

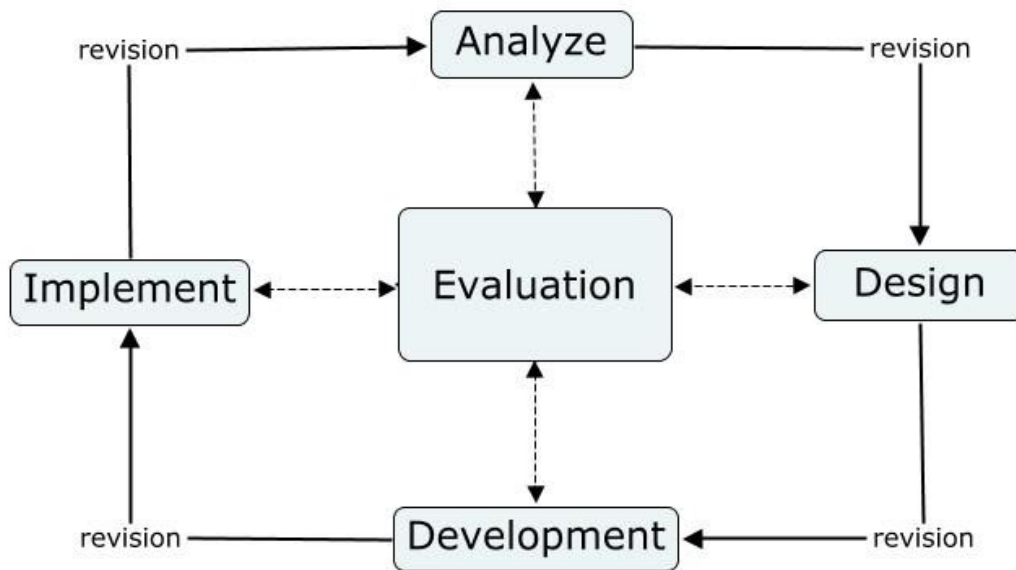
Instructional Design Evaluation

Cheryl Calhoun, Shilpa Sahay, & Matthew Wilson

Evaluation sits at the center of the instructional design model. It provides feedback to all other stages of the design process to continually inform and improve our instructional designs. In this chapter we will discuss the Why, What, When, and How of evaluation. We will explore several of the most cited evaluation models and frameworks for conducting formative, summative, and confirmative evaluations. It is important to note that instruction can occur in formal instructional settings or through the development of instructional products such as digital learning tools. Throughout this chapter we will discuss interchangeably instructional programs and/or products. Effective evaluation applies to all of these forms of instructional design.

Figure 1

ADDIE Model of Design (Fav203, 2012)



Why Do We Evaluate?

Evaluation ensures that the instruction being designed both meets the identified need for instruction and is effective in achieving the intended learning outcomes for participants. It helps to answer questions such as:

- Are our instructional goals aligned with the requirements of the instructional program?
- Are our lesson plans, instructional materials, media, and assessments, aligned with learning needs?
- Do we need to make any changes to our design to improve the effectiveness and overall satisfaction with the instruction?
- Does the implementation provide effective instruction and carry out the intended lesson plan and instructional objectives?
- Have the learners obtained the knowledge and skills that are needed?
- Are our learners able to transfer their learning into the desired contextual setting?

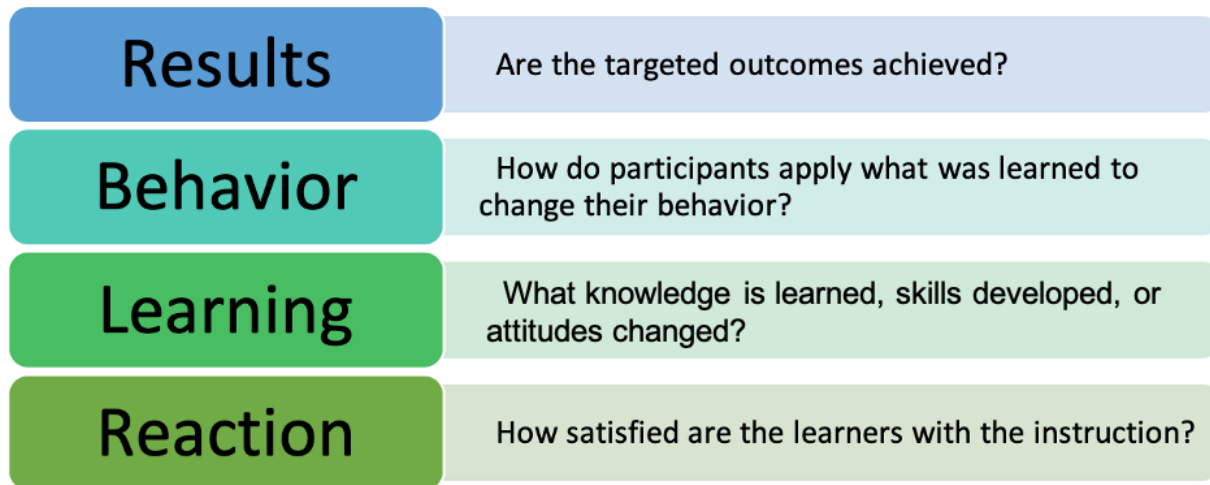
These questions help shape the instructional design, confirm what and to what extent the learner is learning, and validates the learning over time to support the choices made regarding the design—as well as how the program holds up over time.

