically nonsignificant and report $p > .05$ and no further information is provided. Given a customer-centric approach and the previous discussion about setting a value for α that considers the relative trade-offs between Type I and Type II errors, a reader may decide that if $\alpha = .08$, the null hypothesis should be rejected. However, this reader is not able to make a decision regarding the statistical significance of the result in the absence of the actual p value observed.

In summary, our 5-year content analysis demonstrated that the vast majority of articles published in some of the most prestigious journals in five business disciplines use conventional cutoff values for deciding whether a null hypothesis should be rejected and do not report the actual observed probability of obtaining the data in hand given a true null hypothesis. A noteworthy finding is that these practices are more pervasive in the field of management than in management information systems, marketing, accounting, and finance. Specifically, 99% of management articles use conventional definitions for α and only 1.2% of articles report actual observed p values even once. Problems related to these methodological practices include a lack of consideration of trade-offs between falsely rejecting the null hypothesis (Type I error) and not detecting an existing effect (Type II error) and loss of information by dichotomizing a continuous variable (i.e., p values). A customer-centric approach to reporting research results argues for providing information to readers so they can make their own assessment regarding the rejection of the null hypothesis. Thus, we described a procedure for setting an α level that is based on a consideration of the relative seriousness of making a Type I versus a Type II error. Customers of research (i.e., readers of journal articles) can use this procedure to set their own independent α level. Also, we described the advantages of reporting actual p values so that readers can judge the statistical significance of the results using their own α level.

Step 2: Reporting Estimated Effect Magnitude

Reporting statistical significance says nothing about the strength of the effect (in the case of experiments in which causality can be inferred) or relationship (in the case of nonexperimental designs in which only covariation between variables can be inferred). The only conclusion we derive from a statistical significance test is whether the data obtained are unlikely given a true null hypothesis. Thus, the second step in our customer-centric approach includes reporting an estimate of the magnitude of the effect or relationship. In the remainder of the article, we refer to this as the estimated effect magnitude (cf. Kirk, 1996).

Problem #3: Confusion Between Statistical Significance and Magnitude of the Effect or Relationship

As noted earlier, there is a human factors problem in the scientific community because many view the rejection of the null hypothesis as the ultimate goal of any empirical study. For example, Yates, a contemporary of Fisher noted that "the emphasis on tests of significance, and the consideration of the results of each experiment in isolation, have had the unfortunate consequence that scientific workers often have regarded the execution of a test of significance on an experiment as the ultimate objective" (1951, pp. 32-33).

Unfortunately, there are important problems with viewing the rejection of the null hypothesis as the ultimate goal (e.g., Cohen, 1990, 1994; Kirk, 1996; Schmidt, 1996). First, rejecting the null hypothesis does not tell researchers what they really want to know. What researchers want to know is whether the null hypothesis is true given the data in hand—pH_o/D. However, null hypothesis significance testing tells us the probability of obtaining the data in hand (or more extreme data) if the null hypothesis is true—pD/H_o. Unfortunately, $pH_o/\mathrm{D} \neq p\mathrm{D}/H_o$. As noted by Cohen (1994, p. 997), a test of statistical significance "does not tell us what we want to know, and we so much want to know what we want to know that, out of desperation, we nevertheless believe that it does!"

Second, most null hypotheses are false at some level. As noted by Tukey (1991, p. 100), "the effects of A and B are always different—in some decimal place—for any A and B. Thus asking 'Are the effects different' is foolish." From this perspective, finding that the null hypothesis should be rejected simply means that sample size was large enough or, more generally, the research design was sufficiently adequate and had sufficient statistical power (i.e., probability of correctly rejecting the null hypothesis or 1—Type II error). Some equate smaller p values as though they represent greater (i.e., more meaningful) effects sizes. Note that the larger the effect in the population, the more likely it is that the null hypothesis will be rejected. In addition, the larger the sample size, the more likely it is that null hypothesis will be rejected, even if the population effect is miniscule (Aguinis & Harden, **IN PRESS**).

The confusion between statistical significance and magnitude of the effect also has implications for theory development. Statistical significance is a necessary but not a sufficient yardstick to evaluate the precision and accuracy of a theory. Statistical significance does not tell us the extent to which the variables we investigated are sufficient to explain a particular phenomenon satisfactorily. For example, statistical significance does not tell us the extent to which team collaboration is related to team performance. If the result is statistically significant, we conclude that these variables are related to each other, but we do not know to what extent. To do so, we need to compute estimates of the magnitude of their relationship.

Because of the need to go beyond null hypothesis significance testing, journals in several disciplines require that authors report not only the result of the significance test but also a measure of effect magnitude. For example, the 2001 edition of the American Psychological Association's Publication Manual notes that "it is almost always necessary to include some index of effect size or strength of relationship ... The general principle to be followed ... is to provide the reader not only with information about statistical significance, but also with enough information to assess the magnitude of the observed effect or relationship" (pp. 25-26). In addition, 24 journals in psychology and education, including the *Journal of Applied Psychology* and *Career Development Quarterly*, require or strongly recommend that authors report estimates of effect magnitude (Thompson, 2006).

How much attention do business journals give to effect magnitude reporting? To answer this question, in June 2008, we examined the submission guidelines for the same 13 business journals

Table 3. Formulae for Computing Common Effect Magnitude Estimates

Effect Magnitude Estimate	Formula	Comments
Group differences		
Cohen's *d*	$\frac{\bar{X}_{\text{Group A}}-\bar{X}_{\text{Group B}}}{SD}$	One of the most common and frequently used estimates of differences between two groups. It denotes the difference between mean scores across groups in standard deviation units. *SD* is the common standard deviation. A value of 0 means no difference between the groups.
Hedges's *g*	$\frac{\bar{X}_{\text{Group A}}-\bar{X}_{\text{Group B}}}{SD_{\text{pooled}}}$	This is a "corrected" *d* estimate based on a pooled *SD* across the groups.
Variance explained		
R^2	*SS* regression/ *SS* total	One of the most common and frequently used estimates of variance explained. It denotes the proportion of variance in a criterion explained by the predictors given a linear predictor-criterion relationship. A value of 0 means no variance explained.
Adjusted R^2	$1-(1-R^2)\frac{N-1}{N-k-1}$	This is a "corrected" R^2 estimate that considers capitalization on chance because of a small sample *N* and large number of predictors *k*.

