

FROM THE EDITORS

PUBLISHING LABORATORY RESEARCH IN *AMJ*: A QUESTION OF WHEN, NOT IF

Making a theoretical contribution is a core element of *AMJ*'s mission statement and its criteria for evaluating manuscripts. In describing what theory is and is not, Sutton and Staw (1995: 378) wrote, "Theory is about the connections among phenomena, a story about why acts, events, structure, and thoughts occur. Theory emphasizes the nature of causal relationships, identifying what comes first as well as the timing of such events" (1995: 378). The causal relationships aspect of theory is especially important in the applied sciences, where theories may eventually inspire interventions designed to improve some business outcome. For example, a scholar with expertise in motivation might advise an organization to use specific and difficult goals in its performance management system. That scholar would feel more comfortable making such a recommendation if he or she truly believed that specific and difficult goals *cause* increases in task performance (Locke & Latham, 1990).

Unfortunately, inferring causality is one of the most difficult aspects of scientific research. Drawing on J. S. Mill, Cook and Campbell (1979) suggested that inferring causality depends on three criteria: (1) covariation between cause and effect, (2) temporal precedence of the cause, and (3) elimination of alternative explanations for the possible cause-effect connection. Consider a study in which employees were surveyed about the specificity and difficulty of their work goals in one time period, with their supervisors asked to rate their task performance in a subsequent time period. A correlation between those two measures would provide some support for the first two criteria for inferring causality, but what of the third? It may be that high task performance causes employees to be given more specific and difficult goals, meaning that the "effect" actually causes the "cause." It may also be that a third variable impacts both work goals and task performance, creating a spurious relationship between the two. For example, leaders high in "initiating structure" may assign spe-

cific and difficult goals more frequently while also engaging in behaviors that benefit employee performance. Alternatively, conscientious employees may seek out more challenging goals while also performing tasks in an organized and diligent fashion because their perseverance and hard work has raised their performance on ordinary goals. Those two examples represent the classic unmeasured variables problem, a problem that occurs whenever an unmeasured variable is (1) correlated with a presumed cause, and (2) predictive of the presumed effect (James, 1980). Given the complexity of the phenomena studied in management research, the unmeasured variables problem is virtually unavoidable, with one exception. In laboratory research, the experimenter controls which units receive a particular level of the independent variable at a particular time. Such choices are made using random assignment, with assignment to conditions performed with the roll of a die, the use of a random number table, or some similar procedure. Because those choices will, by definition, lack any pattern, no unmeasured variable can be meaningfully correlated with the independent variable of interest. Thus, although there are still unmeasured variables, there is no longer an unmeasured variables problem (James, 1980). For this reason, random assignment has been termed "the great *ceteris paribus*" of causal inference (Cook & Campbell, 1979: 5; italic in original). Moreover, the fact that the experimenter controls the levels of the independent variable rules out the possibility that the outcome actually caused the predictor in a given study.

For these reasons, and others described below, lab studies have been a vital component of the theory-centered research published in a number of journals. For the purposes of this review, *lab studies* are defined as studies involving undergraduate participants that occur in an environment that was created for research purposes. It is true that other combinations of samples and settings are possible. For example, what is often termed a "field experiment" places actual employees in an environment where randomly assigned conditions have been created. What might be termed a "student survey" examines undergraduates in an actual classroom environment. In practice, however, lab studies are much more common than either field experiments

I would like to thank Duane Ireland, Mike Pratt, Dave Ketchen, Peter Bamberger, and Micki Kacmar for their helpful comments on a draft of this work. I would also like to thank Jessica Mueller for her assistance in the literature search.

or student surveys. Moreover, discussions of the relative merits of lab studies typically focus on the simultaneous use of undergraduate participants and artificial environments. This essay will therefore be consistent with those discussions, many of which are cited in the paragraphs to follow.