What Is Evaluation?

Evaluation is the process of reviewing both the instructional components and the resulting outcomes of instruction to determine whether instruction achieves the desired outcomes. Kirkpatrick’s model of evaluation proposes four levels of evaluation: reaction, learning, behavior, and results (Kirkpatrick & Kirkpatrick, 2016). While this is a fairly simplistic model, it provides a framework for understanding evaluation and has provided a significant model of evaluation to the field of instructional design.

Figure 2

Kirkpatrick's Model of Evaluation



Reaction

In order to have effective instruction, one requires frequent feedback from the learners to check learning progress and monitor efficacy of the pedagogical process selected for instruction (Heritage, 2007). An instructional designer can evaluate both the teacher and the learner’s reaction to a new pedagogical instruction. Once it is determined that there is engagement by the learners, one may

assume that learners will not drop out due to their reaction to the quality or applicability of instruction. It also helps the evaluator to control the pace of the program as one moves ahead in the training phase. It leaves less frustration and vagueness in the evaluator's mind if one knows that all the learners are positively oriented towards undertaking the training.

Learning

Evaluating learning is an ongoing process in instructional development. It is important to evaluate whether materials developed solve the problems that were identified. When learners master the content of the training or exhibit proper learning through assessment, one can assume the effectiveness of the program and identify what did not work if the learning outcomes show adverse results. Several studies in the field of educational measurement have suggested that assessments and evaluations lead to higher quality learning. Popham (2008) called this new aspect of assessment in the evaluation process as "Transformative Assessment" where an evaluator identifies learning progression of the learners by analyzing the sequence of skills learned over the period of study program. This also helps the evaluator or the instructional designer to develop methods to assess how much the learners mastered the learning material.

Behavior

Attitudes and behavior are important indicators towards the acceptance and success of an instructional program. Dick, Carey, and Carey (2015) mentioned that an evaluator needs to write directions to guide the learner's activities and construct a rubric (e.g. a checklist or a rating scale) in order to evaluate and measure performance, products, and attitudes. A learner develops several intellectual and behavioral skills, and an evaluation can uncover what changes have been brought in the attitude and behavior of the learners.

Results

With every instructional product, evaluating results is the most significant task by an evaluator, and is done to determine how closely one has been able to achieve success in the implementation of the program. An evaluator conducts an evaluation in order to test the effectiveness of the instruction to create the desired learning outcome (Morrison et al., 2019). Morrison et al. (2019) suggested evaluators measure the efficiency of learning by comparing the skills mastered with the time taken; cost of program development; continuing expenses; reactions towards the program; and long-term benefits of the program.

When Do We Evaluate?

Three commonly used types of evaluation for instruction are formative, summative, and confirmative (Morrison et al., 2019; Ross & Morrison, 2010). Formative evaluation is conducted during the design process to provide feedback that informs the design process. Summative evaluation is conducted at the end of the design process to determine if the instructional product achieves the intended outcomes. Confirmative evaluation is conducted over time to determine the lasting effects of instruction. Each of these stages of evaluation is examined in detail here, both through the definition of the form itself and through a discussion of some of the key tools within each.

"When the cook tastes the soup that's formative; when the guests taste the soup, that's summative." -

Robert E. Stake (M. Scriven, 1991, p. 169)

Formative

Formative evaluation occurs during instructional design. It is the process of evaluating instruction and instructional materials to obtain feedback that in turn drives revisions to make instruction more efficient and effective. One way to think about this is to liken it to a chef tasting his food before he sends it out to the customer. Morrison et al. (2019) explained that the formative evaluation process utilizes data from media, instruction, and learner engagement to formulate a picture of learning from which the designer can make changes to the product before the final implementation.

Boston (2002, p. 2) stated the purpose of formative evaluation as “all activities that teachers and students undertake to get information that can be used diagnostically to alter teaching and learning.” Formative evaluation results in the improvement of instructional processes for the betterment of the learner. While making formative changes are best conducted during earlier stages of the design process, these changes may come later if the situation dictates it. According to Morrison et al., (2019), when summative and confirmative evaluations demonstrate undesirable effects, then the results may be used as a formative evaluation tool to make improvements.