Note: SS = Sum of squares. Additional effect magnitude estimates for group differences are, *g*', *h*, *q*, *w*, Π, and Φ; additional effect magnitude estimates for variance explained are r^2_{pb}, η^2, partial η^2, f^2, ω^2, and ϵ^2. For sources of formulae and computation details of each of these effect magnitude estimates, please see Kirk (1996); Pierce, Block, and Aguinis (2004); Snyder and Lawson (1993); and Vacha-Haase and Thompson (2004).

listed in Tables 1–2. Our review revealed that *none* of the author and/or submission guidelines for the 13 journals include any specific statements about the reporting of effect magnitude estimates. Thus, it seems that the business disciplines lag behind other fields in terms of formal journal guidelines regarding the need to report estimates of effect magnitude.

Solution for Problem #3: Report Magnitude of Effect

The convention in reporting statistical significance is a serious problem—so serious that some have suggested the elimination of the use of statistical significance altogether. One journal, *The American Journal of Public Health*, once banned statistical significance tests completely (although the ban only lasted 2 years). However, it is generally acknowledged that the elimination of statistical significance will be difficult, if not impossible, to achieve for three reasons. First, it does provide the useful information that chance sampling influences is not a likely explanation of findings (Chow, 1998). Second, many of the problems of significance testing can be solved by simple, incremental changes in design methods, reporting, and interpretation of statistical significance (Abelson, 1997). Third, its usage has been ingrained as a staple of hypothesis testing (Hubbard & Ryan, 2000), and there is a "scientific community's persistence in the use of particular methods" (Podsakoff & Dalton, 1987, p. 433).

Most critics of statistical significance believe that reporting effect sizes can mitigate some of the problems associated with sole reliance on null hypothesis significance testing. Several sources exist that provide detailed guidelines and formulae regarding the computation of various types of effect magnitude estimates (e.g., Huberty, 2002; Kirk, 1996; Kline, 2004; Sink & Stroh, 2006; Snyder & Lawson, 1993; Vacha-Haase & Thompson, 2004).

Table 3 includes formulae for some of the most common types of effect magnitude estimates, which most of the common commercially available statistical analysis software packages (e.g., SPSS, SAS) include as part of the output. As shown in Table 3, effect magnitude estimates can be classified into two types as follows: (a) group difference estimates and (b) variance-explained estimates. What these two types of effect magnitude estimates have in common is that they are standardized, independent of the metric used in any particular study (e.g., economic performance measured in US dollars or Euros, job satisfaction measured on a 3- or 7-point Likert-type scale), and can therefore be compared across studies. Group difference estimates are useful when the null hypothesis refers to differences in mean scores across groups. For example, a researcher may test the null hypothesis that a Web-based training program will decrease biases in job analysis ratings and may want to report the magnitude of the difference in average ratings between trained and nontrained groups (Aguinis, Mazurkiewicz, & Heggestad, 2009). Variance accounted for estimates are usually reported in a squared metric (e.g., R^2) and are used when the null hypothesis refers to relationships among variables and are interpreted as the proportion of variance in one variable explained by another. For example, if the relationship between cognitive abilities and job performance is $r = .35$, then cognitive abilities explains 12.25% of variance in performance scores (Aguinis & Smith, 2007).

Note that the formulae included in Table 3 allow researchers to compute sample-based estimates of effect magnitude. After all, researchers usually have a sample in hand only and not the entire population of scores. So, sampling error affects the accuracy of these estimates, and when samples are small and regression models include many predictors, there is more capitalization on chance and, hence, effect magnitude estimates are inflated. Thus, some researchers recommend "correcting" observed estimates and some of these corrections are included in Table 3 as well. Note that the major commercially available statistical packages include some of these corrected estimates and label them using the qualifier "adjusted" (e.g., adjusted R^2).

In addition to sample size and number of predictors, the resulting estimates are also affected by the reliability of the measurement instruments used (i.e., lower reliability leads to smaller estimates), the type of scales used to measure each of the variables (i.e., fewer scale anchors lead to smaller estimates), heterogeneity of the sample used, variance in the scores (i.e., less heterogeneity and variance lead to smaller estimates), and research design (i.e., nonexperimental designs usually lead to smaller estimates; Aguinis et al., **IN PRESS**; Vacha-Haase & Thompson, 2004). Thus, it is important the researchers not only include information on which specific estimate has been computed but also whether it is an adjusted estimate or not, and also a confidence interval around the estimate (Algina & Keselman, 2003; Cumming & Finch, 2001).

In sum, a customer-centric approach to reporting research results involves reporting an estimate of the effect magnitude. Members of the academic target audience would benefit from knowing not only the probability that the data in hand would be obtained given a true null hypothesis but also the estimated magnitude of the effect or relationship. This type of reporting has important implications for theory building and for evaluating the quality of a theory (Bacharach, 1989). For example, knowing that a certain set of variables explains only 5% of variance in an outcome (e.g., probability of a merger) leads to the conclusion that a particular theoretical model is underspecified and more and better predictors should be investigated. On the other hand, knowing that a certain set of variables explains 80% of variance in an outcome would lead to the conclusion that the conceptual model is quite good. Once a researcher has established that a finding is not likely to be because of chance alone and the size of a particular effect, it is time to discuss its practical significance.

