Journal of Applied Psychology
2006, Vol. 91, No. 1, 6–8

0021-9010/06/$12.00 DOI: 10.1037/0021-9010.91.1.6

Choosing Between a Rock and a Hard Place When Data Are Scarce and the Questions Important: Reply to Hollenbeck, DeRue, and Mannor (2006)

Randall S. Peterson
London Business School

D. Brent Smith
Rice University

Paul V. Martorana
University of Texas at Austin

This article replies to J. R. Hollenbeck, D. S. DeRue, and M. Mannor's (2006) comment critiquing R. S. Peterson, D. B. Smith, P. V. Martorana, and P. D. Owens's (2003) use of a large number of statistical tests in research with a small sample. Although Hollenbeck et al.'s point of view is valid, it paints a one-sided picture of the trade-offs inherent in empirical research when data are scarce and the questions important. This reply specifically discusses the dilemmas Peterson et al. faced in conducting empirical research in a nascent area and suggests that theory development in such a situation can be well served by studies that use alternative or new methods with small samples. Theory development scholarship using small-sample research methods (e.g., case studies and Q sorting from archival sources) can be useful for stimulating ideas, theory, and research programs that can be tested with large-sample quantitative research.

Keywords: top management team, personality, research methods

Where to begin? Every new area of research has to begin from some point of reference, often with no obvious place to start. In deciding to pursue a potential new area of research, scholars typically put a great deal of time, effort, energy, and often money at risk in the hope of finding a reasonable place to initiate theory development on that topic (Murphy & Myors, 1998; Sutton & Staw, 1995). Those risks are particularly high in fields of research in which data are difficult to obtain. How many data points are sufficient to consider publication in this circumstance? And what happens if you discover a new research method along the way that you believe would make data collection for others in the field easier, a method that could potentially open that area of research to a much wider audience than ever before? At that point, you face a dilemma between continuing to collect data until the typical standards of statistical power in the field are met or attempting to publish those results in the hopes that the article would show the way to many additional researchers and lead more quickly to large-scale quantitative tests. This is the dilemma that we faced and that we asked the reviewers of the *Journal of Applied Psychology* to assess when we submitted our article (Peterson, Smith, Martorana, & Owens, 2003). This is also, we believe, the dilemma that Hollenbeck, DeRue, and Mannor (2006) did not fully accept in their critique of Peterson et al. (2003). We maintain that sometimes data on important questions are extremely difficult to obtain and require that tough trade-off decisions be made about how long to wait before publishing results. We respond to Hollenbeck et al.'s critique here by first articulating where we agree with them, and then we offer what we believe to be a more realistic assessment of the trade-offs faced by researchers working in nascent areas of psychology.

We begin by clarifying where we believe the two articles agree and disagree on the contributions of Peterson et al. (2003) and on the appropriate directions for future research. We agree substantially with Hollenbeck et al. (2006) about the contributions of our original article. We appear to agree that the article made three key contributions: (a) illustrating a potentially important research method for extracting quantitative information from qualitative data sources, (b) showing future researchers studying top management teams (TMTs) a methodology that can usefully be applied to additional research and, we hope, to larger sample research, and (c) developing a theory that is consistent with prior empirical results and offers a base for further research in the area. In acknowledging these contributions of our article, Hollenbeck et al. seemed to implicitly recognize why our article may have been worth publishing. It is also worthwhile to note two additional areas of agreement that are not explicit in our original article. First, we fully concur that published empirical articles are required in order for future meta-analyses to be performed. Second, we acknowledge that we should have been more measured in our interpretation of results in the discussion section of our article. We should have explicitly discussed the small sample as a concern for the stability of our results, along with the other limitations and alternative explanations we did discuss. In that respect, we fully accept one of the substantive critiques made by Hollenbeck et al.

Randall S. Peterson, London Business School, London, United Kingdom; D. Brent Smith, Jones Graduate School of Management, Rice University; Paul V. Martorana, McCombs School of Business, University of Texas at Austin.

We thank Michael Clegg-Butt, Lisa Moynihan, Madan Pillutla, and Sarah Ronson for thoughtful reviews of earlier versions of this article.