Instructional designers should consider a variety of data sources to create a full picture of the effectiveness of their design. Morrison et al. (2019) proposed that connoisseur-based, decision-oriented, objective-based, and constructivist evaluations are each appropriate methodologies within the formative process. More recently Patton (2016) introduced developmental evaluation which introduces innovation and adaptation in dynamic environments.

Types of Formative Evaluation

Connoisseur-Based

Employs subject matter experts (SMEs) in the review of performance objectives, instruction, and assessments to verify learning, instructional analysis, context accuracy, material appropriateness, test item validity, and sequencing. Each of these items allow the designer to improve the organization and flow of instruction, accuracy of content, readability of materials, instructional practices, and total effectiveness (Morrison et al., 2019).

Decision-Oriented

Questions asked may develop out of the professional knowledge of an instructional designer or design team. These questions subsequently require the designer to develop further tools to assess the question, and as such should be completed at a time when change is still an option and financially prudent (Morrison et al., 2019).

Objective-Based

Through an examination of the goals of a course of instruction, the success of a learner’s performance may be analyzed.

Constructivist

Takes into account the skills students learned during the learning process as well as how they have

assimilated what is learned into their real lives.

Developmental

Responsive to context and more agile, allowing for quicker response and support of innovative designs (Patton, 2011).

Summative

Dick et al. (2015, p. 320) claimed the ultimate summative evaluation question is "Did it solve the problem?" That is the essence of summative evaluation. Continuing with the chef analogy from above, one asks, "Did the customer enjoy the food?" (M. Scriven, 1991). The parties involved in the evaluation take the data and draw a conclusion about the effectiveness of the designed instruction. However, over time, summative evaluation has developed into a process that is more complex than the initial question may let on. In modern instructional design, practitioners investigate multiple questions through assessment to determine learning effectiveness, learning efficiency, and cost effectiveness, as well as attitudes and reactions to learning (Morrison et al., 2019).

Learning Effectiveness

Learning effectiveness can be evaluated in many ways. Here we are trying to understand:

- How well did the student learn?
- Are the students motivated to change behavior?
- Did we engage the intended population of learners?
- Even, did we teach the learner the right thing?

Measurement of learning effectiveness can be ascertained from assessments, ratings of projects and performance, observations of learners' behavior, end of course surveys, focus groups, and interviews. Dick et al. (2015) outlined a comprehensive plan for summative evaluation throughout the design process, including collecting data from SMEs and during field trials for feedback.

Learning Efficiency and Cost-Effectiveness

While learning efficiency and cost-effectiveness of instruction are certainly distinct constructs, the successfulness of the former impacts the latter. Learning efficiency is a matter of resources (e.g., time, instructors, facilities, etc.), and how those resources are used within the instruction to reach the goal of successful instruction (Morrison et al., 2019). Dick et al. (2015) recommended comparing the materials against an organization's needs, target group, and resources. The result is the analysis of the data to make a final conclusion about the cost effectiveness based on any number of prescribed formulas.

Attitudes and Reactions to Learning

The attitudes and reactions to the learning, while integral to formative evaluation, can be summatively evaluated as well. Morrison et al. (2019) explained there are two uses for attitudinal evaluation: evaluating the instruction and evaluating outcomes within the learning. While most objectives within learning are cognitive, psychomotor and affective objectives may also be goals of learning. Summative evaluations often center on measuring achievement of objectives. As a result, there is a natural connection between attitudes and the assessment of affective objectives.

Conversely, designers may utilize summative assessments that collect data on the final versions of their learning product. This summative assessment measures the reactions to the learning.

Confirmative

The purpose of a confirmative evaluation is to determine if instruction is effective and if it met the organization's defined instructional needs. In effect, did it solve the problem? The customer ate the food and enjoyed it. But, did they come back? Confirmative evaluation goes beyond the scope of formative and summative evaluation and looks at whether the long-term effects of instruction is what the organization was hoping to achieve. Is instruction affecting behavior or providing learners with the skills needed as determined by the original goals of the instruction? Confirmative evaluation methods may not differ much from formative and summative outside of the fact that it occurs after implementation of a design. Moseley and Solomon (1997) described confirmative evaluation as maintaining focus on what is important to your stakeholders and ensuring the expectations for learning continue to be met.