Step 3: Reporting Practical Significance

Although this third step in our customer-centric approach is the one that addresses the specific needs of practitioners, this third step is not possible unless Steps 1 and 2 are present. We first need to understand that it is highly improbable that we would obtain the data we did if the null hypothesis were true (i.e., Step 1), and we also need to know the estimated effect magnitude (i.e., Step 2). Once we have completed these two steps, we can attempt to understand the practical significance of our results.

Problem #4: Confusion Between Magnitude of the Effect and Practical Significance

Calls for reporting the practical significance of research results are certainly not new. Cohen (1977) has been a pioneer in alerting researchers that using null hypothesis significance testing is not sufficient. Several such calls have been made over the past few decades such as Kirk's warning that "we must get over our obsession with null hypothesis significance tests and focus on the practical significance of our data" (1996, p. 757).

However, there is confusion in the literature regarding the meaning of practical significance, and there is the general belief that reporting an estimate of the magnitude of the effect is the same as reporting a study's practical significance. Since Fisher (1925) proposed the use of eta squared (η^2), many researchers have issued calls for reporting effect magnitude estimates as if it were a direct indicator of a study's practical significance. However, assume that a preemployment test is correlated with job performance at $r = .30$. How can a practitioner understand the practical significance of a preemployment test that explains $.3 \times .3 = 9\%$ of the variance in future performance scores? Is this result practically significant? Should this practitioner recommend the use of the new preemployment test to prescreen job applicants in the future? Will using this preemployment test lead to better hiring decisions? How much better? If we decide to use this test, what are the consequences of not being able to explain the remaining 91% of variance in performance scores? In general, as noted by Cortina and Landis (2009), "the myth is that general statements such 'X and Y are strongly related' are observation sentences vis-à-vis particular effect sizes irrespective of the contexts in which the values were generated" (p. 305).

To answer the above questions and understand the practical significance of results, researchers have used yardsticks to evaluate whether a particular effect magnitude is sufficiently important. Thus, researchers tend to use descriptive labels such as "strong" and "weak" in referring to effect sizes (Cortina & Landis, 2009). The most well-known such taxonomy was suggested originally by Cohen (1962; see Hemphill, 2003, for a taxonomy specific to the area of psychological assessment and treatment). For example, researchers have used the values of $r = .10$ as a "small" effect, $r = .30$ as a "medium" effect, and $r = .50$ as a "large" effect (for examples of the use of the small, medium, and large values proposed by Cohen, see Aguinis & Stone-Romero, 1997; Brews & Tucci, 2004; Kim, Hoskisson, & Wan, 2004; Morgeson & Campion, 2002; Raver & Gelfand, 2005).

Where did these values come from? How did Cohen determine that, for example, $r = .10$ is "small"? Aguinis, Beaty, Boik, and Pierce (2005) reviewed the history behind these values. Cohen's first published description of specific magnitudes for effects appeared in his 1962 *Journal of Abnormal and Social Psychology* article. In this article, Cohen reported results of a review and content analysis of articles published in the 1960 volume of this same journal. In the Method section of his article, when describing the effect sizes he used for his power analysis, Cohen (1962, p. 147) noted that "the level of average population proportion at which the power of the test was computed was the average of the sample proportions found" and "the sample values were used to approximate the level of population correlation of the test." For the correlation coefficient, Cohen defined .40 as medium because this seems to have been close to the average observed value he found in his review. Then, he chose the value of .20 as small and .60 as large. In other words, Cohen's definitions of small, medium, and large effect sizes are based in part on observed values as reported in the articles published in the 1960 volume of *Journal of Abnormal and Social Psychology*, and in part on his own subjective opinion. A few years later, Cohen (1988) decided to lower these values to .10 (small), .30 (medium), and .50 (large) because the originally defined values seemed a bit too high. Given the history behind the conventional values for small, medium, and large effects, it is not surprising that Cohen (1992) himself acknowledged that these definitions "were made subjectively" (p. 156).

In short, the effect size values that many management researchers use as benchmarks to understand whether their results are practically significant were originally derived from the observed effect sized published more than 40 years ago in a journal in a completely different discipline. This is why Thompson noted that if we use these values to determine practical significance "with the same rigidity that the $\alpha = .05$ criterion has been used, we would merely be being stupid in a new metric" (2002, p. 68).

Solution for Problem #4: Report Practical Significance

As noted earlier, merely reporting effect magnitude estimates is not sufficient to describe a study's practical significance. Also, using one-size-fits all benchmark values to try to understand whether an effect is practically significant may be convenient for academics and may convey the impression of objectivity and standardization. However, this information is not necessarily useful to those who make recommendations about the management or development of people in organizational settings or give advice to those who do.

How often do management researchers discuss the issue of practical significance? We conducted a literature search for the phrase "practical significance" in the Abstracts of the *Academy of Management Journal, Academy of Management Perspectives, Academy of Management Review*, and *Academy of Management Conference Proceedings*. The result was a surprisingly low hit of eight and the majority of the sources were from the Proceedings. Moreover, none of the eight articles described how to assess practical significance and, instead, described how a certain set of results had "practical significance."

For a study to report results in a way that is customer-centric in terms of the needs of practitioners, it must contextualize and properly interpret the results and conclusions (cf. Cortina & Landis, 2009). Practical significance involves a value judgment made by the consumer of research about the implications of a set of results and the consequences of a particular decision (Vaske, Gliner, & Morgan, 2002). Practical significance means asking the question of whether the results are noteworthy and whether they are big enough to matter (Armstrong & Henson, 2004). In other words, "data should be described in a way that fits with how practitioners would describe the situation being addressed in the study" (Baldridge et al., 2004, p. 1073). Including a token section in an article's Discussion section describing "implications for practice" has proven ineffective at bridging the science- practice gap and demonstrating the significance of research results in the eyes of practitioners. Also, including brief commentaries by individual executives as follow-ups to articles written by academics has not been successful, as demonstrated by the abysmally low rate at which practitioners read the *Academy of Management Executive* (now *Academy of Management Perspectives*) despite the fact that this journal has implemented this practice for many years (i.e., average frequency of reading by practitioners is 1.1 on a scale ranging from 1: *never* to 5: *always*; Rynes et al., 2002).

