Dye's Two-Tier Framework of Evaluation:
A Practical Approach for Determining the Effectiveness of HRD Activities

In 1959, Don Kirkpatrick [1, 2, 3, 4] proposed a four-step process for evaluating training interventions: Reaction, Learning, Behavior, and Results. These four steps have become known as the "Four Levels." Forty years after its introduction, the Four Level framework dominates thinking and practice of evaluating human resource development (HRD) activities.

Despite its popularity, practitioners struggle with its usefulness as a means for articulating the value of HRD activities. Several factors contribute to this dilemma: business has changed, and the role of HRD within an organization has changed. Specifically:

1. Business changes such as globalization, technology, accelerating speed of change.
2. Changes in work requirements, such as the expectation of just-in-time responsiveness, "24/7" productivity, cross-training, and the shift from manufacturing to service / information.
3. Acceptance that HRD is a subsystem within the larger open system of a business; serving the greater business purpose and integrally linked to the business.
4. Desire of HRD professionals to be strategic partners with business managers.

The Traditional View of HRD Evaluation (The Four Levels)

The traditional view of HRD evaluation, as proposed by Kirkpatrick in 1959-60, evolved in a business context significantly different from the current context. Like all business processes, it grew from a set of assumptions that were likely true at the time. As the Four Levels became more widely accepted over time, the set of assumptions evolved as well. The following list provides examples of these assumptions.

- The terms "evaluation" and "evaluating effectiveness" are synonymous and can be used interchangeably.
- Organizational decision making, including decisions about HRD activities, is a rational process based on objective facts.
- One approach to evaluating HRD effectiveness (the Four Levels) is sufficient for all purposes; all HRD interventions can be evaluated exactly the same way.

---

1 - Kirkpatrick, D.L., *Techniques for Evaluating Training Programs.* Journal of ASTD, 1959. **13**(11): p. 3-9.

2 - Kirkpatrick, D.L., *Techniques for Evaluating Training Programs: Part 2 - Learning.* Journal of ASTD, 1959. **13**(12): p. 21-26.

3 - Kirkpatrick, D.L., *Techniques for Evaluating Training Programs: Part 3 - Behavior.* Journal of ASTD, 1960. **14**(1): p. 13-18.

4 - Kirkpatrick, D.L., *Techniques for Evaluating Training Programs: Part 4 - Results.* Journal of ASTD, 1960. **14**(2): p. 28-32.

- Stakeholders in the HRD intervention and in its evaluation share the same agendas and outcomes for the intervention and its evaluation; all parties share a similar view of "success".
- Positive results will best serve all the agendas; positive results are more valuable than negative or mixed results.
- Outcomes of each of the Four Levels are related in a linear and direct manner; positive reaction leads to positive learning which leads to positive transfer which leads to positive results.
- All four levels are necessary for effective HRD evaluation.
- Transfer (specific and immediate application on the job) is the only relevant outcome of HRD activities.
- Each "level" of outcomes is more meaningful and valuable.
- Relevant categories of outcomes (reaction, learning, etc.) are the same for all interventions and purposes.
- The quantitative and/or financial expression of outcomes is most effective and most valuable.
- Self-report data is adequate.

Problems with the Traditional View

These assumptions may, in fact, be true in specific instances of evaluating the effectiveness of HRD activities. The difficulty, however, is that the profession, collectively, believes that they are always true in every instance. Most practitioners aren't consciously aware of these assumptions, yet they underpin our profession's literature about evaluating the effectiveness of our work. The danger of blindly accepting these assumptions is that our efforts to evaluate and improve our work will be off-target, incomplete, misleading, or even inaccurate. Seven major problems embedded in these assumptions are:

1. Misuse of the terms "evaluation" and "evaluating effectiveness"
2. Failure to explicitly address the different purposes for evaluating HRD activities
3. Narrow view of stakeholders and their agendas
4. Confusion about what the term "level of analysis" really means
5. Outdated range of subjects for HRD evaluation
6. Vague and often inaccurate definition of the intended outcomes of an HRD intervention
7. Insufficient research methods (definition of units of analysis and tools for understanding them)

The first, and most fundamental, problem with the traditional framework is misuse of two similar terms: "evaluation" and "evaluating effectiveness." The word "evaluation" is a noun that describes a complex business process of determining worth or value, and the phrase "evaluating effectiveness" is a verb phrase describing one means of comparing outcomes to intentions. Evaluation is a much larger process than evaluating effectiveness. Evaluation often considers, as one factor in determining worth or value, the

effectiveness of an action. In contemporary business situations, evaluation is often a complex social process that includes technical facts along with political agendas. To ignore this dimension of evaluation would be naïve if not irresponsible.