Correspondence concerning this article should be addressed to Randall S. Peterson, London Business School, Regent's Park, London NW1 4SA, United Kingdom. E-mail: rpeterson@london.edu

There are, of course, some points over which we do genuinely disagree. Our most important point of disagreement with Hollenbeck et al. (2006) is reflected in our differing reactions to one of the key observations made in Peterson et al. (2003). Hollenbeck et al. expressed disbelief that management and psychology researchers

> consistently use low levels of statistical power and that the level of power in these literatures has not changed over time—all despite numerous calls for formal power analyses as a basic requirement in the methods sections of studies that apply quantitative analyses. (p. 3)

We are not as surprised by this observation because we believe that statistical power is one important consideration, but not the only consideration, in establishing the value or contribution of any given article. Indeed, we believe that occasionally the best way to proceed in developing a new and difficult line of research is to lead the way by publishing a small-sample study, with an eye to encouraging later large-scale quantitative replication as well as assessment of the robustness of effects through meta-analysis (e.g., see the impact of House, Spangler, & Woycke, 1991; and Miller, Kets de Vries, & Toulouse, 1982). We believe such pathbreaking is particularly worthwhile when results are combined with methodological innovation, so that the advance also suggests *how* large sample replication might be achieved. In short, we believe that it is possible to have a significant and positive effect on a field of study by publishing a small-sample study, especially in the interests of theory development and methodological innovation.

Considering the Broader Trade-Offs

Hollenbeck et al. (2006) suggested two primary and related criticisms concerning Peterson et al. (2003). First, they noted that our reliance on a small sample suggests that our statistical power at the parameter-level is too low to allow us to correctly reject the null hypothesis in the presence of a population effect. Although they are less explicit regarding this particular criticism, it is certainly the central point of much of the literature they cited (cf. Maxwell, 2004; Mone, Mueller, & Mauland, 1996; Schmidt, 1996). Second, they noted that because of the large number of statistical significance tests we reported in the study, we should have been concerned about alpha inflation or cumulative Type I error and should have appropriately corrected our significance tests using an inflation-correction procedure (e.g., Bonferroni's adjustment). As a result of these concerns, Hollenbeck et al. argued that the parameter estimates we presented in the article are too unstable and therefore inherently misleading. To demonstrate the instability in our parameter estimates, Hollenbeck et al. conducted a simulation using our data to demonstrate that the subtraction of any one chief executive officer (CEO) from our sample frequently led to changes in the significance of parameters.

The issue of alpha inflation and power did most certainly occur to us. However, as noted, we believe that in an arena thin on empirical examinations, it was better to publish a study with low power and perhaps unstable parameters that was based on good theory and was consistent with prior empirical observations. We hoped that this would encourage large-scale replication that could lead to future meta-analysis and ultimately could provide better estimates of the population effects. After all, without this kind of large-sample research, it would be impossible to conduct a future meta-analysis. More important, for the advancement of knowledge, it may be not only necessary but also desirable to relax our traditionally stringent statistical constraints (relating to power and alpha levels) to allow research to be conducted in underexplored research areas. Recall the anecdote of the man searching for his lost keys under a streetlight where the lighting was better, despite clearly recalling that his keys were misplaced out in the darkness. Although it is easier to search where the resources are plentiful, the more fruitful search sometimes lies where the light is scarce.

Hollenbeck et al. (2006) did not fully articulate the trade-offs (and risks) associated with a focus on statistical power when sample sizes are inherently limited and data are difficult to access. As Murphy and Myors (1998) noted,

> The most serious cost that might be associated with the widespread use of power analysis is an overemphasis on scientific conservatism. If studies are hard to carry out, and require significant resources (time, money, energy), there may be less willingness to try new ideas and approaches, or to test creative hypotheses. The long-term prospects for scientific progress are not good if researchers are unwilling or unable to take risks or try new ideas. (p. 90)

We concur and are concerned that an unqualified emphasis on the importance of statistical power could have a stultifying effect on new and innovative research in the field.

Hollenbeck et al. (2006) went on to suggest that there is no countervailing value to reporting the results (parameters and significance tests) of small-sample studies given their instability. The risk, they argued, is that the scholarly audience will uncritically accept the findings as fact. We are clearly more optimistic than Hollenbeck et al. in our belief in the ability of the average reader of the *Journal of Applied Psychology* to properly qualify the results of a single small-sample study. We certainly did not intend to suggest that our study should be the definitive word on the nature of the relationship between CEO personality, TMT group process, and firm performance. Rather, given the lack of research in this area (i.e., relative darkness), we hoped to provide tentative initial results along with a creative methodology for conducting future research, thus showing other scholars how they might proceed in conducting the large-sample studies all researchers agree are desirable. We did not, however, state this explicitly, and perhaps we should have. Given appropriate qualification, we strongly believe that the benefits of publishing a study such as ours in an underresearched area substantially outweigh the risks of misinterpretation.