To demonstrate a study's practical significance, there is a need to describe results in a way that makes sense for practitioners. We suggest that this can be achieved by including practitioners in each research project as part of a qualitative study. Specifically, we suggest that the practical importance of results be reported using practitioners' own discourse. In essence, we are proposing a turn to language and interpretation (Willig & Stainton-Rogers, 2008) so that the voice of practitioners is heard. Qualitative methodology is particularly appropriate because its goal is to understand and describe phenomena. Also, qualitative research gives voice to the participants and places importance on their understanding and interpretation of a given research study (e.g., Gilligan, 1982). Note that we are not suggesting that practitioners write a section of the resulting manuscript or that practitioners conduct the study, but that they become participants in a qualitative study.

There are several types of qualitative methods that can be used to understand the extent to which a study's results are practically significant to practitioners (for recent reviews, see Green, Camilli, & Elmore, 2006; Willig & Stainton-Rogers, 2008). These methods include the case study, cross-case analysis, ethnography, field notes analysis, historical research, narrative inquiry, practitioner inquiry, conversation analysis, focus groups, discourse analysis, social representation analysis, and visual analysis. Because of space constraints, we describe three of these approaches as follows: ethnography, conversation analysis, and narrative inquiry. Our goal is not to provide a comprehensive review of qualitative methodology. Rather, our intention is to demonstrate how researchers can use qualitative methods to gather information regarding the practical significance of their results. In other words, using qualitative methods can help provide an answer to the question of the extent to which a particular set of results actually matter.

Ethnography. Ethnography is a type of analysis that assumes the existence of shared cultural meaning in a social group (for reviews, see Anderson-Levitt, 2006; Griffin & Bengry-Howell, 2008). It was originally developed in the field of anthropology and is also used in sociology and education. It is a type of method that attempts to uncover meaning from the perspective of those involved. It involves studying a phenomenon in its natural setting and involves collecting data using several types of instruments such as semi-structured interviews and diaries. Researchers spend time in the field working alongside participants so they can immerse themselves in their world. Researchers become familiar with the values, norms, and cultural environment of the participants. As noted by Steier (1991), an ethnography study has the goal to build bridges of understanding between groups of people who use different meanings and languages. In short, ethnography can be used to understand the meaning of particular research results for individuals, such as human resource managers, employees, or consultants.

In terms of implementation, an ethnographic study should be seen as a conversation with a purpose. Our goal is to understand, from the perspective of participants, the extent to which research results matter and why this is so. What are the reactions of the participants to the research results? What do the results mean to them? How would they use the results in their work environments? How would the results be beneficial?

Conversation analysis. Conversation analysis focuses on everyday conversation or institutional talk. Originally, the methodology was developed in the fields of psychology, sociology, linguistics, and communication studies. The talk data are usually not collected by the researcher directly but, instead, are collected by the participants themselves using audio or video recording technology (Wilkinson & Kitzinger, 2008). Conversation analysis focuses not only on what people say but also how they say it.

For example, Wilkinson and Kitzinger (2006) focused on what they labeled reaction tokens such as "Wow!," "My Goodness!," "Ooh!," and "That's amazing!" Wilkinson and Kitzinger (2006) wanted to answer the question of "what are these exclamatory imprecations doing?" Conversation analysis involved first noticing a conversational phenomenon of interest (i.e., these "reaction tokens"), assembling a preliminary collection of instances of the phenomenon, and finally identifying the largest and most important subset within the collection, analyzing the clearest, less transparent, and deviant cases (Wilkinson & Kitzinger, 2008).

Conversation analysis allows for "a clearly specified, yet nuanced definition of intersubjectivity, as depending on displayed understandings of prior talk" (Wilkinson & Kitzinger, 2008, p. 67). One great practical advantage of this method is that researchers can rely on pre-existing data derived from various institutional contexts.

Narrative inquiry. Narrative inquiry is a methodology used to think about experiences through the lens of stories and originates from the view that individuals have storied experiences (Connelly & Clandinin, 2006; Hiles & Čermák, 2008; Phillips, Sewell, & Jaynes, 2008). Data used in narrative inquiry are collected using personal journals, stories, photographs, artifacts, annals, chronologies, interviews, and conversations that are used to tell a narrative of people. Stated differently, these data are used to tell their stories. In this way, narrative inquiry can be used to tell a story of how and why a certain study result matters. We use storytelling to describe the practical significance of a result.

There are three important features of a narrative inquiry process (Connelly & Clandinin, 2006). First, the time dimension is important in narrative inquiry. A certain phenomenon is described in terms of its past, present, and future potential. Second, narrative inquiry pays attention to both personal and social conditions surrounding the phenomenon. Personal conditions include feelings, hopes, and desires of the individual and social conditions include the organizational environment and other social forces that form the individual's context. Third, it is important to identify the specific concrete, physical boundaries of place where the phenomenon takes place. In the interest of generalizability, many researchers do not pay attention to the place dimension, specifying location is crucial to understand the phenomenon of interest.