The second problem with the traditional view of HRD evaluation is its failure to explicitly address the different purposes for evaluating our work. Three purposes are stated in the original Kirkpatrick work: improving the product (a training program), demonstrating the value of the program or the training department, and helping the trainer feel good. Current terminology for these would be, respectively, formative evaluation (continuous improvement), summative evaluation (the overall worth), and impression management (how an entity is perceived in the organization). Contemporary business needs may add a fourth reason to evaluate: certification of workers, processes, or work groups. All four purposes are legitimate, but each calls for a slightly different approach. This is especially true in the American workforce if there are job actions (hiring, promotion, termination) that result from any of these types of evaluation. When evaluation is for continuous improvement, the most important data is about errors or exceptions. It is important to know what works well, but improvement means correcting errors. When evaluation is for determining overall worth, the most important data is a comparison of outcomes to intentions, including political outcomes and intentions as well as technical outcomes and intentions. It is important to know the technical component (to what degree were skills improved?), but organizational sensemaking is more than a simple technical answer. It is a social process that considers political agendas, organizational outcomes and relationships, and negotiation. Impression management is largely ignored in the explicit discussion of HRD evaluation, but it is a strong undercurrent in the reasoning for all evaluation. When evaluation is for certification, it is important that the measures be trusted by outside agencies, including the legal or regulatory system.

The third difficulty with the traditional view of evaluation is a narrow definition of stakeholders and their agendas. This is closely linked to the simplistic view of the purposes of evaluation. This position assumes that learners, managers, HRD practitioners, and senior executives are the relevant stakeholders in the intervention and its outcomes, and that these constituent groups all share the same desire for positive outcomes of both the intervention and its evaluation. This may be true in situations where the organization has previously discussed and negotiated its evaluation strategy and where there is agreement about and confidence in the evaluation processes. More often, however, evaluation of HRD activities does not define stakeholders based on the subject and purpose of the evaluation, nor does it consider the full range of potential stakeholders (internal and external) and their possible agendas (positive, neutral, negative).

The fourth difficulty with the traditional approach to HRD evaluation is inaccurate use of the term, "level of analysis." I suspect that in its original usage, the term was used generically to mean "type of analysis." Unfortunately, HRD practitioners are using a precise research term in a generic and incorrect

manner. This practice is confusing, and it perpetuates the myth that research and practice are unrelated. The fields of business and organizational research use the term "level of analysis" to refer to groupings of individuals within the organization of study. Four levels are discussed: micro level (individual), meso level (group), macro (organization), and mega (organization and its environment). Theory development in the social sciences (including business and organizational research) warn against the temptation to aggregate upwards the outcomes of these levels of analysis, the practice recommended in the Four Level approach to evaluating HRD activities.

The fifth difficulty, outdated range of subjects for HRD evaluation, is a natural outgrowth of the evolution of our profession. In 1959-60, training was a relatively new field, training was likely the only (or predominant) HRD intervention, and the types of training were relegated primarily to specific skills training or generic management training. The literature on evaluation practice, including research journals as well as practitioner "how to" books, remains focused on techniques for evaluating these types of interventions. While these types of HRD interventions are still a significant part of HRD work, there are approximately 50 interventions currently recognized as HRD interventions. More importantly, business professionals are interested in evaluating people (participants, managers, HRD practitioners) and processes (work processes, design and development processes, transfer processes, etc.) as well as interventions.

Closely related to the problem of subject of evaluation is the narrow understanding of the outcome of the intervention. If skills and management training are the predominant subjects, then transfer (specific and near-term use on the job) are reasonable outcomes. Transfer, as commonly understood, is an inappropriate outcome for many of the newer HRD interventions, for example decision making, innovation, and diversity sessions. The outcomes of the intervention must be determined by the original assessment process used to determine the intervention. Many, if not most, HRD interventions are "prescribed" by soft assessment – the perception of management that a problem or opportunity exists. Since effectiveness is a comparison of outcomes to intentions, this may be problematic. The final comparison is only as rich as the leanest type of data. Specifically, if perception data drives the choice of intervention and its intended outcomes, then the organization is limited to using perception data to compare outcomes to intentions. Organizations that want quantifiable and precise outcome measures must be willing to invest in quantifiable and precise front-end assessment.