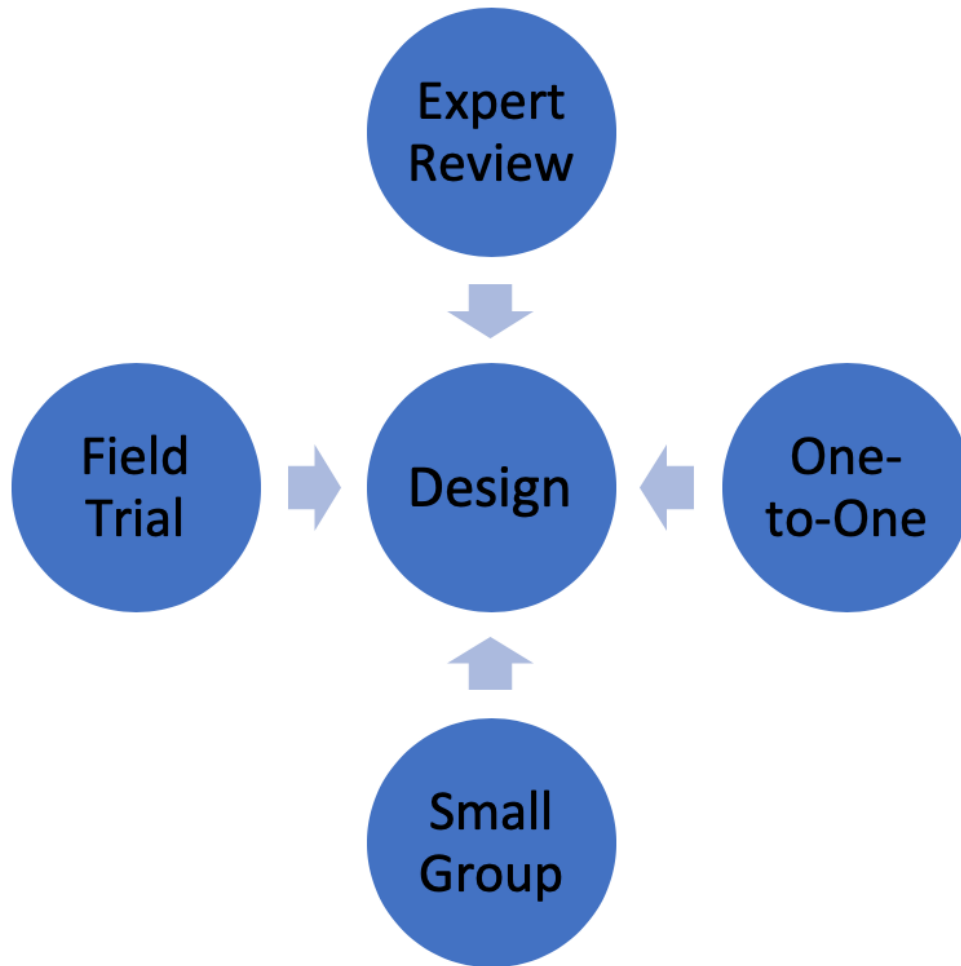
How Do We Evaluate?

Formative Evaluation

Formative evaluation is an iterative process that requires the involvement of instructional designers, subject matter experts, learners, and instructors. Tessmer (2013) identified four stages of formative evaluation including expert review, one-to-one, small group, and field test evaluation. Results from each phase of evaluation are fed back to the instructional designers to be used in the process of improving design. In all stages of evaluation, it is important that learners are selected that will closely match the characteristics of the target learner population.

