Elwood F. Holton III

I appreciate Kirkpatrick's reaction to my article "The Flawed Four-Level Evaluation Model." In the spirit of advancing the profession, I will respond to the two issues he raised that relate to practice and research in the profession: the distinction between a model and a taxonomy and the value of evaluation research.

Kirkpatrick states that he never called the four-levels framework "a model" and that he doesn't care whether it is called a model or a taxonomy, suggesting that the distinction is merely one of semantics. But the distinction between a taxonomy and a model is crucial for both practice and research. If the four levels are truly proposed as a model, then there should be relationships among the levels. Kirkpatrick's response to my article clarifies that he does indeed believe there are. This is helpful because, as I discussed in my article's introduction, his writing has been unclear in this area. If the relationships he refers to are the linear relationships among the four levels he has suggested (see Kirkpatrick, 1994, p. 27 and p. 51, for example) then, unfortunately, the research does not support them. There are variables and relationships missing from the four-level framework, resulting in an underspecified model (Campbell, 1990). If simple linear relationships among the levels are not intended by Kirkpatrick, then a clear specification of intervening variables and relationships is needed. Finally, for the record, he has labeled it a model: "The reason I developed this four-level model..." (Kirkpatrick, 1994, p. xiii).

I do not intend to demean the value of taxonomies, which are quite appropriate for intermediate stages of theory development (Bobko and Russell, 1991). They are very useful, and I have developed them myself (Holton, in press). However, it is important to realize that they are just an intermediate stage in making complex phenomena more understandable. Therefore, whether considered a model or a taxonomy, the four-level evaluation framework needs updating. If it is a model, it is underspecified; if it is a taxonomy, then a true model is needed.

This issue has critical implications for practitioners, even though they incorrectly use the terms synonymously. The purpose of evaluation is to make...
decisions about human resource development (HRD) intervention effectiveness and to decide on a course of action to change an intervention if it is not effective. Taxonomies, by definition, do not provide information about causes. Fully developed models do provide causal information, but underdeveloped models lead to wrong decisions or confusion.

Suppose a practitioner finds that learning outcomes in a training program are acceptable but the learning is not used on the job. Then what? Should the training program be canceled? Or should a meeting be called with the supervisors to have them reinforce the training? Should the training be redesigned to include more practical examples? Or should goal setting be included to increase trainee motivation to transfer? The practitioners I work with need to make these kinds of decisions and do not have the tools to do so, even if they implement the four levels.

This issue leads to the second issue I wish to address: the value of evaluation and HRD research. Kirkpatrick suggests in his response that the four levels must be correct because they are widely used. History has shown that the extent to which something is practiced is no indication that it cannot be improved upon. He further seems to suggest that scholarly research articles (such as those published in *Human Resource Development Quarterly*) are not practical or helpful. However, to quote Kirkpatrick, "Progress in evaluation of training will result if all of us will freely exchange information on objectives, methods, and criteria" (1960, p. 17). Isn't that what research does? Is he now suggesting that we should not capitalize on the excellent research that has been conducted to improve upon the four levels by building a more fully specified model? Is he suggesting that we should not conduct research on the four levels because they have been widely used for thirty-five years, essentially without validation or modification? Is he suggesting that the work of scholars in HRD is not important if that work is not first published in *Training and Development*? I hope he doesn't mean to suggest any of these things. Certainly some articles in *HRDQ* may need further refinement before widespread application, but "there is nothing so practical as good research" (Passmore, 1984, p. 24). I have confidence that practitioners in our field are capable of understanding and using models that depict the complex world of human performance in which we work, particularly if they lead to more effective HRD interventions.

In conclusion, if I appear not to have given full credit to Kirkpatrick for the contributions the four levels have made in the thirty-five years since they were first published, then let me acknowledge them right now. As Newstrom (1995) noted, Kirkpatrick is the person who long ago focused practitioners on evaluating outcomes. His four-level framework will always be a classic in the field. But, as Newstrom further noted, HRD research has advanced to the point that we can improve upon the four levels. Furthermore, the complexity of performance improvement practice today demands that we do so. I hope that we can continue to debate and discuss new approaches as scholars in our shared pursuit to improve HRD practice continuously.
References


Elwood F. Holton III is assistant professor of human resource development, Louisiana State University, Baton Rouge.