

# What Works and What Doesn't: Evaluation Beyond Kirkpatrick

by Roger Kaufman, John Keller, and Ryan Watkins

Roger Kaufman is professor and director of the Center for Needs Assessment & Planning at Florida State University and a research professor at New Jersey Institute of Technology/ Newark College of Engineering. He may be reached at (904) 644-6435 or 644-1708, by FAX at (904) 644 -9647, or e-mail: [rkaufman@cnap.fsu.edu](mailto:rkaufman@cnap.fsu.edu).

John Keller is Professor of Educational Research with a specialty in the Instructional Systems program at the Florida State University. He may be reached at the Department of Educational Research at 206 Dodd Street, Tallahassee, Florida 32306.

Ryan Watkins is a graduate student in the Instructional design program at Florida State University as well as a graduate research assistant at the Center for Needs Assessment and Planning. He may be reached at the Center at 206 Dodd Street, Tallahassee, Florida 32306.

Our professional future depends on doing the right thing at the right time, but most people cringe when a conversation shifts to evaluation. Like it or not, finding out what has worked and what has not is essential for human performance improvement and organizational success. Without evaluating, we don't know which performance interventions to stop, modify, continue, or improve.

Evaluation is important, pragmatic, and practical even though it carries both promise and threat. The promise that comes from determining what works and what does not prompts you to continuously improve. The threat, which flows from misuse, comes from the fear that performance data will be used for blaming and not for fixing or learning. One must realize the promise of evaluation, however, as well as the information it can provide for continuous improvement. We can use it wisely to answer the questions that face every organization.

Unfortunately, evaluation is an underused and misunderstood process that is often inappropriately or unwisely applied, thus yielding harmful results. There are at least five reasons for this:

- The definitions and models of evaluation that we use are often too restrictive.
- We do not ask the right questions.
- We do not phrase the questions appropriately.
- The relationship between ends and means is not made clear in the planning, implementation, and evaluation processes.

- The data from evaluation is not used for continuous improvement.

An effective evaluation approach should be based on a paradigm that has been expanded beyond our conventional frameworks. A useful evaluation framework would identify questions that concern results for evaluation programs and interventions as well as for accurately assessing individual performance (Westgaard, 1993). These questions would be keyed to important, practical, and specific objects and relate directly to those who receive our deliverables.

### What Is Evaluation?

Literature concerning evaluation is bountiful, and definitions of evaluation range from supplying information to decisionmakers to comparing results with intentions. We prefer the latter. Supplying information to decisionmakers is important, but it is only part of the evaluation story. To complete the story, it is necessary to ask some questions:

- How do you evaluate the results of the decisionmakers after information has been supplied to them?
- How do you determine whether the decisionmakers have results-oriented information related to all the vital questions they and their organization must ask and answer?
- How do you use performance data for continuous improvement rather than as snapshots of a dynamic process?

Going beyond the definition of evaluation as “supplying information to decisionmakers,” we suggest that all evaluation questions compare results with intentions and delve into the usefulness of methods and resources so that we may move toward the required results (termed “quasi evaluation” by Kaufman, 1992). Evaluation-related questions include:

- How well have we met our precise and rigorous objectives?
- What is the value of the methods and resources that were used for getting required results?

### Kirkpatrick Plus: Beyond Training Evaluation

For more than 30 years the most popular, and probably the primary, model for evaluation among human-resource development professionals has been the four-level training evaluation

Another reason for applying the levels of evaluation outside of training is to incorporate other performance improvement interventions. For example, organizational development, strategic planning, quality management/continuous improvement, career pathing, diversity, integrated work teams, and

---

## **We suggest an expanded framework that ensures that what we do and what we deliver are worthy both within and outside of the organization.**

approach proposed by Kirkpatrick in 1967. Kirkpatrick provided a rational and accepted, though sometimes criticized, four-level evaluation framework as follows:

- Level 1: Reaction (how learners feel about instruction);
- Level 2: Learning (learner performance on in-class tests);
- Level 3: Behavior (extent to which learners implement, or transfer, what they learned in class); and
- Level 4: Results (organizational benefits, stated in terms of organizational performance or return on investment derived from a course).

