

Difficult but Doable: Increasing the Internal Validity of Organizational Change Management Studies

The Journal of Applied Behavioral Science
201X, Vol XX(X) 1–5
© 2013 NTL Institute
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0021886313515614
jabs.sagepub.com



**Eric Barends¹, Barbara Janssen¹, Wouter ten Have¹,
and Steven ten Have¹**

The purpose of our study was to conduct a systematic review of the quality of the body of evidence on the effectiveness of organizational change interventions. Our findings—that this body of evidence is sparse and low in quality—have helped to stimulate discussion among leading scholars in organizational change management (OCM). The commentaries by Woodman, Beer, and Schwartz and Stensaker raise many interesting and important points. Our response to their thought-provoking commentaries attempts to respond to their questions, correct misinterpretations, and elaborate on the implications they raise for improving the internal validity of OCM studies.

The Background of Our Study: Enhancing Evidence-Based Practice in OCM

As members of the Evidence-based Management Collaborative (www.CEBMa.org), we seek to promote the use of evidence-based practice in management, including OCM. The basic idea of evidence-based practice is that management decisions should incorporate the “best available” evidence. Evidence means various types of information. It may come from scientific research, internal business data and even personal experience. Anything can count as evidence if it’s judged valid, reliable, and relevant. Evidence-based practice is *not* about prizing one source of evidence as more valid than other. In fact, it would be naive to assume that the outcome of controlled scientific research alone could provide clear and comprehensive answers on how to tackle a managerial problem.

¹VU University of Amsterdam, Amsterdam, Netherlands

Corresponding Author:

Eric Barends, VU University of Amsterdam, De Boelenlaan 1105, Room 2A17, 1081 HV Amsterdam, Netherlands.

Email: e.barends@hotmail.com

An important principle of evidence-based practice is the notion of “best” available evidence: the most valid and reliable evidence given the type of question or problem at hand. For instance, the most valid and reliable information on the holiday destination with the least chance of rain in early August obviously comes from statistics on the average rainfall per month and not from the personal experience of a colleague who visited the destination once. The same counts for questions regarding the effectiveness of a change intervention. As our article points out, and Woodman further explicates in his commentary, in determining the validity and reliability of evidence regarding the effectiveness of change interventions (does it work?), internal validity is the most important indicator. For example, when making a decision whether to use Six Sigma to reduce medical errors in a hospital, the outcome of several controlled, longitudinal studies with a large combined sample size provides stronger evidence than a single case study. However, the “best” types of studies are not always available. In those situations, a manager has no other option than to make a decision in part based on a studies with a lower internal validity, because that constitutes “the best available” evidence. In evidence-based practice, the goal is to increase the likelihood of a favorable outcome from an intervention, so some evidence, if relevant and reliable, is better than no evidence at all.

As practitioner-scholars in the field of OCM with a combined managerial experience of more than 80 years, we often encounter practical concerns about the effect of change interventions. These interventions involve topics such as downsizing, incentivizing employees, encouraging entrepreneurship, managing mergers, and improving performance. These are all “does it work?” questions. In these cases, results from scientific studies with a high internal validity provide the “best” evidence. As noted in our article, our search for relevant studies in scientific databases turned up relatively few studies with high internal validity. This observation fueled our professional curiosity and led us to our systematic review questions: What is the quality of the body of scientific evidence underlying organizational change interventions? Does it provide change managers with quality evidence that supports effective decision making? Or should we be skeptical regarding the usability and value of the research evidence in OCM? The results of our systematic review lead us to recommend skepticism regarding the evidence for the effectiveness of OCM interventions. Thus we seek to generate discussion of what could be done to change this. Accepting the current state of affairs would mean that managers, employees, and other stakeholders would never conclusively know whether OCM practices work.