Figure 3

The Cycle of Formative Evaluation

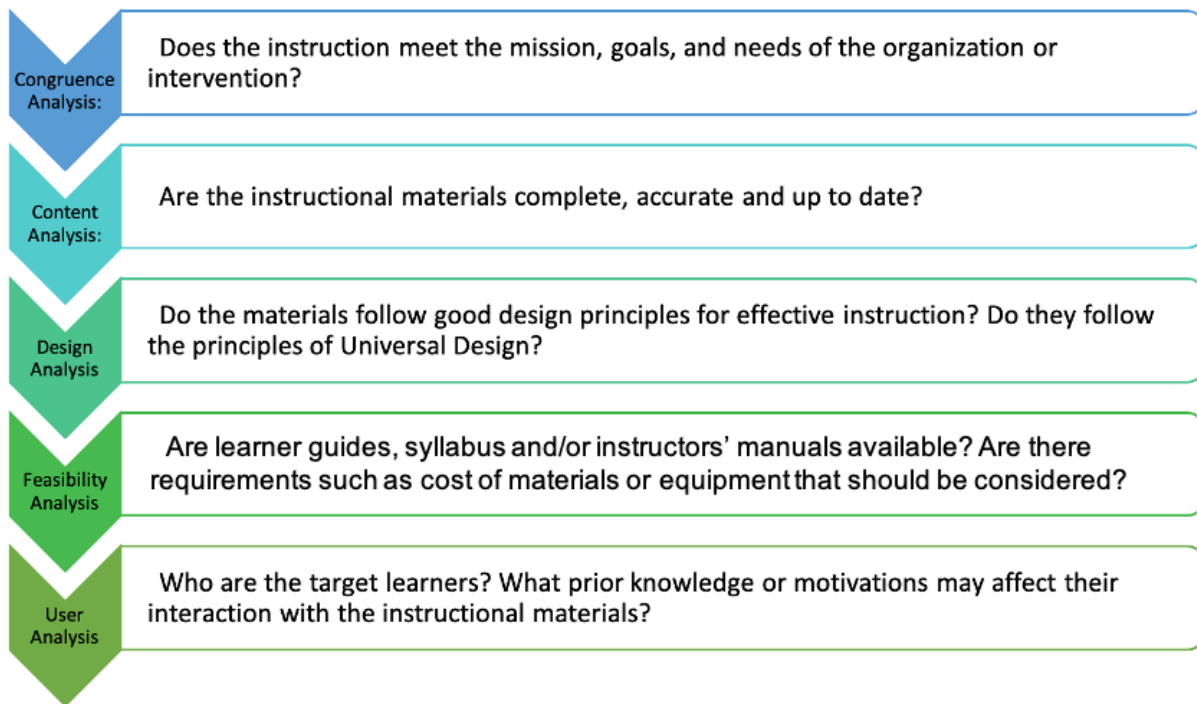


Expert Review

The purpose of the expert review is to identify and remove the most obvious errors and to obtain feedback on the effectiveness of the instruction. The expert judgment phase can include congruence analysis, content analysis, design analysis, feasibility analysis, and user analysis. Results from expert review can be used to improve instructional components and materials before a pilot implementation. This phase is conducted with the instructional designer, the subject matter experts, and often an external reviewer. Target learners are not involved in this stage of evaluation.

Figure 4

The Expert Judgment Phase (Dick et al., 2015)



One-to-One

The one-to-one evaluation is much like a usability study. During this evaluation, IDs should be looking for clarity, impact, and feasibility (Dick et al., 2015, p. 262; Earnshaw, Tawfik, & Schmidt, 2017). The learner is presented with the instructional materials that will be provided during the instruction. The evaluator should encourage the learner to discuss what they see, write on materials as appropriate, and note any errors. The ID can engage the learner in dialog to solicit feedback on the materials and clarity of instruction. There are many technological tools that can facilitate a one-on-one evaluation. The principles of Human Computer Interaction and User Center Design can inform the instructional design review (Earnshaw et al., 2017). In *Don't Make Me Think*, Krug (2014) described a process of performing a usability study for website development. The steps he provided are a good guide for performing a one-to-one evaluation. Krug recommended video recording the session for later analysis. If instruction is computer based, there are tools available that can record the learner interaction as well as the learner's responses. Morae from Techsmith (<https://edtechbooks.org/-oPnH>) is a tool that allows you to record user interactions and efficiently analyze the results.