This framework reflects important and useful areas to assess. We suggest that these levels are incomplete, however, in terms of assessing performance and consequences, and as such they have encouraged many to focus narrowly on evaluating training (Kaufman and Keller, 1994). This focus is too restrictive and often produces laments about the difficulty of estimating use on the job (Level 3), organizational performance, and the worth of the organization’s deliverables (Level 4) because of the complex factors influencing transfer and organization.<sup>1</sup> Because of the perceived complexity of all but the first level, Kirkpatrick’s four levels have not been widely used despite their overall appeal.

mentoring can be included in evaluation designs to assess how effectively they have been implemented (an aspect of Level 3 evaluation) and whether they have positive organizational results (Level 4 evaluation). Including interventions other than training into the general evaluation design will require some modification to the four levels while incorporating existing aspects of program evaluation.

We suggest an expanded framework that builds on the traditional four-level training evaluation framework, but adds useful considerations to ensure that what we do and deliver are worthy both within and outside of the organization (Kaufman and Keller, 1994). We propose that the principles and processes of training evaluation be expanded to consider all interventions associated with strategic and tactical planning, performance improvement, organizational development, customer satisfaction/total quality, and societal contributions. We add levels of impact because the traditional four-level approach targets only training and does not adequately address all of the key evaluation questions organizations face.

Training is only one performance improvement avenue, although it should be (but usually is not) based on

contributions to performance inside as well as outside the organization.

### Questions Every Organization Should Ask—And Answer

We must define and justify the ends before we can ask and answer sensible evaluation questions. Previous formulations of ends, often called objectives, often focus on self-serving organizational purposes rather than results that contribute to external client and societal usefulness. Kaufman proposes the following questions any organization, public or private, must ask and answer (Kaufman 1992, 1995; Kaufman and Grise, 1995):

- Do you commit to delivering organizational results that will have beneficial impact, consequences, and payoffs for external clients, including positive payoffs for society?
- Do you commit to delivering timely organizational outputs that are of desired quality to external clients?
- Do you commit to the timely delivery of the quality products that individuals or small groups within your organization require?

When attempting to answer these three basic questions, we must also consider that there are three levels of planning and results based on who the primary clients and beneficiaries of the deliverables are. The categories are as follows:

- Mega: the primary client and beneficiary of what any organization (public or private) delivers is society now and in the future;
- Macro: The primary client and beneficiary of what is planned and delivered is the organization itself; and
- Micro: the primary client and beneficiary of what is planned and delivered is an individual or small group.

The first question, a mega-level question, targets the societal usefulness of what an organization delivers and its impact on the external world. Issues of health, continued profits, pollution levels, safety, and well-being are central. The basis for mega-level concerns is an ideal vision, which is a measurable

statement of the kind of world required for the health, safety, and well being of tomorrow's children. The second question, a macro-level question, concerns satisfying the client's expectations and standards and thus focuses on the organization's short-term success and market share. The third question, a micro-level question, is concerned with the expectations and standards within an organization as products move from one group to another before being deliverable (such as training results, an advertising campaign, the progressive treatment of a patient from intake to discharge) outside of the organization (quality outputs to clients, increased sales and satisfied customers, discharged patients who return for the same medical reasons).

These three questions and their three levels all focus on results, but we know that results come from resources that are converted through processes, methods, and means. Here are two important questions concerning means and resources:

- Do you commit to providing efficient and acceptable methods and procedures to be used by an individual or small group within the organization?
- Do you commit to the availability or quality of resources that will be used by an individual or small group within the organization?

Each of these five questions, regardless of whether they are formally recognized, must be answered. If we address all levels, evaluation will provide data to determine whether we have achieved our goals and what might be done to improve our performance. These five questions provide the basis for defining ends and relating them to means and resources. They in turn provide the basis for designing the various levels of evaluation, which leads to our third suggestion: All levels of evaluation planned and designed prior to implementing an intervention and all levels of evaluation should be defined in terms of expected outcomes, meaning results and positive consequences at the mega level.

For example, before deciding whether there is an organizational problem important enough to warrant developing and implementing an intervention, we ought to identify the expected organizational changes and consequences that will occur from the intervention and estimate the costs of the investment. Likewise, before choosing a particular performance improvement intervention, we ought to identify the expected changes in intervention. We should identify how the changed performances will be observed and what the magnitude of internal and external (including societal) consequences will be in relation to the change.