More important, we believe it is wholly inappropriate to call for a moratorium on research when there exists the possibility that the findings *might* be misinterpreted. We have personally read many studies on teams that assume the generalizability of laboratory research to field settings. Should we, therefore, ban laboratory research for fear that it might be misinterpreted or, worse yet, might obstruct field investigations of organizational groups? Of course, the answer is *no.* Nor should we call for a blanket moratorium on the publication of small-sample studies for fear that some in the audience might uncritically accept the findings as fact. We believe that science is self-correcting in the long run and that ultimately our new empirical results, like all results, will stand or fail on the basis of replication or lack thereof (Popper, 1959/2002).

Finally, Hollenbeck et al. (2006) implied that a qualitative approach would have been more suitable to our research area. Our initial reaction to their recommendation for qualitative analysis of small-sample studies was puzzlement—why should qualitative analysis be any more stable than quantitative analysis when the issue is a small sample size? Surely a qualitative analysis would be equally unstable with the removal of individual observations. At a minimum, quantitative analysis allows for correction through meta-analysis, whereas a qualitative analysis does not. However, there are several additional reasons why we explicitly chose a quantitative path. First, we hoped to introduce a means of extracting quantitative data from qualitative information. We agree with Hollenbeck et al. that this is the contribution of Peterson et al. (2003). Second, we hoped to introduce this line of research to the readership of the *Journal of Applied Psychology*. Arguably, this would not have been possible with a purely qualitative investigation (please note the lack of qualitative studies published in this journal over the past decade). Although the null hypothesis significance testing paradigm is significantly flawed, from a pragmatic standpoint, it represents the point of entry for many major psychology journals.

Conclusion

In summary, we believe that Hollenbeck et al. (2006) were correct in arguing that Peterson et al.'s (2003) results were not adequately qualified. However, we also believe that Hollenbeck et al. did not sufficiently acknowledge the trade-offs associated with an emphasis on statistical power and controlling Type I error. Such an emphasis, we believe, presents a high-risk strategy for theory development. If we hold to the ideal standards of publication from Hollenbeck et al.'s point of view, we are troubled by the risk of discouraging scholars from taking risks and conducting innovative research with small samples. Such research, although flawed, may open new areas of theory and show the way for future researchers to conduct the kind of large-sample research studies that all psychologists agree are needed to have full confidence in a set of findings. At the same time, we believe most applied psychologists would agree that CEO personality and TMT dynamics are critically important areas to study—important enough that we as applied psychologists *ought* to be studying them regardless of the limited data available. Therefore, we believe that the real question to ask of scholars in the field is the following: Is it better to take some risk in publishing small-sample studies that could inspire theory development and large-scale quantitative studies but that also open the field to the possible need for correction through meta-analysis, or is it better to face the risk of having virtually no research at all published on an important topic? Here we come down clearly on the side of publishing more small-sample studies of the best possible quality, with the expectation of possible long-term correction through meta-analysis, rather than imposing a standard on a field of research that may effectively render it nonexistent. Thus, although we agree with Hollenbeck et al. that the ultimate goal should be developing CEO personality research that has established theory supported by reliable and valid data, it seems we disagree on the most appropriate way to get there.

References

Hollenbeck, J. R., DeRue, D. S., & Mannor, M. (2006). Statistical power and parameter stability when subjects are few and tests are many: Comment on Peterson, Smith, Martorana, and Owens (2003). *Journal of Applied Psychology, 91,* 3–7.

House, R. J., Spangler, W. D., & Woycke, J. (1991). Personality and charisma in the U.S. presidency: A psychological theory of leader effectiveness. *Administrative Science Quarterly, 36,* 364–390.

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods, 9,* 147–163.

Miller, D., Kets de Vries, M. F. R., & Toulouse, J. M. (1982). Top executive locus of control and its relationship to strategy-making, structure and environment. *Academy of Management Journal, 25,* 237–253.

Mone, M. A., Mueller, G. C., & Mauland, W. (1996). The perceptions and usage of statistical power in applied psychology and management research. *Personnel Psychology, 49,* 103–120.

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

Peterson, R. S., Smith, D. B., Martorana, P. V., & Owens, P. D. (2003). The impact of chief executive officer personality on top management team dynamics: One mechanism by which leadership affects organizational performance. *Journal of Applied Psychology, 88,* 795–808.

Popper, K. R. (2002). *The logic of scientific discovery.* London: Routledge. (Original work published 1959)

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

Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly, 40,* 371–384.

Received December 15, 2004
Revision received August 22, 2005
Accepted September 10, 2005 ■