Historically speaking, lab studies have been somewhat rare in *AMJ*'s pages. A search of the past decade reveals that about 5 percent of the articles published in *AMJ* are lab studies. That search included 60 issues, 24 of which included at least one lab study. That rate therefore indicates that a lab study appears about every second or third issue of *AMJ*. To provide some context for that 5 percent rate, note that 22 percent of the articles published in the *Journal of Applied Psychology* over the same time period were conducted in the lab. Of course, it should be noted that *Journal of Applied Psychology* specializes in micro research, whereas *AMJ* is a "big tent" journal encompassing both micro and macro domains. Thus, the two publication rates would not be expected to be similar. Still, those rates indicate that lab studies are not published in *AMJ* with great frequency, which is one of the reasons that our editorial team wanted to prepare a "From the Editors" on this subject.

There may be (at least) two reasons for the relative shortage of lab studies at *AMJ*. First, there may be a bias among editors, editorial board members, or ad hoc reviewers against lab studies that results in higher rejection rates. Second, authors of lab studies may presume that such a bias exists and therefore decide to send their submissions elsewhere. In the absence of any compelling data, all I can draw on is my own firsthand experience as an *AMJ* associate editor and as an *AMJ* author. My experience as an associate editor provides more support for the second possibility than the first. Despite the fact that my research lies in areas in which the laboratory has a significant presence, I have only edited three lab studies in my first year on the job. As an author, I have published eight lab studies in top-tier journals, including one in *AMJ*. Of the other seven, one was submitted to (and rejected by) *AMJ*, with the reviews failing to indicate any sort of bias. The remaining six were never sent to *AMJ* for several reasons, most involving the nature of their theoretical contribution.

The purpose of this essay is to reiterate the fact that *AMJ* does indeed welcome laboratory studies. *AMJ*'s mission statement notes that, "All empirical methods—including, but not limited to, qualitative, quantitative, field, laboratory, meta-analytic, and combination methods—are welcome." Of course, the impact of a mission statement pales in comparison to what is actually in the pages of the journal. With that in mind,

this "From the Editors" has two other goals. First, I wish to spotlight some of the lab studies that have been accepted by *AMJ* in the past few years, to illustrate how laboratory methods can be used to conduct the sort of theory-centered research that is so valued by the journal. Second, I wish to describe how reviewers should go about evaluating lab studies, particularly when their experience in doing so is limited. To borrow the titular phrase from Ilgen's (1985) classic treatment of these issues, my hope is to show that laboratory research in *AMJ* is a question of *when*, not *if*.

The *AMJ* Lab Study: Some Exemplars

Although the lab studies published in *AMJ* over the past few years have involved different literatures and multiple author teams, they do share some elements that were likely instrumental to their publication. Of those shared elements, the following may be the most critical to efforts to plan and conduct laboratory work with a view toward publication in this journal:

Theoretical contribution. Empirical articles can make a theoretical contribution by testing, extending, or building theory (Colquitt & Zapata-Phelan, 2007). Many of the lab studies published in *AMJ* make a theoretical contribution in multiple respects. For example, Schweitzer, Ordóñez, and Douma (2004) built new theory by examining a previously unexplored relationship: the relationship between goal setting and unethical behavior. The authors also tested existing theory by grounding their hypotheses in an integration of goal setting theory and an existing model of deception. Similarly, Fong (2006) built new theory by linking emotional ambivalence to creativity, grounding that new relationship with logic from informational models of emotion. Seijts, Latham, Tasa, and Latham (2004) extended both goal setting theory and research on the goal orientation construct, becoming the first study to contrast specific and difficult "learning goals" with specific and difficult "performance goals." As a final example, Johnson, Hollenbeck, Humphrey, Ilgen, Jundt, and Meyer (2006) built new theory by introducing a new construct: "cutthroat cooperation." They also grounded their predictions using structural adaptation theory, an integrative perspective that they created by drawing from structural contingency theory and social interdependence theory.

Psychological realism. Berkowitz and Donnerstein (1982) noted that two forms of realism in lab studies can be described. One is "mundane realism," which reflects whether the experimental setting and procedure resemble things that occur in

the real world. For example, a study that uses a business memo task to assess creativity would have more mundane realism than a study that uses a structure-building task. A more important form is “experimental realism,” a sort of psychological realism. Here the experimental setting and procedure capture the *intended essence* of the constructs of interest. It is this sort of psychological realism that impacts the construct validity of experimental manipulations and the ultimate replicability of experimental findings.