In addition to planning for Kirkpatrick's four levels, there are more evaluation levels that are implied by our list of results and questions concerning ends and means. We urge that these be added to the conventional levels of evaluation.

### Four Levels and Five Questions

Figure 1 compares Kirkpatrick's four levels with the basic organizational questions and foci we suggest. Viewed this way, Kirkpatrick's four-level evaluation framework is missing mega-level evaluation. It is also missing the determination of the usefulness and availability of organizational resources.

Kirkpatrick's fourth level, organization, refers to organizational performance. It refers to the organizational benefits of an intervention, but not its societal costs and consequences. We believe that return on investment should remain within the context of Level 4 evaluation along with other organizational results, as in Kirkpatrick's formulation. It should be applied to societal return on investment as well.

Performance improvement specialists and their supervisors must be concerned with means and resources such as the availability and quality of assets, personnel, materials, funds, and the efficiency of their use. Traditional Level 1

Kirkpatrick's Levels	Organizational Questions/Focus	Compatibility	Suggested Levels*
	F. Societal contributions	Missing from Kirkpatrick's framework	Mega
4. Results	E. Organizational payoffs	Yes	Macro
3. Behavior (performance)	D. Individual or small group payoffs	Yes	Micro
2. Learning (acquisition)	C. Individual or small group payoffs	Yes	Micro
1. Reaction	B. Process acceptability and efficiency	Partial; efficiency and quality are missing from Kirkpatrick's framework.	Process
	A. Resource availability and quality		Input
			*Based on Kaufman's Organizational Elements Model (1992, 1995).

Figure 1. A comparison of Kirkpatrick's four-level evaluation framework and the five levels of critical organizational concern and evaluation (based in part on Kaufman and Keller, 1994).

evaluation focuses on whether participants have positive attitudes toward the content of instruction and the quality of delivery. We suggest expanding Level 1 to include the efficient use of the quality resources available. Level 1 evaluation will become more rigorous and useful and its application will extend to contexts other than training.

Kirkpatrick's framework has, under most circumstances, served conventional evaluators quite well, especially those concerned only with micro-level targeted interventions such as training. As seen in Figure 1, concern for a resource's societal impact and appropriateness is not considered. We suggest that mega-level results wherein society is the primary client and beneficiary are vital for all organizations (as recently supported by Drucker, 1993; Popcorn, 1991; and Marshall and Tucker, 1992). A missing fifth level should be added to the conventional four: societal consequences and payoffs.

We also suggest expanding Level 1 to include the value and worth of resources, methods, and tools and the efficiency of their use. Because evaluation best compares results with expectations, consideration of the usefulness and appropriateness of our resources

and methods is also relevant and important. In fact, the "happiness" displayed by learners in Level 1 evaluations may depend mightily on their perceptions of the value and worth of the methods and resources used.

Assessing Level 1 results will depend on the objectives regarding those affected by the intervention. This expansion of Level 1 would include the role, usefulness, appropriateness, and contributions of methods and resources. Levels 2 and 3, in keeping with the suggestions of Keller and Stevens (1988), can also be redefined to include more than training interventions. In this expanded view of evaluation, Level 2 becomes more an assessment of acquisition, yet it continues to be an evaluation that depends on training being implemented properly as shown by the resultant learning. With other interventions it focuses on whether the intervention was implemented properly based on the objectives.

Level 3 evaluation becomes application based on actual performance instead of on transfer. It still identifies the training's level of transfer, but the emphasis is on whether the newly acquired knowledge and skills are being applied on the job. As with interventions, it applies to whether the training is working as

intended. For example, we could assess a new career pathing process in terms of whether people perceive it to be a desirable program that uses appropriate amounts of resources (Level 1); whether it is being implemented according to its design and objectives as measured by acquired skills, knowledge, attitudes, and abilities (Level 2); and whether it succeeds in getting people into the appropriate career paths (Level 3).

The purpose of Level 4 evaluation, according to our suggestions, would remain essentially unchanged. The goal of this level is to assess organizational results in terms of benefits, performance improvements, and cost benefit/costs consequences analysis, such as timely, quality outputs.

