

---

## Is There a Fly in the “Systematic Review” Ointment?

---

MICHAEL J. BURKE  
*Tulane University*

Briner and Rousseau’s (2011) characterization of a systematic review as a penultimate procedure for evidence-based practice in applied psychology is misleading in that a systematic review is likely to unnecessarily discard useful data leading to potentially erroneous conclusions that can adversely affect workers, organizations, and the public. Furthermore, although there is much to like about Briner and Rousseau’s identification of shortcomings in some current efforts to generate cumulative knowledge within practice-oriented domains of applied psychology, their discussion confounds the usefulness of meta-analysis as a research methodology with researchers’ use of meta-analytic methods. This distinction between methodology and methodologist is important, as current meta-analytic methodology, along with some of the recommendations in Briner and Rousseau’s commentary, has considerable potential for advancing evidence-based practice beyond that of the so-called systematic review.

More specifically, in their unbridled statements that a systematic review is more practically useful than a meta-analysis, Briner and Rousseau asserted that meta-analysis (a) can only answer questions addressed many times over by researchers and (b) cannot address questions that arise from practice problems. Their statements

related to these points are not only incorrect but (c) systematic reviews often exclude studies altogether or from final analyses based on arbitrary criteria, including questionable “study quality” considerations. The manner and degree to which study inclusion/exclusion criteria are considered in a systematic review leads to potentially more biased, misleading evidence for practice in comparison to the findings from a meta-analysis in the same domain. I will comment on each of these points below.

First, a meta-analysis, like the systematic review described in Briner and Rousseau, can address questions including practice-oriented research questions that the authors of the primary studies were unable to address. The reason for this occurrence is that the meta-analytic researcher can take into account and systematically study contextual information found in the respective set of primary studies. Along with scores of other examples in the literature, in our research program alone we have examined expectations related to the influence of situational variables that heretofore were difficult to examine or could not be examined by primary researchers. Examples of our recent efforts within the health and safety domain would include meta-analyses that examined effects associated with the level of safety training engagement, the nature of workplace hazards, organizational safety climate, and national cultural values on safety training effectiveness (Burke, Chan-Serafin, Salvador, Smith, & Sarpy, 2008; Burke et al., in press; Burke et al., 2006).

---

Correspondence concerning this article should be addressed to Michael J. Burke.  
E-mail: mburke1@tulane.edu  
Address: Freeman School of Business, Tulane University, New Orleans, LA 70118

Second, meta-analyses, like systematic reviews, can and do occasionally address questions of considerable practical importance. For instance, the Burke et al. (2006) meta-analysis addressed a practical question that had been posed, in varied forms, for decades by health professionals, training specialists, union personnel, and private and public policy makers alike: What is the effect of safety training engagement on knowledge acquisition, safe work behavior, and safety and health outcomes? This question focuses explicitly on how best to train workers in relation to health and safety concerns, with the expectation that training programs that are more engaging (e.g., those that involve more dialogue, practice, and reflection) will be more effective. To underscore the practical importance of this question at the intersection of applied psychology and public health, one need only look at recent news headlines concerning worker safety training for the largest oil spill recovery effort in the U.S. history. For instance, Assistant Secretary of Labor for the Occupational Safety and Health Administration (OSHA) David Michaels was quoted as saying, "We have reports that some [in reference to British Petroleum's oil recovery safety training programs] are offering training in significantly fewer than 40 hours, showing video presentations instead of requiring hands-on training and offering only limited instruction" (Mirza, 2010). These workers' insufficient training was reflected in their behaviors: Some of the clean-up workers were initially working without personal protective equipment (such as gloves) and in their regular clothes, meaning that they were not only coming into direct contact with contaminants, but they also may have been bringing those contaminants into their households (also see Krisberg, 2010).

Third, systematic reviews follow a process (see stages of the systematic review process in Briner and Rousseau) that leads to the use of somewhat arbitrary study inclusion and exclusion criteria, including numerous assessments of study quality for entering a primary study's data/findings

into particular analyses. To illustrate the potential downside of this stringent study inclusion and exclusion process, I will contrast Burke et al.'s (2006) meta-analysis with a systematic review in the same domain by Robson et al. (2010). The Burke et al. meta-analysis included all field experimental studies with useable data (i.e., 95 studies) since the passage of the Occupational Safety and Health Act in 1971 through 2003 to address the above practice-oriented question. In addition, Burke et al. identified a key methodological "study quality" consideration: whether the effectiveness of safety training differed for supposedly higher quality between-subjects study designs that employed a control or comparison group in relation to within-subjects study designs that only had pre- and postintervention measures and conducted moderator analyses (rather than exclude the within-subjects studies) to examine the possible influence of study design on safety training–outcome relationships. Their results indicated that as the method of safety and health training becomes more engaging (going from passive, less engaging methods such as lecture to experiential-based, highly engaging methods such as hands-on training that encourage dialogue, practice, and reflection), the effect of training is greater for knowledge acquisition, safety performance, and the reduction in accidents and injuries. Furthermore, methodological quality did not have a consistent and meaningful effect on study findings, which allowed for combining study effects from within-subjects and between-subjects designs to produce more robust findings.