Misinterpretations and Clarifications

All the commentators mention the classification scheme we used to identify the most valid and reliable research designs for demonstrating a cause-and-effect relationship between an intervention and an outcome: the levels of internal validity. This classification scheme is not new. In evidence-based medicine, education, public policy, criminology, and (recently) management, it is widely used to determine the “best available” evidence regarding the effectiveness of a treatment, teaching method, policy, or management intervention. Beer as well as Schwartz and Stensaker infer from this classification scheme that research designs lower in the hierarchy, such as case

studies or qualitative research, are “poor,” “flawed,” or “not valuable.” This is a misconception. Value can come in many forms. A dinner at a three-star restaurant is likely to be higher in terms of culinary sophistication than a meal at the local pub, but this does not mean that a pub meal is poor, flawed, or of low value.

One cause of this misconception might be the notion of “outcome.” Surely, outcome is not only about effectiveness (what works?). It can also be about process (how does it work?), theory (why does it work?), prevalence (how often/how many?), procedure (in which way?), or attitude (how does the target group feel about the intervention?). Cross-sectional studies and qualitative research can provide high-quality and valuable evidence regarding *these* types of outcomes. Moreover, we agree with Beer as well as Schwartz and Stensaker that, for other outcomes beside effectiveness, evidence need not only come from controlled scientific studies. Our study only focused on the outcome of effectiveness: What is the likelihood that the intervention will indeed work and to what extent? To answer that particular question, the outcome of a randomized controlled study with a pretest obviously constitutes the most valid and reliable evidence.

Woodman is correct when he points out that sample size is an important condition: Sample size and effect size tend to be negatively correlated (e.g., Levine, Asada, & Carpenter, 2009; Slavin & Smith, 2009). It is therefore an important indicator of a study’s statistical validity. However, that was not the focus of our study. His observation that a pretest is not essential to establish internal validity is absolutely right, provided that the individuals, teams, or organizations were randomly assigned to the intervention and control group. The designs in our study that were classified as “controlled posttest-only design” did not use randomization and were therefore placed lower in the hierarchy.

Finally, Beer argues that one of our study’s weaknesses is that we did not provide information about whether the researchers who conducted studies lower in the hierarchy discussed their study’s methodological and other limitations. This observation is correct, but given our study’s objective, that information is irrelevant. But more important, we fail to see that reporting possible biases or methodological limitations could, as Beer suggests, increase confidence in the validity of the findings.

Increasing Internal Validity: Difficult but Doable

All the commentaries point out repeatedly that experimental and quasi-experimental studies are difficult to execute. We agree, up to a point. Yes, it is difficult to gain senior management’s commitment to do research, and yes, the field of change management has a lot of constraints that make it difficult to use randomization or control groups. We are very aware of this, but the fact that doing better research is difficult is no defense of poor-quality research, especially regarding important practical questions. Let’s try not to emulate the chef who complains that it is really hard to make a good soufflé after a customer’s complains his soufflé has not risen. True, but the soufflé is still flat.

Conducting studies more appropriate to the question “what works” is challenging but certainly not impossible. It is difficult but doable. First of all, as Beer points out, there is increasing receptivity among senior managers and corporate leaders to the idea

of conducting controlled studies within their companies. The zeitgeist is changing. New developments such as evidence-based practice, big data, pressures for accountability, and Internet access to research outcomes (e.g., Google Scholar) are making managers more aware that the well-being and effectiveness of employees and organizations depend to a large extent on the use of high-quality evidence in making decisions. For this reason, we suspect that the call by Woodman and by Beer for an increased use of collaborative management research, engaged scholarship, and practitioner-researcher teams will be honored. It may eventually lead to more systematic assessment of the outcomes of (change) management decisions and a process of constant experimentation and critical reflection about what works and what does not.