Level 5, the additional level that we suggest, adds a vital, practical concern for the contributions an organization must make to its clients and society. Is what we deliver contributing to the good of society in general as well as satisfying to the client?

This broader view of the levels of evaluation should make evaluation more responsive to the questions organizations and their decisionmakers must face. The suggested five levels of

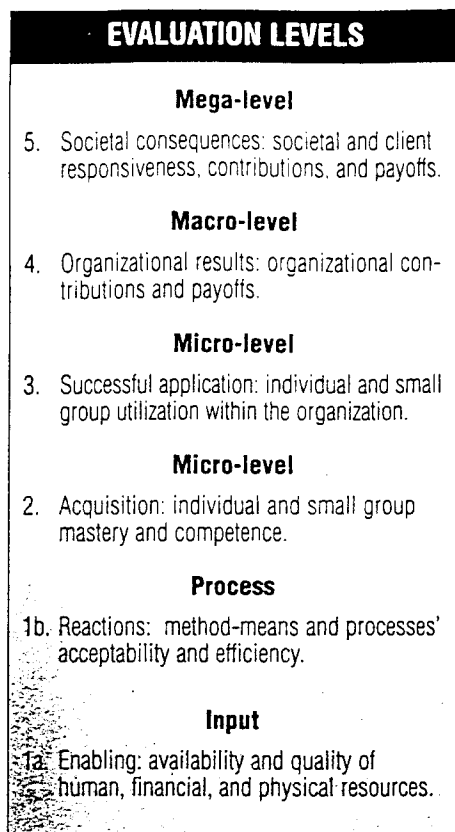


Figure 2. Five levels of evaluation of human performance improvement interventions.

evaluation, including an expanded first level, are shown in Figure 2. By expanding Kirkpatrick's four-level evaluation formulation, we derive a holistic and realistic framework. This adds practicality to the formulation and allows the framework to be useful not only for training, but for any organizational intervention. The "evaluation plus" framework suggested here will deliver a way to determine whether useful results were achieved, whether the means and resources used were efficient and appropriate, and whether there was a proper and useful alignment between what the organization used, produced, and delivered. ■

## References

Alliger, G., and E. Janak, "Kirkpatrick's Levels of Training Criteria: Thirty Years Later." *Personnel Psychology*, vol. 42, no. 3 (1989).

Drucker, P.F., *Post Capitalist Society*. New York: Harper Business, 1993.

Hartz, R., R. P. Niemeck, and H. J. Walberg, "The Impact of Management Education," *Performance Improvement Quarterly*, vol. 6, no. 1., (1993).

Hamblin, A.C., *Evaluation and Control of Training*. London: McGraw-Hill, 1974.

Hepworth, A., "Evaluation Research in Perspective," *Management Education and Development*, vol. 3, no. 2, (1972) pp. 49-62.

Kaufman, R., *Strategic Planning Plus: An Organizational Guide*. Newbury Park, CA: Sage (Revised, 1992).

Kaufman, R., and D. Zahn, *Quality Management Plus: The Continuous Improvement of Education*. Thousand Oaks, CA: Corwin Press (1993).

Keller, J.M., and J. Moller, *Evaluation In Training: Much Lauded, Seldom Applauded*. Proceeding of the Joint Conference of the International Federation of Automatic Control and the International Federation for Information Processing, Leiden, The Netherlands.

Keller, J.M., and G. Stevens, *Evaluating the Organizational Results of Training*. Department of Education Research, Florida State University, Tallahassee, FL: Occasional paper.

Kirkpatrick, D., "Evaluation." In R.L. Craig and L.R. Bittel (Eds.), *Training & Development Handbook*. American Society for Training and Development, New York: McGraw-Hill (1967).

Marshall, R., and M. Tucker, *Thinking for a Living: Education and the Wealth of Nations*. New York: Basic Books (1992).

Popham, W.J., (Ed.), *Evaluation in Education*. Berkeley, CA: McCutchan (1974).

Popcorn, F., *The Popcorn Report*. New York (1991).

Westgaard, O., *Good Fair Tests: Test Design and Implementation*. Amherst, MA: HRD Press (1993).

## Endnote

1. See Kaufman and Grise, 1995, for a costs-consequences framework.