A number of factors can promote psychological realism, including placing participants in real rather than hypothetical situations, using vivid and engrossing manipulations and tasks, and creating real stakes by using monetary or credit-based contingencies. Porath and Erez’s (2007) study of the behavioral consequences of rudeness provides a good illustration of psychological realism. The authors manipulated rudeness by having an experimenter criticize a group of participants after a confederate arrived late for the study. The sense of disrespect that the participants felt captured the intended essence of the rudeness construct in a way that a hypothetical scenario could not. In another example, Ellis (2006) examined the effects of stress on team performance, manipulating stress by videotaping performance while noting that the videos of poorly performing teams might be broadcast in their classes. That manipulation captured the essence of a stressful event, lending a psychological realism to the study.

Measures of actual behavior. Aside from random assignment, one of the most important strengths of laboratory studies is that they supply a forum for measuring actual behavior. This is in contrast to most field studies, which rely heavily on self-reports of behavior, or peer or supervisor reports (Dipboye & Flanagan, 1979). Porath and Erez’s (2007) study of the behavioral effects of rudeness did not assess declines in helping by giving participants a self-report version of a citizenship scale. Instead, the experimenter “accidentally” knocked over a jar with ten pencils, operationalizing helping as the number of pencils picked up by the participants. In another example, Schweitzer et al.’s (2004) study of goal setting and unethical behavior assessed the criterion by allowing participants to score their own word creation task in a context in which the honesty of their scores could be tracked. This ability to measure actual behavior in the lab also allows studies to focus on a number of mediating and process variables. For example, studies of individual and team performance have tracked a number of intervening mechanisms, including information search, information sharing, mental models, and transactive memory (Ellis, 2006; Johnson et al.,

2006; Seijts et al., 2004). Indeed, the feasibility of using interviews, behavioral observation, content analysis, verbal protocol analysis, and other techniques makes laboratory studies a valuable venue for building theory, not just testing theory.

Evaluating Lab Studies in *AMJ*: Some Guidelines

Having spotlighted some of the lab studies that have already appeared in *AMJ*’s pages, I turn now to this question: How should reviewers go about evaluating lab studies that are submitted to the journal? In view of the discussion above, I submit that a study’s theoretical contribution, psychological realism, and (depending on the research question) use of behavioral measures all provide a good place to start. Beyond that, lab studies should be evaluated in the same manner as field studies. *AMJ*’s reviewer evaluation form includes nine criteria for evaluating manuscripts. The following three criteria warrant specific discussion with respect to lab studies:

Appropriateness for *AMJ*. The first criterion on the evaluation form is “appropriateness for *AMJ*.” If a reviewer has a bias against lab studies, applying this criterion is one place where it might manifest itself. I hope this “From the Editors” conveys that lab studies and field studies should not automatically differ on the appropriateness count. That judgment should be based on the content of the study and the degree to which it matches the mission of the journal, as described in “Information for Contributors.”

Technical adequacy. The third criterion on the evaluation form is “technical adequacy.” Although this criterion encompasses a wide array of issues, Cook and Campbell’s four types of validity—construct validity, statistical conclusion validity, internal validity, and external validity—offer one means of examining the issue (Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002). I begin with external validity because that issue is particularly salient with lab studies. Assume, for the sake of argument, that a reviewer included the following sentence in his/her review: “These results are interesting, but I know that lab studies have lower external validity.” Drawing on Kerlinger and Lee’s (2000) discussion of the methods of knowing, we might ask how this reviewer came to that conclusion. Was it the *method of intuition*, where it was simply self-evident to him or her that lab studies are not generalizable? Or was it the *method of authority*, with the reviewer’s advisor or doctoral training informing his or her view? Or perhaps the relevant method is the *method of tenacity*, whereby a view is held without even knowing where it came from.