Discussion

There has been an ongoing concern for several decades that knowledge generated by management researchers is not relevant and does not affect the work of practitioners. Echoing Hambrick's (1994) concern that management researchers have minimal visibility and minimal impact, Latham (2007, p. 1031) issued a severe warning that "we, as applied scientists, exist largely for the purpose of communicating knowledge to one another. One might shudder if this were also true of another applied science, medicine." In spite of suggestions on how to bridge the science-practice gap, the gap remains and no noticeable improvements have taken place over the past two decades despite regular calls to narrow it and that the mission statements of our empirical journals include making a contribution to management practice (e.g., Welcome to the Academy, 2009).

We believe this is the right time to call for an important change in how academics report research results. We need to report results adopting a customer-centric approach that considers the needs of both academics and practitioners. This involves reporting on (a) whether the phenomenon is likely to exist (i.e., statistical significance), (b) how strong it is (i.e., effect size), and (c) the extent to which it matters, to whom, and in which context (i.e., practical significance). This type of reporting embraces the concepts of rigor, relevance, and practical impact. It is rigorous because it relies on statistical significance testing but does not adhere to arbitrary cutoffs for deciding whether a result is significant or not. Instead, it involves a process of carefully weighing costs and benefits of making a Type I versus a Type II error. Second, it is relevant because it involves reporting the precise p value and allows researchers to make their own decision in terms of whether a particular result is statistically significant given their own choice for a Type I error rate. It is also relevant because it involves reporting an estimate of the magnitude of the effect, which allows researchers to understand the accuracy, precision, and quality of particular theories. Third, it is impactful because it involves reporting results of a qualitative study to assess the practical significance of the results. Moreover, in anticipating the conduct of a qualitative study including practitioners to assess practical significance, researchers will be forced to think about impact before the study even begins. This point is related to the idea of beginning the research journey with a purpose in mind (Hakel, Sorcher, Beer, & Moses, 1982). Thus, our proposed approach goes beyond the assumption of a communication problem between scientists and practitioners and addresses the question of whether academics are studying issues that are relevant to practitioners.

Our proposed approach imposes additional demands on researchers. First, researchers need to think more carefully about the consequences of making a Type I versus a Type II error. Currently, the practice of using one-size-fits-all cutoffs for Type I error makes it easier and gives the illusion that we are using ‘‘objective’’ criteria. Second, researchers need to compute an estimate of the effect. Currently, none of the 13 journals in the field of business that we examined mentions this type of reporting as mandatory in the submission guidelines. Third, researchers, particularly in new areas of research, would need to conduct a qualitative study to gather some evidence regarding the practical significance of their findings. Currently, most journals include a token section describing ‘‘implications for practice,’’ but this is done from the perspective of researchers and not practitioners using practitioners’ language and context. In spite of this additional work, we believe the payoff can be great and elevate our profession from the category of cottage industry to a field that helps improve human well-being and society (cf. Cascio & Aguinis, 2008). After all, universities also need to behave in socially responsible ways, and if researchers conduct work and disseminate results in a way that is useful outside of academic circles, this goal would be accomplished (Cummings, 2007). As noted by management scholar William Starbuck, ‘‘People should do management research because they want to contribute to human welfare. Those who are professors of management are people of superior abilities and they should use these abilities for purposes greater than themselves . . . I also observe that many doctoral students and junior faculty are focusing on achieving social status and job security and are viewing research methods as tools to construct career success. Few of them seem apt to initiate or even to participate in significant reorientations’’ (Barnett, 2007, pp. 126-127).

Our suggestion about conducting a qualitative study is only one step toward the understanding of practical significance. As any other research effort, one qualitative research study will not be able to prove unequivocally the universal worth of certain results. Moreover, there is an ongoing debate regarding what exactly constitutes a high-quality qualitative study (e.g., Easterby-Smith, Golden-Biddle, & Locke, 2008; Pratt, 2008; Savall, Zardet, Bonnet, & Péron, 2008). Thus, replicability and transparency are crucial not only regarding statistical significance and effect magnitude estimation but also regarding practical significance. Do the results have the same value for consultants and first-line managers? Do the results have the same value for human resource managers in the pharmaceutical and the telecommunications industries? More than one qualitative study will be needed to answer these questions and establish the external validity of a study’s practical significance. This is particularly true in substantive areas where there are few qualitative studies or areas where researchers have limited practical experience.

We are aware that our approach will not be effective at bridging the science-practice divide if professional associations, reviewers, and journal editors do not require the new type of customer-centric reporting that we propose. Journal editors play a crucial role in establishing scientific norms given that journal publishing is a central activity in an academic’s professional life (Starbuck, Aguinis, Konrad, & Baruch, 2008). Change is difficult and it is likely that many professional associations, reviewers, and journal editors will see a customer-centric approach to reporting significant results as a threat to established norms (e.g., Bedeian, Van Fleet, & Hyman, 2009). We suggest evolutionary rather than revolutionary change in which there is a transition period during which the adoption of our proposed approach is slowly introduced as a recommendation but not a matter of mandatory policy. This type of approach to change is necessary given that our proposal involves a change in professional norms that is guaranteed to lead to some resistance (Gephart, 2004; Pratt, 2008). Nevertheless, it is painfully obvious that unless there is a tangible change in how we approach the reporting of our research, the science-practice gap will remain. It is time that we make a conscious and mindful decision to change the way we report our results so that our research begins to matter in the eyes of practitioners and society at large. As noted by Hambrick (1994), ‘‘it is time for us to matter.’’

Authors' Note

We thank Don Hambrick, Susan Jackson, Bob Keller, Jim Phillips, and Chuck Pierce for constructive comments on previous drafts. This research was conducted in part while Herman Aguinis held the Mehalchin Term Professorship in Management at the University of Colorado Denver and a visiting appointment at the University of Puerto Rico.

Declaration of Conflicting Interests

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

Funding

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

References

Abelson, R. P. (1997). On the surprising longevity of flogged horses: Why there is a case for the significance test. *Psychological Science*, *8*, 12-15.