The seventh, and final, difficulty with the traditional method of evaluating HRD activities is the limited range of tools / methods in use. Common practice in HRD evaluation is consistent with common practice in most organizational research: heavy use of self-reported opinion data about reaction to the intervention, perceived value of the intervention, perceived changes in the workplace, and perceived contribution to the organization. There is also some use of objective testing (paper and pencil, completion of

job tasks) and observation (role plays). Most evaluation could be classified as quantitative (as opposed to qualitative) in its approach. There is very little precise description of methods in the literature. Self-report opinion data results in perceptions of outcomes and perceptions of effectiveness. This isn't necessarily wrong, but it will disappoint an organization that requires objective data.

A Two-Tier Framework for Developing a More Effective Evaluation Strategy

If we accept the importance of differentiating the complex social process of evaluation from the more technical process of evaluating (measuring) effectiveness, it is clear that an organization must first understand its context for evaluating the HRD intervention before it designs specific studies to measure effectiveness.

Given this distinction, it is helpful to consider two groups of factors that contribute to truly useful practices for evaluating the effectiveness of HRD interventions. One set of factors, termed *evaluation factors*, deals with the organizational sense making dimension of evaluation. The second set of factors, termed *effectiveness factors*, deals with the technical components of assessing the intervention. The following table presents the elements of each set of factors.

Evaluation Factors	Examples of Viable Options
• What is the subject of the evaluation?	• Processes • People • Interventions
• What is the purpose of the evaluation?	• Formative evaluation (continuous improvement) • Summative evaluation (overall effectiveness) • Certify workers, processes, products • Impression management
• Who are the stakeholders in the evaluation?	Internal • HRD professionals • Management (various levels) • Participants External • Union officials (various levels) • Customers • Regulators
• What are the agendas of these stakeholders?	• Positive • Neutral • Negative

Effectiveness Factors	Examples of Viable Options
• What is the appropriate level (or levels) of analysis?	• Micro level (individual worker, intervention, or process) • Meso (groups of workers, interventions, or processes) • Macro (the organization, its interventions, or its processes) • Mega (organization in relation to its environment)
• What are the intended outcomes of the subject of the evaluation?	• Specific outcomes expressed immediately • Specific outcomes expressed in the future • General outcomes expressed immediately • General outcomes expressed in the future
• What are the specific units of analysis?	• Individual readiness for the intervention, skill change, design of the intervention, etc. (micro) • Effectiveness of an HRD group's service, changes in departmental productivity, etc. (meso) • Organizational performance, organizational stability, total investment in HRD, etc. (macro) • Marketplace value of training, measured by a company's share price (mega)
• What are the appropriate research methods and measures?	• Overarching evaluation paradigm • Data collection methods • Data analysis methods

Process for Developing Effective Evaluation

The process for developing effective evaluation activities loosely follows the two tiers of factors. HRD managers or practitioners will most likely initiate and drive this process. It involves two phases:

- Phase 1 articulates what the organization needs in order to make a value judgment about HRD activities.

- Phase 2 sets forth a specific plan to capture or "measure" only that data the organization needs.

(NOTE: the term "needs" will be used throughout the remainder of this document to mean both "wants" and "needs." There may be real and material differences in what these terms mean in any given context, but the one term will be used here simply for convenience.)

Phase 1: Develop an HRD Evaluation Plan

Phase 1 depends on understanding what the organization needs so that it can make informed judgments about the worth or value of HRD. The scope of the evaluation plan could cover anything from the entire HRD function and all its products / interventions, processes, and people to only one or two specific courses or interventions. Carefully and explicitly explore the following questions; and pay attention to clues and indirect information from the organization.

- What does the organization currently evaluate? What does it measure?
- What is the purpose of evaluation? How is the information used?
- Who participates in evaluation? Who measures? Who gets the information? Who cares about the results?
- How effective is current practice? What requests do we get? Complaints?
- Who are real stakeholders in HRD evaluation? What are their agendas? What do *they* want/need to know about?
- How will stakeholders and/or the organization use the information?
- What other considerations are relevant?