Small Group

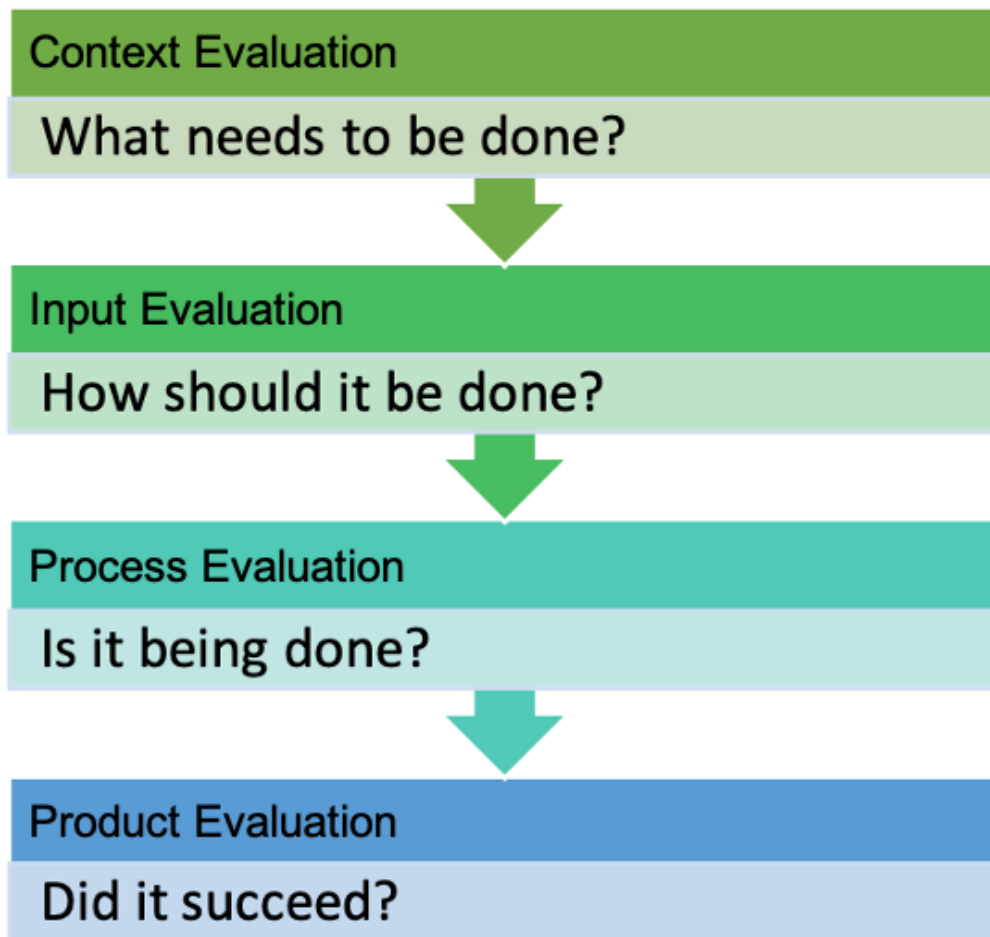
Small group evaluation is used to determine the effectiveness of changes made to the instruction following the one-to-one evaluation and to identify any additional problems learners may be experiencing. The focus is on consideration of whether learners can use the instruction without interaction from the instructor. In a small group evaluation, the instructor administers the instruction and materials in the way they are designed. The small-group participants complete the lesson(s) as described. The instructional designer observes but does not intervene. After the instructional lesson is complete, participants should be asked to complete a post-assessment designed to provide feedback about the instruction.

Field Trial

After the recommendations from the small group evaluation have been implemented, it is time for a field trial. The selected instruction should be delivered as close as possible to the way the design is meant to be implemented in the final instructional setting, and instruction should occur in a setting as close to the targeted setting as possible. Learners should be selected that closely match the characteristics of the intended learners. All instructional materials for the selected instructional section, including the instructor manual, should be complete and ready to use. Data should be gathered on learner performance and attitudes, time required to use the materials in the instructional context, and the effectiveness of the instructional management plan. During the field trial the ID does not participate in delivery of instruction. The ID and the review team will observe the process and record data about their observations.

Figure 5

Field Trial



Summative Evaluation

The purpose of a summative evaluation is to evaluate instruction and/or instructional materials after they are finalized. It is conducted during or immediately after implementation. This evaluation can be used to document the strengths and weaknesses in instruction or instructional materials, to decide whether to continue instruction, or whether to adopt instruction. External evaluators for decision makers often conduct or participate in summative evaluation. Subject matter experts may be needed to ensure integrity of the instruction and/or instructional materials. There are several models we can consider for summative evaluation including the CIPP Model, Stake's Model, and Scriven's Model.

CIPP Model

The CIPP evaluation model by Stufflebeam (1971) describes a framework for proactive evaluation to serve decision making and retroactive evaluation to serve accountability. The model defines evaluation as the process of delineating, obtaining, and providing useful information for judging decision alternatives. It includes four kinds of evaluation: context, input, process, and product. The first letters of the names of these four kinds of evaluation gave the acronym - CIPP. The model provides guidelines for how the steps in evaluation process interact with these different kinds of evaluation.

The CIPP Model of Evaluation by Mallory Buzun-Miller



Watch on YouTube <https://edtechbooks.org/rTn>

Stake's Model

Stake in 1969 created an evaluation framework to assist an evaluator in collecting, organizing, and interpreting data for the two major operations of evaluation (Stake, 1967; Wood, 2001). These include (a) complete description and (b) judgment of the program. W. J. Popham (1993) defined that Stake's

schemes draw attention towards the differences between the descriptive and judgmental acts according to their phase in an educational program, and these phases can be antecedent, transaction, and outcome. This is a comprehensive model for an evaluator to completely think through the procedures of an evaluation.