Aguinis, H. (2004). *Regression analysis for categorical moderators.* New York: Guilford.

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology*, *90*, 94-107.

Aguinis, H., & Harden, E. E. (2009). Sample size rules of thumb: Evaluating three common practices. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in the organizational and social sciences* (pp. 269-288). New York: Routledge.

Aguinis, H., & Harden, E. E. (**IN PRESS**). Cautionary note on conveniently dismissing chi-square goodness-of-fit test results: Implications for strategic management research. In D. J. Ketchen & D. A. Bergh (Eds.), *Research methodology in strategy and management* (vol. 5). San Diego, CA: Elsevier.

Aguinis, H., Mazurkiewicz, M. D., & Heggestad, E. D. (2009). Using Web-based frame-of-reference training to decrease biases in personality-based job analysis: An experimental field study. *Personnel Psychology*, *62*, 405-438.

Aguinis, H., Pierce, C. A., Bosco, F. A., & Muslin, I. S. (2009). First decade of Organizational Research Methods: Trends in design, measurement, and data-analysis topics. *Organizational Research Methods*, *12*, 69-112.

Aguinis, H., Pierce, C. A., & Culpepper, S. A. (**IN PRESS**). Scale coarseness as a methodological artifact: Correcting correlation coefficients attenuated from using coarse scales. *Organizational Research Methods*.

Aguinis, H., & Smith, M. A. (2007). Understanding the impact of test validity and bias on selection errors and adverse impact in human resource selection. *Personnel Psychology*, *60*, 165-199.

Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology*, *82*, 192-206.

Algina, J., & Keselman, H. J. (2003). Approximate confidence intervals for effect sizes. *Educational and Psychological Measurement*, *63*, 537-553.

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

Anderson-Levitt, K. M. (2006). Ethnography. In J. L. Green, G. Camilli & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 279-295). Mahwah, NJ: Lawrence Erlbaum.

Anderson, N. (2007). The practitioner-researcher divide revisited: Strategic-level bridges and the roles of IWO psychologists. *Journal of Occupational and Organizational Psychology*, *80*, 175-183.

Armstrong, S. A., & Henson, R. K. (2004). Statistical and practical significance in the IJPT: A research review from 1993-2003. *International Journal of Play Therapy*, *13*, 9-30.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review*, *14*, 496-515.

Baldridge, D. C., Floyd, S. W., & Markóczy, L. (2004). Are managers from Mars and academicians from Venus? Toward an understanding of the relationship between academic quality and practical relevance. *Strategic Management Journal*, *25*, 1063-1074.

Barnett, M. L. (2007). (Un)learning and (mis)education through the eyes of Bill Starbuck: An interview with Pandora's playmate. *Academy of Management Learning and Education*, *6*, 114-127.

Bartunek, J. M. (2007). Academic-practitioner collaboration need not require joint or relevant research: Toward a relational scholarship of integration. *Academy of Management Journal*, *50*, 1323-1333.

Bedeian, A. G., Sturman, M. C., & Streiner, D. L. (**IN PRESS**). Decimal dust, significant digits, and the search for stars. *Organizational Research Methods*.

Bedeian, A. G., Van Fleet, D. D., & Hyman, H. H., III. (2009). "Circle the wagons and defend the faith": Slicing and dicing the data. *Organizational Research Methods*, *12*, 276-295.

Brews, P. J., & Tucci, C. L. (2004). Exploring the structural effects of internetworking. *Strategic Management Journal*, *25*, 429-451.

Cascio, W. F., & Aguinis, H. (2001). The Federal Uniform Guidelines on Employee Selection Procedures (1978): An update on selected issues. *Review of Public Personnel Administration*, *21*, 200-218.

Cascio, W. F., & Aguinis, H. (2008). Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology*, *93*, 1062-1081.

Cashen, L. H., & Geiger, S. W. (2004). Statistical power and the testing of null hypotheses: A review of contemporary management research and recommendations for future studies. *Organizational Research Methods*, *7*, 151-167.

Chow, S. L. (1998). Précis of statistical significance: Rationale, validity, and utility. *Behavioral and Brain Sciences*, *21*, 169-239.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, *65*, 145-153.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York: Academic Press.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304-1312.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Cohen, J. (1994). The earth is round ($p < 0.05$). *American Psychologist*, *49*, 997-1003.

Connelly, F. M., & Clandinin, D. J. (2006). Narrative inquiry. In J. L. Green, G. Camilli & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 477-488). Mahwah, NJ: Lawrence Erlbaum.

Cortina, J. M, & Folger, R. G. (1998). When is it acceptable to accept a null hypothesis: No way, Jose? *Organizational Research Methods*, *1*, 334-350.

Cortina, J. M., & Landis, R. S. (2009). When small effect sizes tell a big story, and when large effect sizes don't. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in the organizational and social sciences* (pp. 287-308). New York: Routledge.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, *61*, 532-574.

Cummings, T. G. (2007). 2006 Presidential address: Quest for an engaged academy. *Academy of Management Review*, *32*, 355-360.

Duriau, V. J., Reger, R. K., & Pfarrer, M. D. (2007). A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements. *Organizational Research Methods*, *10*, 5-34.

Easterby-Smith, M., Golden-Biddle, K., & Locke, K. (2008). Working with pluralism: Determining quality in qualitative research. *Organizational Research Methods*, *11*, 419-429.

Fan, X. (2001). Statistical significance and effect size in education research: Two sides of a coin. *Journal of Educational Research*, *94*, 275-282.

Fisher, R. A. (1925). *Statistical methods for research workers.* Edinburgh, Scotland: Oliver and Boyd.

Fisher, R. A. (1959). *Statistical methods and scientific inference* (2nd ed.). New York: Hafner.