The key outcome of this phase of work is to understand what the *organization* wants regarding evaluation – not what HRD thinks the organization wants. Practitioners who remain bound by the Four Level framework (reaction, learning, behavior, results) are likely to remain outside the key decision processes of the organization. HRD practitioners may well need a specific and more detailed set of measures to evaluate specific elements of its work, but these internal processes are not always of interest to the key stakeholders. Know the difference between evaluating internal processes and evaluating organizational issues. Giving stakeholders exactly what they want increases the likelihood that they will engage in a meaningful way.

The needs of the organization should be recorded in a draft document that is then circulated among stakeholders. Engage them in conversation about the accuracy of the organizational context, about their individual and collective needs, and about the completeness or redundancy of information. The underlying question is *What does the organization need to know to make a judgment about the worth or value of HRD?* Organizations frequently make value judgments without measuring anything. Revise the document so that it accurately captures the collective needs of the organization. This often-overlooked step increases buy-in among stakeholders. Remember that evaluation is about determining the worth or value of something – not just about measuring its effectiveness.

The final output of Phase 1 is a document, accepted by the stakeholder group, called the Evaluation Strategy. It should articulate the stakeholders and their needs, and it should provide strong and clear direction for the types and frequency of specific measures. Note that the strategy document is specific for whatever

scope was defined, and that complex organizations may have several interdependent or independent evaluation strategies.

Phase 2: Develop a Measurement Plan

The Measurement Plan follows directly from the Evaluation Plan. It sets forth the specific data the organization needs in order to demonstrate the effectiveness of specific HRD activities. Remember that effectiveness is a comparison of actual outcomes to intended outcomes. Depending on the needs of the organization, this data may be quantitative (hard) or qualitative (soft) data, or a combination of both. It is worth repeating that HRD professionals may need a more detailed type of data for internal evaluation. This data may not be of interest to other stakeholder groups and should not be "forced" onto those who aren't interested.

While specific measurement plans will vary greatly and will depend on the specific needs of individual organizations, two trends have surfaced:

1. Few organizations use rigorous, quantitative methods for assessing HRD needs. Many organizations use "soft" assessment data such as self-report of perceived need, and many organizations don't systematically conduct needs assessments at all. This limits the methods for establishing the effectiveness of the activity: no matter how precise the outcome measures are, the overall comparison can only be as rigorous as the input (assessment) data.
2. Few stakeholders care about internal HRD processes; they care, instead, about macro "accounting" factors such as cost per employee per hour of training or fair market value of internal services.

Many organizations use a "scorecard" concept to quickly and accurately track measures of effectiveness for business processes, for example sales and marketing. This concept works well for reporting HRD effectiveness. Give the organization what it wants – no more and no less. This process may require HRD to partner with MIS / IT professionals to capture and report data that already exists within the organization's databases. Discuss the plan with stakeholders and verify that it provides complete and accurate information. If possible, prepare a draft or mock-up of the report and circulate it among the stakeholders or stakeholder groups (if specialized reports are needed). The right handful of indicators will be far more meaningful than pages of the wrong indicators.

Practical Example: Applying the new Framework for HRD Evaluation

This example discusses Big Bank, a large, fictitious, regional bank with a strong focus on shareholder earnings. The firm boasts 25 consecutive years of double-digit shareholder returns. Strong cost controls and a decentralized entrepreneurial spirit drive the success of the firm.

Big Bank has a unique culture that it imposes onto all its acquired firms. Big Bank executives understand that there are costs associated with this strategy and that, on occasion, the organization has converted a more efficient acquired system to the less efficient Big Bank system. This has occurred with hard as well as "soft" systems. Costs associated with both employee and customer turnover are considered in the overall cost of acquisition and conversion.

Big Bank's aggressive growth strategy and corporate pressures for operating efficiencies necessitate that the organization have an effective and efficient HRD group. This group plays a central role in integrating new acquisitions as well as providing for ongoing employee training and development at all levels of the organization (entry level through executive).

Big Bank has developed a standard and highly effective program to train employees from acquired firms on its Customer Service processes and equipment. The 5-day training program accommodates two 3-hour custom-developed segments. Big Bank has used the same training program for five years, and modifies the approach and content only when there are changes to work processes. This training is critical to the success of an acquisition: Big Bank trains the employees of a location the week immediately preceding total conversion of the Customer Service system. Big Bank's goal is a seamless and flawless conversion with no noticeable impact on the customer.