Scientists, of course, are trained to rely on the *method of science*, which relies on conclusions

based on the careful analysis of valid data. The increased use of meta-analysis has actually given us a burgeoning set of data on the external validity of lab studies. Anderson, Lindsay, and Bushman (1999) reviewed 38 different meta-analyses that included lab vs. field breakdowns, correlating the 38 lab effect sizes with the 38 field effect sizes. Their analysis revealed a .73 correlation between the two sets of effect sizes, suggesting a high degree of generalizability from lab to field. Such comparisons can be done more informally within any literature that includes a meta-analysis with lab-field breakdowns. For example, Cohen-Charash and Spector's (2001, 2002) meta-analysis of the justice literature provides such breakdowns for 10 different effect sizes. The correlation between the lab effect sizes and the field effect sizes is .97. That level suggests that justice findings in the lab are likely to generalize to field settings.

Of course, such convergence may vary across literatures, and it is certainly the case that some research questions are more "labable" than others. In the absence of relevant data, a scientific analysis would require building a theory that explains why lab-field effects might be expected to diverge. For example, one could consider the constructs that might differ for undergraduate and employee populations, such as age, work experience, and so forth. One could also consider the constructs that might differ in artificial settings and actual organizations, such as the magnitude of stakes, the existence of history, the possibility of a shared future, and so forth. The analysis would then explore whether there are theoretical reasons to expect one of those constructs to moderate the relevant predictor \rightarrow outcome relationships.

Consider, for example, a study examining perceived ability as an antecedent of trust, defined here as the willingness to be vulnerable (Mayer, Davis, & Schoorman, 1995). It is difficult to provide a conceptual rationale for an ability by age or an ability by work experience interaction effect. Those two variables rarely explain variance when they appear in regression equations as controls, and it seems unlikely that those null effects are actually masking crossed interactions. (There may be little to be gained from using MBA students in lab studies for just this reason, though the use of MBAs is not common enough to allow quantitative analysis of the issue.) Regardless, now consider a study examining the shared emotional connection between trustor and trustee as an antecedent of trust (McAllister, 1995). One could indeed provide a conceptual rationale for an emotional connection by history interaction effect, given that long-standing bonds are likely to be more impactful than bonds forged in a laboratory. Ultimately, however,

such a difference is an empirical question that could best be answered by examining the relationship using both lab and field methods.

That suggestion to use multiple methods raises an important observation made by Cook and Campbell in their discussion of external validity (Cook & Campbell, 1979; Shadish et al., 2002). They suggested that external validity is best viewed as a characteristic of a *stream of research*, not a single study. The primary means of increasing the external validity of a single study is random sampling of participants and contexts, or at least deliberately choosing heterogeneous samples of participants and contexts (Cook & Campbell, 1979; Shadish et al., 2002). Neither technique is used with any frequency in management research, in the lab or in the field. Indeed, Dipboye and Flanagan's (1979) analysis of field studies suggested that they are almost as homogeneous in their samples and methods as lab studies, typically relying on survey studies of predominantly managerial, technical, or professional samples.

Turning to the other three kinds of validity that underlie technical adequacy, internal validity will be a strength whenever random assignment is utilized, freeing reviewers from having to focus on potential unmeasured variables problems. Statistical conclusion validity is subject to the same threats as a field study, as the extent to which one makes correct inferences about the existence and strength of covariation still depends primarily on statistical power. The lab environment can reduce noise and distractions, however, resulting in a reduction of extraneous variance. Construct validity is a particularly critical issue in lab studies, as the validity of both the manipulations and measures must be judged. Do the manipulations adequately represent the constructs of interest? Do they foster psychological realism? Do the manipulation checks show that the inductions were perceived psychologically by the participants? These sorts of concerns represent construct validity issues that are unique to the laboratory environment.

Interestingness, innovativeness, and novelty.

The sixth criterion on the *AMJ* reviewer evaluation form is "interestingness, innovativeness, and novelty." *AMJ*'s emphasis on those standards is perhaps best summarized by this sentence from its mission statement: "Authors should strive to produce original, insightful, interesting, important, and theoretically bold research." Lab studies that are largely replications, or that merely "add a variable" to previously validated models, may not be viewed as theoretically bold on the part of reviewers. Indeed, the lab studies published in *AMJ* over the past several years are notable for the multifaceted nature of their theoretical contributions. Al-

most all of them included some combination of testing, extending, and building theory. That theoretical boldness may therefore be particularly critical to publishing lab studies in *AMJ*.