Dr. Robert Stake by Education at Illinois



Watch on YouTube <https://edtechbooks.org/-UpSY>

Scriven's Goal-Free Model

Scriven provides a transdisciplinary model of evaluation in which one draws from an objectivist view of evaluation (Michael Scriven, 1991a, 1991b). Scriven defined three characteristics to this model: epistemological, political, and disciplinary. Some of the important features of Scriven's goal free evaluation stress on validity, reliability, objectivity/credibility, importance/timeliness, relevance, scope, and efficiency in the whole process of teaching and learning. Youker (2013) expanded on the model to create general principles for guiding the goal-free evaluator. Younker proposed the following principles:

1. Identify relevant effects to examine without referencing goals and objectives.
2. Identify what occurred without the prompting of goals and objectives.
3. Determine if what occurred can logically be attributed to the program or intervention.
4. Determine the degree to which the effects are positive, negative, or neutral.

The main purpose of the goal-free evaluation is to determine what change has occurred that can be attributed to the instructional program. By conducting the evaluation without prior knowledge of learning outcomes or goals, the evaluator serves as a check to see if the program produced the outcome desired by the instructional designer(s).

The Past, Present, and Future of Evaluation: Possible Roles for the University of Melbourne, by The University of Melbourne



Watch on YouTube <https://edtechbooks.org/-yCH>

Confirmative Evaluation

The focus of confirmative evaluation should be on the transfer of knowledge or skill into a long-term context. To conduct a confirmative evaluation, you may want to use observations with verification by expert review. You may also develop or use checklists, interviews, observations, rating scales, assessments, and a review of organizational productivity data. Confirmative evaluation should be conducted on a regular basis. The interval of evaluation should be based on the needs of the organization and the instructional context.

Conclusion

Evaluation is the process of determining whether the designed instruction meets its intended goals. In addition, evaluation helps us to determine whether learners can transfer the skills and knowledge learned back into long-term changes in behavior and skills required for the target context. Evaluation provides the opportunity for instructional designers to ensure all stakeholders agree that the developed instruction is meeting the organizational goals.

In this chapter we reviewed what evaluation looks like and its relationship within the instructional design process. We looked at several models of evaluation including Kirkpatrick's Model and the four levels of evaluation: Evaluating Reaction, Evaluating Learning, Evaluating Behavior, and Evaluating Results. We also looked at the three phases of evaluation including formative, summative, and confirmative evaluation, and introduced several different models and methods for conducting evaluation from many leading evaluation scholars.

Discussion

- Where does evaluation stand in the instructional design model? How will your flow chart look when you describe evaluation in relation to the other stages of instructional design?
- Describe the three stages of evaluation. Give an example to explain how an instructional designer will use these three stages in a specific case of a learning product.
- Which are the five types of formative evaluation methods mentioned in the chapter that assist in collecting data points for the initial evaluation? Which two of these methods will be your preferred choice for your formative evaluation and why?
- What will be the parameters to evaluate the success of the instructional training?
- What are some of the techniques to conduct formative and summative evaluation?
- Several models of evaluation have been discussed in the chapter. Discuss any two of these models in detail and explain how you will apply these models in your evaluation process.

Application Exercises

For the following exercises, you may use an instructional module that you are familiar with from early childhood, k-12, higher ed, career and technical, corporate, or other implementation where instructional design is needed. Be creative and use something from an educational setting that you are interested in. Be sure to describe your selected instructional module as it relates to each of these exercises. You may need to do some additional online research to answer these questions. Be sure to include your references in your responses.

1. Describe how you would conduct the three phases of the formative evaluation. Define your strategies, populations, and methodologies for each stage within the process.
2. Draw a diagram of the iterative formative evaluation process. What specific pieces of the instructional intervention are considered within each stage of the process? How is the data gathered during this process employed to improve the design of instruction?
3. Describe the context and learner selection process you would use for setting up a formative evaluation field trial. What special considerations need to be made to conduct this stage of evaluation effectively?
4. What materials should the designer include in a field trial? How do the materials used for field trials contrast with the one-to-one and small group evaluations?

You have been asked to serve as an external evaluator on a summative evaluation of a training model designed by one of your colleagues. Explain the phases of the summative evaluation that you may be asked to participate in as an external reviewer. Imagine you have created a rubric to help you evaluate the instructional intervention. What items might that rubric contain to help you effectively and efficiently conduct a review?

Group Assignment

Conduct an evaluation study to understand how successful an instructional intervention has been in achieving the goals of the designed instruction. Keep in mind the group project conducted in the previous development and implementation chapters and conduct an evaluation study to assess the success of achieving the goals and objectives of the instruction. To achieve these goals, you should conduct several rounds of evaluation:

- Conduct a one-on-one evaluation with a student from the target population. Make observations of the student's actions within the instruction and reactions to the materials, content, and overall design. Propose changes to the instructional intervention based on the sample student's feedback.
- Conduct a small group evaluation with a group of 3 to 5 learners. This evaluation should reflect the changes you made after the one-to-one stage and evaluate nearly operational materials and instruction. You must recruit an instructor to deliver instruction. Make observations and have the instructor administer a summative assessment after instruction. Analyze the data gathered and create a one-page report on the results.
- Implement a field test with an instructor and a student population of at least 15 people. Instructional materials, including the instructor manual, should be complete and ready to use. Gather data on learner performance and attitudes, as well as time required for instruction and the effectiveness of the instructional management plan. Observe the process and record data, and create a final report detailing the full evaluation cycle.