If Big Bank followed Kirkpatrick's Four Level framework, it would engage in the following evaluation activities:

- Level 1: Assess the reaction of each employee who attended the training.
- Level 2: Test each employee on how much they learned in order to attribute learning to the training course.
- Level 3: Monitor employee use of the new skills for a specific period of time to determine how much of what employees learned transferred to the job.
- Level 4: Estimate the organizational impact of the new skills on overall business effectiveness; this would likely be a "negative" measure of assessing how much low skill transfer cost as compared to high and effective skill transfer.

In this situation, Kirkpatrick's Four Level framework would be ineffective and inefficient. Since Big Bank is not interested in modifying its current and proven systems training course, even if employees from a newly acquired firm don't like it, Level 1 evaluation is not necessary. Level 2 evaluation, testing the change in employee knowledge that is due to the training course, is also unnecessary since it is reasonable to assume that *all* knowledge about the new system will be due to the training course. Concerning transfer to the job, this is also a "given" since the old system will not be available once the new system is operational. All

employees will be required to apply their knowledge to the new system. Finally, Level 4 evaluation could be important, but only as an indicator of financial costs to convert.

If Big Bank followed the new proposed framework, it would begin by asking evaluation questions:

- What is the subject of the evaluation?
- What is the purpose of the evaluation?
- Who are the stakeholders in the evaluation?
- What are the agendas of these stakeholders?

The real subject of the evaluation is the effectiveness of the overall systems conversion. The training course is simply a vehicle to facilitate smooth conversion. The purpose of the evaluation is to identify potential problems with the conversion. This purpose could be considered a special version of summative evaluation: assessing the cumulative skills of specific workers as a tool to predict difficulties with "live" operation after the conversion. The stakeholders are the conversion team, the management team of the acquired location, and the IT/Systems support team that will provide troubleshooting for conversion difficulties. The agendas of all stakeholders are relatively consistent: smooth conversion. It is reasonable that some employees will balk at the swift conversion or will use the conversion as a vehicle to vent frustration or dissatisfaction with the acquisition and its resulting changes. While these are important local concerns, Big Bank considers this in its costs of acquisition and conversion but will not modify its conversion process to minimize acquired employee dissatisfaction.

These evaluation decisions suggest the following effectiveness decisions:

- The appropriate level of analysis would likely be the aggregate at the local unit being converted (meso), for example an individual location or a metro area that includes multiple locations.
- The intended outcomes of the evaluation would be to predict potential difficulties with the conversion so that remedial and/or corrective measures could be taken with minimal disruption to customer service.
- The specific units of analysis would be collective performance on specific skills or skill sets. Big Bank might also want to look at overall skills or skill sets for individual employees so that those with substandard skills can be given remedial training prior to working "live" with customers.
- Skill tests, including paper-and-pencil and computer simulations, are appropriate evaluation methods. Results from these tests would allow technical support specialists to prepare targeted approaches to minimize disruption when the system goes "live."
- As mentioned above, Big Bank might also want to monitor the costs of employee errors during conversion to ensure that the financial calculations remain accurate when considering future acquisitions.

In this example, the new approach to HRD evaluation would eliminate unnecessary, but standard, evaluation activities (Level 1, Level 3, and possibly Level 4). With a different investment in time and resources, the organization would be better equipped to use evaluation as a tool to mobilize the necessary resources to anticipate and meet the challenges at one or more locations on conversion day.

Summary

Evaluation is the feedback loop that allows the organization to monitor various factors of effectiveness – how well are we doing compared to what we intended to do? In the complexity of today's business, and the complexity and sophistication of HRD, our feedback processes must be equally sophisticated. We must understand that evaluation is a complex social process and that measurement of specific outcomes is only one of many factors that determine how an organization values HRD activities.

Although there may be many specific situations where the 40-year-old Four Level approach is still relevant and meaningful, it is generally ineffective as an overarching paradigm for today's business.

The most important revision to our understanding of evaluation practice is to understand that evaluation is a complex process of organizational decision-making. Measuring effectiveness (the traditional definition of evaluation) is a technical issue.