Gelade, G. A. (2006). But what does it mean in practice? The Journal of Occupational and Organizational Psychology from a practitioner perspective. *Journal of Occupational and Organizational Psychology*, *79*, 153-160.

Gephart, R. P. Jr. (2004). Qualitative research and the Academy of Management Journal. *Academy of Management Journal*, *47*, 454-462.

Gigerenzer, G. (1998). We need statistical thinking, not statistical rituals. *Behavioral and Brain Sciences*, *21*, 199-200.

Gilligan, C. (1982). *In a different voice.* Cambridge, MA: Harvard University Press.

Gómez-Bezares, F., & Gómez-Bezares, F. R. (2006). On the use of hypothesis testing and other problems in financial research. *Journal of Investing*, *15*, 64-67.

Green, J. L., Camilli, G., & Elmore, P. B. (Eds.). (2006). *Handbook of complementary methods in education research*. Mahwah, NJ: Lawrence Erlbaum.

Griffin, C., & Bengry-Howell, A. (2008). Ethnography. In C. Willig & W. Sainton-Rogers (Eds.), *The Sage handbook of qualitative research in psychology* (pp. 15-31). Thousand Oaks, CA: Sage.

Hakel, M. D., Sorcher, M., Beer, M., & Moses, J. L. (1982). *Making it happen—Designing research with implementation in mind.* Beverly Hills, CA: Sage Publications.

Hambrick, D. C. (1994). 1993 Presidential Address: What if the Academy actually mattered? *Academy of Management Review*, *19*, 11-16.

Hemphill, J. F. (2003). Interpreting the magnitude of correlation coefficients. *American Psychologist*, *58*, 78-79.

Henkel, R. E. (1976). *Tests of significance.* Beverly Hills, CA: Sage Publications.

Hiles, D., & Čermák, I. (2008). Narrative psychology. In C. Willig & W. Sainton-Rogers (Eds.), *The Sage handbook of qualitative research in psychology* (pp. 147-164). Thousand Oaks, CA: Sage Publications.

Hillman, A. (2009). Editor's comments. *Academy of Management Review*, *34*, 7-10.

Hubbard, R., & Ryan, P. A. (2000). The historical grown of statistical significance testing in psychology—and its future prospects. *Educational and Psychological Measurement*, *60*, 661-681.

Huberty, C. J. (1993). Historical origins of statistical testing practices: The treatment of Fisher versus Neyman-Pearson views in textbooks. *Journal of Experimental Education*, *61*, 317-333.

Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement*, *62*, 227-240.

Hunter, J. E. (1997). Needed: A ban on the significance test. *Psychological Science*, *8*, 3-7.

Iacobucci, D. (2005). From the editor. *Journal of Consumer Research*, *32*, 1-6.

Kim, H., Hoskisson, R. E., & Wan, W. P. (2004). Power dependence, diversification strategy, and performance in keiretsu member firms. *Strategic Management Journal*, *25*, 613-636.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*, 746-759.

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

Koontz, H. (1961). The management theory jungle. *Journal of the Academy of Management*, *4*, 174-188.

Krueger, L. E. (1998). The Ego has landed!: The 0.05 level of statistical significance is soft (Fisher) rather than hard (Neyman/Pearson). *Behavioral and Brain Sciences*, *21*, 207-208.

Latham, G. P. (2007). A speculative perspective on the transfer of behavior science findings to the workplace: "The times they are a-changin." *Academy of Management Journal*, *50*, 1027-1032.

Little, J. (2001). Understanding statistical significance: A conceptual history. *Journal of Technical Writing and Communication*, *31*, 363-372.

McCloskey, D. N. (1985). The loss function has been mislaid: The rhetoric of significance tests. *American Economic Review*, *75*, 201-205.

Miller, D. J., Fern, M. J., & Cardinal, L. B. (2007). The use of knowledge for technological innovation within diversified firms. *Academy of Management Journal*, *50*, 308-326.

Morgeson, F. P., & Campion, M. A. (2002). Minimizing tradeoffs when redesigning work: Evidence from a longitudinal quasi-experiment. *Personnel Psychology*, *55*, 589-612.

Murphy, K. R., & Myors, B. (1998). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests.* Mahwah, NJ: Erlbaum.

Phillips, N., Sewell, G., & Jaynes, S. (2008). Applying critical discourse analysis in strategic management research. *Organizational Research Methods*, *11*, 770-789.

Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement*, *64*, 916-924.

Podsakoff, P. M., & Dalton, D. R. (1987). Research methodology in organizational studies. *Journal of Management*, *13*, 419-441.

Pratt, M. G. (2008). Fitting oval pegs into round holes: Tensions in evaluating and publishing qualitative research in top-tier North American journals. *Organizational Research Methods*, *11*, 481-509.

Raver, J. L., & Gelfand, M. J. (2005). Beyond the individual victim: Linking sexual harassment, team processes, and team performance. *Academy of Management Journal*, *48*, 387-400.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276-1284.

Rynes, S. L. (2007). Let's create a tipping point: What academics and practitioners can do, alone and together. *Academy of Management Journal*, *50*, 1046-1054.

Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management*, *41*, 149-174.

Rynes, S. L., Giluk, T. L., & Brown, K. G. (2007). The very separate worlds of academic and practitioner periodicals in human resource management: Implications for evidence-based management. *Academy of Management Journal*, *50*, 987-1008.

Sauley, K. S. & Bedeian, A. G. (1989). 05: A case of the tail wagging the distribution. *Journal of Management*, *15*, 335-344.

Savall, H., Zardet, V., Bonnet, M., & Péron, M. (2008). The emergence of implicit criteria actually used by reviewers of qualitative research articles: Case of a European journal. *Organizational Research Methods*, *11*, 510-540.