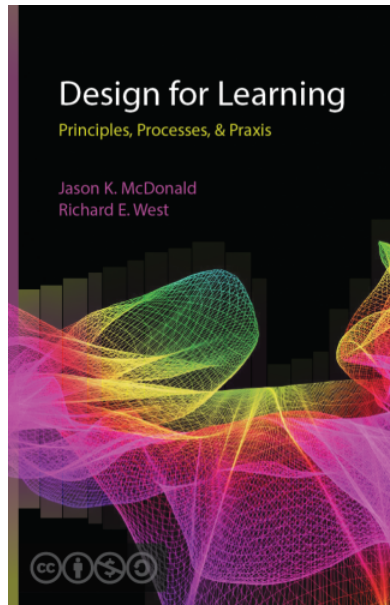
Editor's Note

This is a remixed version of an earlier chapter on [evaluation in instructional design](#) that can be found at the [ADDIE Explained](#) website.

References

- Boston, C. (2002). The concept of formative assessment. In: ERIC Clearinghouse on Assessment and Evaluation.
- Dick, W., Carey, L., & Carey, J. (2015). The systematic design of instruction (8th ed.). USA: Pearson.
- Earnshaw, Y., Tawfik, A. A., & Schmidt, M. (2017). User experience design. In Foundations of Learning and Instructional Design Technology.
- Fav203. (2012). ADDIE_Model_of_Design.jpg. In. Wikipedia.com: is licensed under CC BY-SA 3.0 via Wikimedia Commons. Retrieved from <https://edtechbooks.org/-kyA>
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? Phi Delta Kappan, 89(2), 140-145.
- Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). Kirkpatrick's four levels of training evaluation: Association for Talent Development.
- Krug, S. (2014). Don't Make Me Think, Revisited (Vol. 3): New Riders.

- Morrison, G. R., Ross, S. J., Morrison, J. R., & Kalman, H. K. (2019). *Designing effective instruction*: Wiley.
- Moseley, J. L., & Solomon, D. L. (1997). Confirmative Evaluation: A New Paradigm for Continuous Improvement. *Performance Improvement*, 36(5), 12-16.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*: Guilford Press.
- Patton, M. Q. (2016). The state of the art and practice of developmental evaluation. *Developmental evaluation exemplars*, 1-24.
- Popham, W. (2008). *Transformative Assessment*: Association for Supervision and Curriculum Development. 1703 North Beauregard Street, Alexandria, VA 22311-1714. In: Tel.
- Popham, W. J. (1993). *Educational Evaluation*: Allyn and Bacon. Retrieved from <https://edtechbooks.org/-mnxy>
- Ross, S. M., & Morrison, G. R. (2010). The Role of Evaluation in Instructional Design. In *Handbook of Improving Performance in the Workplace: Instructional Design and Training Delivery*.
- Scriven, M. (1991a). Beyond formative and summative evaluation. *Evaluation and education: At quarter century*, 10(Part II), 19-64.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed. ed.). Newbury Park, CA: Sage Publications.
- Scriven, M. (1991b). Prose and cons about goal-free evaluation. *Evaluation Practice*, 12(1), 55-62.
- Stake, R. E. (1967). *The countenance of educational evaluation*: Citeseer.
- Stufflebeam, D. L. (1971). The relevance of the CIPP evaluation model for educational accountability. Retrieved from <https://edtechbooks.org/-NIX>
- Tessmer, M. (2013). *Planning and conducting formative evaluations*: Routledge.
- Wood, B. B. (2001). Stake's countenance model: evaluating an environmental education professional development course. *The Journal of Environmental Education*, 32(2), 18-27.
- Youker, B. W. (2013). Goal-free evaluation: A potential model for the evaluation of social work programs. *Social Work Research*, 37(4), 432-438.



Calhoun, C., Sahay, S., & Wilson, M. (2020). Instructional Design Evaluation. In J. K. McDonald & R. E. West (Eds.), *Design for Learning*. EdTech Books. Retrieved from https://edtechbooks.org/id/instructional_design_evaluation



CC BY-NC-SA: This work is released under a CC BY-NC-SA license, which means that you are free to do with it as you please as long as you (1) properly attribute it, (2) do not use it for commercial gain, and (3) share any subsequent works under the same or a similar license.