We strongly support Woodman's call for more quasi-experimental and (controlled) observational research. Even in medicine, randomization is not always feasible, and as a result most clinical evidence comes from quasi-experimental and observational research. As in management, medicine faces practical constraints on the use of randomization as well as ethical considerations. In these situations quasi-experiments or observational research are used. As our article points out, such research can also provide evidence with high internal validity, particularly when studies are replicated and under varying conditions. During the many workshops and seminars we have given on evidence-based practice, we have noticed that researchers in the field of OCM are not always familiar with the concept of quasi-experimental or observational research. This unfamiliarity is reflected by the fact that of the 563 included studies, only 17 (3%) studies used a cohort study design and 14 studies (2%) used a case control design. For this reason we would like to use this opportunity to promote the use of case control designs.

A case control study is a longitudinal retrospective study comparing one group of employees, teams, divisions, or organizations with a particular outcome (e.g., above-average performance) retrospectively with a control group without this outcome. A specific feature of this design is that it takes the outcome as a starting point, instead of the intervention. Given the fact that most organizations generate a wide range of (quantitative and qualitative) outcome data that can be analyzed retrospectively, such as financial data (cash flow, solvability), business outcomes (return on investment, market share), customer/client impact (customer satisfaction, medical errors), performance indicators (occupancy rate, productivity, failure frequency), HR metrics (absenteeism, employee engagement, turnover rates), and so on, a case control study is a controlled design that could be applied in management relatively easily. An illustrative example of a case control design in the field of OCM is the study by Medin, Ekberg, Nordlund, and Eklund (2008). The objective of this study was to explore whether organizational change and job-related stress are associated with a higher risk of heart failure among employees. The researchers selected a total of 65 cases with first-ever stroke from four hospitals. During the same period, 103 persons in the same age-group were randomly selected out of the general population. A validated questionnaire was used to collect data on organizational change, job-related stress, and traditional medical risk factors. The results showed that organizational change indeed increased the likelihood of heart failure. Although a case control design cannot show causality—the association of organizational change with heart failure does not prove that organizational change causes

heart failure—it does have a higher internal validity than uncontrolled designs. The key requirements for this higher internal validity are that the selection of cases and controls are based on objective and validated criteria and that objective and validated measurement methods are used to measure the outcome.

Finally, we agree with Woodman's comment that the field of OCM "should not give up too quickly on the value of subjecting specific findings or studies to repeated tests of their 'truth' value." In fact, we would argue that *replication* offers the most viable approach to increase the internal validity of the body of evidence on the effectiveness of organizational change interventions. The importance of replication has become clear again in a recent study by Fanelli and Loannidis (2013). This study demonstrated, based on a systematic review of 82 meta-analyses, that studies in disciplines where there is little replication are more likely to report extreme effects and falsified findings than studies in disciplines where the value of replication is well accepted. The findings of our systematic review have demonstrated that there is a severe lack of replication in OCM. Replications should therefore be considered to be the gold standard, especially since randomization is so difficult to execute in OCM. In the past decades, much has been said about the importance of replication. Although welcoming further discussion on this issue, we feel it's time to act. We appreciate Schwartz and Stensaker's reference to the Rolling Stones ("You Can't Always Get What You Want"), but we also believe that there is wisdom in the song Elvis Presley recorded almost 50 years ago: "A Little Less Conversation, A Little More Action Please."

References

- Fanelli, D., & Loannidis, J. (2013). US studies may overestimate effect sizes in softer research. *PNAS*. Advance online publication. doi:10.1073/pnas.1302997110
- Levine, T., Asada, K., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: Evidence and implications of a publication bias against non-significant findings. *Communication Monographs*, 76, 286-302.
- Medin, J., Ekberg, K., Nordlund, A., & Eklund, J. (2008). Organisational change, job strain and increased risk of stroke? A pilot study. *Work*, 31, 443-449.
- Slavin, R. E., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, 31, 500-506.

Author Biographies

Eric Barends is a researcher at the VU University of Amsterdam and the director of the Center for Evidence-Based Management.

Barbara Janssen is a management consultant and a researcher at the VU University of Amsterdam.

Wouter ten Have is a management consultant and a teacher and researcher at the VU University of Amsterdam.

Steven ten Have is a management consultant and a professor of Strategy and Change at the Strategy & Change at the VU University of Amsterdam.