Conclusion

Returning to the goal setting example that opened this essay, why do scholars feel so comfortable telling managers to incorporate specific and difficult goals into their performance management systems? Is it because of one grand field study by Locke, Latham, and colleagues that was performed on a random sample of employees from a random sample of industries? No—it is because the effects of specific and difficult goals were demonstrated in both lab and field settings using puzzles, proofreading, Lego construction, and juggling—not to mention faculty performance, scientific engineering, energy conservation, and safety behavior (Locke & Latham, 1990). That pairing of lab and field studies within a research stream represents a form of triangulation (Denzin, 1978), giving scholars confidence in the validity of the findings. I hope that this “From the Editors” results in more lab studies being submitted to *AMJ*, so that this sort of triangulation can occur within the pages of the journal.

Jason A. Colquitt
Gainesville, Florida

REFERENCES

- Anderson, C. A., Lindsay, A. J., & Bushman, B. J. 1999. Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science*, 8: 3–9.
- Berkowitz, L., & Donnerstein, E. 1982. External validity is more than skin deep: Some answers to criticisms of laboratory experiments. *American Psychologist*, 37: 245–257.
- Cohen-Charash, Y., & Spector, P. E. 2001. The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 86: 278–321.
- Cohen-Charash, Y., & Spector, P. E. 2002. Erratum to “The role of justice in organizations: A meta-analysis.” *Organizational Behavior and Human Decision Processes*, 89: 1215.
- Colquitt, J. A., & Zapata-Phelan, C. P. 2007. Trends in theory building and theory testing: A five-decade study of the *Academy of Management Journal*. *Academy of Management Journal*, 50: 1281–1303.
- Cook, T. D., & Campbell, D. T. 1979. *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton-Mifflin.
- Denzin, N. K. 1978. *The research act* (2nd ed.). New York: McGraw-Hill.
- Dipboye, R. L., & Flanagan, M. M. 1979. Research settings in industrial and organizational psychology. *American Psychologist*, 34: 141–150.
- Ellis, A. P. J. 2006. System breakdown: The role of mental models and transactive memory in the relationship between acute stress and team performance. *Academy of Management Journal*, 49: 576–589.
- Fong, C. T. 2006. The effects of emotional ambivalence on creativity. *Academy of Management Journal*, 49: 1016–1030.
- Ilgen, D. R. 1985. Laboratory research: A question of when, not if. In E. A. Locke (Ed.), *The generalizability of laboratory experiments: An inductive survey*: 257–267. Lexington, MA: Heath.
- James, L. R. 1980. The unmeasured variables problem in path analysis. *Journal of Applied Psychology*, 65: 415–421.
- Johnson, M. D., Hollenbeck, J. R., Humphrey, S. E., Ilgen, D. R., Jundt, D., & Meyer, C. J. 2006. Cutthroat cooperation: Asymmetrical adaptation to changes in team reward structures. *Academy of Management Journal*, 49: 103–119.
- Kerlinger, F. N., & Lee, H. B. 2000. *Foundations of behavioral research*. Fort Worth, TX: Harcourt.
- Locke, E. A., & Latham, G. P. 1990. *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. 1995. An integrative model of organizational trust. *Academy of Management Review*, 20: 709–734.
- McAllister, D. J. 1995. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38: 24–59.
- Porath, C. L., & Erez, A. 2007. Does rudeness really matter? The effects of rudeness on task performance and helpfulness. *Academy of Management Journal*, 50: 1181–1197.
- Schweitzer, M. E., Ordóñez, L., & Douma, B. 2004. Goal setting as a motivator of unethical behavior. *Academy of Management Journal*, 47: 422–432.
- Seijts, G. H., Latham, G. P., Tasa, K., & Latham, B. W. 2004. Goal setting and goal orientation: An integration of two different yet related literatures. *Academy of Management Journal*, 47: 227–239.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Sutton, R. I., & Staw, B. M. 1995. What theory is not. *Administrative Science Quarterly*, 40: 371–384.

Copyright of *Academy of Management Journal* is the property of Academy of Management and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.