Sawyer, A. G., & Peter, J. P. (1983). The significance of statistical significance tests in marketing research. *Journal of Marketing Research*, *20*, 122-133.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, *1*, 115-129.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, *105*, 309-316.

Shapiro, D. L., Kirkman, B. L., & Courtney, H. G. (2007). Perceived causes and solutions of the translation problem in management research. *Academy of Management Journal*, *50*, 249-266.

Shrout, P. E. (1997). Should significance tests be banned? Introduction to a special section exploring the pros and cons. *Psychological Science*, *8*, 1-2.

Sink, C. A., & Stroh, H. R. (2006). Practical significance: The use of effect sizes in school counseling research. *Professional School Counseling*, *9*, 401-411.

Smiddy, H. F. (1962). Presidential address: Theory and research for improvement of management practise. *Academy of Management Proceedings* (pp. 179-196). Briar Cliff Manor, NY: Academy of Management.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education*, *61*, 334-349.

Special research forum: Knowledge transfer between academics and practitioners. (2001). *Academy of Management Journal*, *44*, 340-440.

Steier, F. (Ed.). (1991). *Research and reflexivity*. London: Sage Publications.

Starbuck, W. H., Aguinis, H., Konrad, A. M., & Baruch, Y. (2008). Tradeoffs among editorial goals in complex publishing environments. In Y. Baruch, A. M. Konrad, H. Aguinis & W. H. Starbuck (Eds.), *Opening the black box of editorship* (pp. 250-270). New York: Palgrave-Macmillan.

Thompson, B. (1999a). Journal editorial policies regarding statistical significance tests: Heat is to fire as p is to importance. *Educational Psychology Review*, *11*, 157-169.

Thompson, B. (1999b). Statistical significance tests, effect size reporting and the vain pursuit of pseudo-objectivity. *Theory and Psychology*, *9*, 191-196.

Thompson, B. (2002). "Statistical," "practical," and "clinical:" How many kinds of significance do counselors need to consider? *Journal of Counseling & Development*, *80*, 64-71.

Thompson, B. (2006). Role of effect sizes in contemporary research in counseling. *Counseling and Values*, *50*, 176-186.

Trieschmann, J. S., Dennis, A. R., Northcraft, G. B., & Niemi, A. W. Jr. (2000). Serving multiple constituencies in business schools: M.B.A. program versus research performance. *Academy of Management Journal*, *43*, 1130-1141.

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods*, *6*, 371-386.

Tukey, J. W. (1991). The philosophy of multiple comparisons. *Statistical Science*, *6*, 100-116.

Tushman, M., & O'Reilly, C. (2007). Research and Relevance: Implications of Pasteur's Quadrant for Doctoral Programs and Faculty Development. *Academy of Management Journal*, *50*, 769-774.

Vacha-Haase, T., Nilsson, J. E., Reetz, D. R., Lance, T. S., & Thompson, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. *Theory & Psychology*, *10*, 413-425.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, *51*, 473-481.

Vaske, J. J., Gliner, J. A., & Morgan, G. A. (2002). Communicating judgements about practical significance: Effect size, confidence intervals and odds ratios. *Human Dimensions of Wildlife*, *7*, 287-300.

Vermeulen, F. (2007). "I shall not remain insignificant": Adding a second loop to matter more. *Academy of Management Journal*, *50*, 754-761.

Welcome to the Academy of Management Journal: Mission statement. (2009). *Academy of Management Journal*. Retrieved January 26, 2009, from http://journals.aomonline.org/amj/

Werner, S., & Brouthers, L. E. (2002). How international is management? *Journal of International Business Studies*, *33*, 583-591.

Wilkinson, S., & Kitzinger, C. (2006). Surprise as an interactional achievement: Reaction tokens in conversation. *Social Psychology Quarterly*, *69*, 150-182.

Wilkinson, S., & Kitzinger, C. (2008). Conversation analysis. In C. Willig & W. Sainton-Rogers (Eds.), *The Sage handbook of qualitative research in psychology* (pp. 54-71). Thousand Oaks, CA: Sage Publications.

Willig, C., & Stainton-Rogers, W. (2008). Introduction. In C. Willig & W. Sainton-Rogers (Eds.), *The Sage handbook of qualitative research in psychology* (pp. 1-12). Thousand Oaks, CA: Sage Publications.

Yates, F. (1951). The influence of "statistical methods for research workers" on the development of the science of statistics. *Journal of the American Statistical Association*, *46*, 19-34.

Zhang, X., Bartol, K. M., Smith, K. G., Pfarrer, M. D., & Khanin, D. M. (2008). CEOs on the edge: Earnings manipulation and stock-based incentive misalignment. *Academy of Management Journal*, *51*, 241-258.

Bios

Herman Aguinis is the Dean's research professor and a professor of Organizational Behavior and Human Resources at Indiana University's Kelley School of Business. He has published five books, about 70 journal articles and 20 book chapters and monographs on a variety of human resource management, organizational behavior, and research methods and analysis topics.

Steve Werner is a professor of Management at the C. T. Bauer College of Business at the University of Houston. His research interests include various aspects of international human resource management and research methods.

JéAnna Lanza Abbott is the Spec's Charitable Foundation professor in Corporate Social Responsibility at the Conrad N. Hilton College of Hotel and Restaurant Management at the University of Houston. Her research interests include negotiation, discrimination, and business law.

Cory Angert is a doctoral student in Management at the C.T. Bauer College of Business at the University of Houston. Cory's research interests include corporate governance, cultural analysis and the entertainment industry.

Joon Hyung Park is a doctoral student in Management at the C.T. Bauer College of Business at the University of Houston. Joon's research interests include stress, proactive behavior, workplace deviance, and work/life balance.

Donna Kohlhausen is a lawyer and adjunct professor at the University of Houston, Downtown. Her primary area of interest is business law.