The conclusions of Burke et al. (2006) stand in contrast to the conclusions of the systematic review of worker safety training effectiveness recently conducted by a joint international team of researchers from Canada's Institute for Work & Health and the U.S.'s National Institute for Occupational Safety and Health (Robson et al., 2010). Robson et al.'s systematic review, which was designed in part to revisit Burke et al.'s conclusions, was a highly selective examination of only 14 safety training

studies culled from 6,469 potentially relevant studies, a fraction of the field experimental studies included in Burke et al. As emphasized in Briner and Rousseau's example (i.e., Joyce, Pabayo, Critchley, & Bambra's 2010 study in which 11,954 articles were pruned to 10 studies for final systematic review), this focus on a very restricted set of studies is ironically typical of the systematic review. And, as in a common systematic review, Robson et al. employed rather arbitrary study inclusion criteria (e.g., to only include safety training research conducted subsequent to the conclusion of the National Institute for Occupational Safety and Health's last systematic review of the safety training literature in 1996) and questionable "study quality" exclusion considerations (e.g., excluding studies with nonrandom assignment due to time constraints for their research team; excluding other studies that reviewers had, on average, lower confidence that the study provided an "unbiased estimate of the true effect" of a training intervention) in arriving at the set of 14 safety training studies. The conduct of the Robson et al. investigation, an exemplar application of the systematic review process by an international team of researchers, introduced considerable second-order sampling into the analyses and findings, unnecessarily lowered statistical power for detecting experimental effects, and precluded tests of assumptions about study quality.

Nevertheless, Robson et al. (2010) stated that "The review team concluded there is insufficient evidence of high engagement training (single session) having a greater impact on Occupational Health and Safety (OHS)-related behaviors compared to low/medium engagement training (single session)." This systematic review now serves as the official federal government research statement/advice on safety training effectiveness for workers and work organizations in Canada and the United States. Notably, in this domain of inquiry, the findings of the systematic review stand against the results of a more comprehensive meta-analytic study in the literature that

explicitly examined key assumptions about study quality and exclusion. Herein may lay the fly in the ointment: The requisite evaluation of study quality in a systematic review and misuse of study quality as an inclusion or exclusion criterion within systematic reviews can work against the quality of a systematic review.

On the other hand, I agree with Briner and Rousseau's point that evidence-based decisions require more focused and tailored reviews of evidence than what we see in a typical meta-analytic study, where both a practice question or problem and the conditions under which the evidence might be applied are taken into account. As a reviewer and editor, I have handled scores of meta-analyses in applied psychology in which the authors devoted little, if any, attention to practice concerns or the conditions under which the meta-analytic findings might be of use. Yet, this situation does not call for adopting a systematic review as described by Briner and Rousseau. This situation does suggest that greater attention be given on the behalf of all involved in the peer-review process (authors of primary and meta-analytic studies, reviewers, and editors) and in other efforts to produce cumulative knowledge to attend more to practice issues. Furthermore, the incorporation of other evidence into traditional meta-analytic reviews including qualitative data and the perspectives of those using or affected by organizational interventions is a tall order yet a useful suggestion on Briner and Rousseau's behalf.

The recommendation to examine the "best evidence" was eloquently expressed by Slavin (1986) at the advent of the meta-analytic era, and it still holds today. Importantly, Slavin's criticisms of the somewhat arbitrary limitations placed on the inclusion of studies in meta-analyses and traditional reviews of his era echo loudly in relation to the current conduct of systematic reviews, to which Briner and Rousseau desire to have us aspire. In short, it is difficult to justify a return to reviews, now under the label of "systematic review," with rather arbitrary, albeit articulated,

study inclusion or exclusion criteria and resulting reliance on very restricted data for important practice recommendations. Slavin's notion of a "best-evidence synthesis" that combines the optimal features of meta-analysis with qualitative data and varied stakeholder input would appear to still provide the most promising means for generating "best evidence" and "best practice" in applied psychology.

## References

- Briner, R., & Rousseau, D. M. (2011). Evidence-based I–O psychology: Not there yet. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 4, 3–22.
- Burke, M. J., Chan-Serafin, S., Salvador, R., Smith, A., & Sarpy, S. (2008). The role of national culture and organizational climate in safety training effectiveness. *European Journal of Work and Organizational Psychology*, 17, 133–154.
- Burke, M. J., Salvador, R., Smith-Crowe, K., Chan Serafin, S., Smith, A., & Sonesh, S. (in press). The dread factor: How hazards and safety training influence learning and performance. *Journal of Applied Psychology*.
- Burke, M. J., Sarpy, S. A., Smith-Crowe, K., Chan-Serafin, S., Salvador, R. O., & Islam, G. (2006). Relative effectiveness of worker safety and health training methods. *American Journal of Public Health*, 96, 315–324.
- Joyce, K., Pabayo, R., Critchley, J. A., & Bambra, C. (2010). Flexible working conditions and their effects on employee health and well-being. *Cochrane Database of Systematic Reviews*, Art. No. CD008009. DOI: 10.1002/14651858.CD008009.pub2.
- Krisberg, K. (2010). U.S. gulf spill poses public health threat: Response targeting workers, residents, food and air quality. *The Nation's Health*, 40, 1–8.
- Mirza, B. (2010). *OSHA: BP responsible for oil spill worker training, protection*. Alexandria, VA: Society for Human Resource Management. Retrieved June 6, 2010, from [www.shrm.org/hrdisciplines/safetysecurity/articles/Pages/OSHABP.aspx](http://www.shrm.org/hrdisciplines/safetysecurity/articles/Pages/OSHABP.aspx)
- Robson, L., Stephenson, C., Schulte, P., Amick, B., Chan, S., Bielecky, A., Wang, A., ... Grubb, P. (2010). *A systematic review of the effectiveness of training and education for the protection of workers* (DHHS (NIOSH) Publication No. 2010-127). Toronto: Institute for Work and Health; Cincinnati, OH: National Institute for Occupational Safety and Health.
- Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, 15, 5–11.