



DISTANCE EDUCATION FOR TEACHER TRAINING:

Modes, Models, and Methods

Mary Burns

Education Development Center, Inc.

Washington, DC

Acknowledgments

Thank you to the following colleagues both near and far for providing information, resources, and fact-checking on this guide.

Sarwat Alam, Director of Learning Systems and Pedagogy, USAID Pre-STEP Project (2013), Pakistan

Dr. Catherine Margaret Beukes-Amiss, Director, Centre for Innovation in Learning and Teaching (CILT), University of Namibia, Windhoek, Namibia

Arjana Blazic, Teacher Trainer, and Course Designer, EduDigiCon, Zagreb, Croatia

Alisa Buchstab, Junior Policy Advisor in the Sector Program Education, Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH (GIZ), Germany

Dr. Eduardo C. Cascallar, University of Leuven, Belgium; Managing Director, Assessment Group International (Europe and USA)

Will Clurman, CEO, and co-founder, eKitabu, Nairobi, Kenya

Valeria Cruz Gomes, Head of Training and Support, ProFuturo, Madrid, Spain

Dr. Robyn A. Defelice, Learning Strategist and Consultant, Bloomsburg, Pennsylvania, USA

Dr. Nathalia Edisherashvili, Researcher, Institute of Education, University of Tartu, Estonia

Concepción Gallego Garcia, Expert on Global Partnerships and Institutional Relations, ProFuturo, Madrid, Spain

Dr. Sophia Gorgodze, Director National Assessment and Examinations Center, Ministry of Education and Science of Georgia

Dr. Sara Hennessy, Professor of Teacher Development and Pedagogical Innovation, Faculty of Education, University of Cambridge, and Research Director, EdTech Hub, Cambridge, England

Shane Ives, Serious gamer, solar electrician, Albuquerque, New Mexico, USA

Eilean von Lautz-Cauzanet, Policy Advisor in the Sector Program, Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH (GIZ), Germany

Stephen McDonald, app developer, Somerville, Massachusetts, USA

Dr. Mary Mendenhall, Teachers College, Columbia University, New York City, New York, USA

Dr. Barbara Moser-Mercer, Visiting Professor, University of Nairobi, Coordinator African Higher Education in Emergencies Network (AHEEN), Nairobi, Kenya

Denise-Henry Orndorff, Instructional Technology Coach, Frederick County Public Schools, Virginia, USA

Ee-Reh Owo, Schools Director, Justice Rising, Democratic Republic of the Congo

Aroa Sánchez Rodríguez, Educational Specialist, Innovation and Products, ProFuturo, Madrid, Spain

Jodi Sansone, Instructional designer, and eLearning specialist, Jodisansone.com, USA

Dr. Beverly Shirley, Academic Program Officer for University of the West Indies Open Campus British Overseas Territories, (Cayman Islands, Anguilla, British Virgin Islands, Monserrat, Turks & Caicos, Bermuda), George Town, Grand Cayman, West Indies

Dr. Carmen Strigel, Director, Education Technology, RTI International, North Carolina, USA

Dr. Torrey Trust, Associate Professor, Learning Technology, College of Education University of Massachusetts Amherst, USA

Freda Wolfenden, Professor of Education and International Development, School of Education, Childhood, Youth and Sports Studies, The Faculty of Wellbeing, Education and Language Studies, United Kingdom Open University, Milton Keynes, England

Dr. Diana Woolis, CEO, Sustainable Learning Strategies, New York City, New York, USA

Nicole M. Zumpano, Director of Instructional Technology Coaching, The Learning Technology Center (LTC) of Illinois, USA

My thanks to Education Development Center (EDC) colleagues for sharing program information and for providing resources, review, and feedback on chapters or content.

Helen Boyle, Vice President, Director of Program Strategy

Susan Bruckner, Senior International Technical Advisor

Nancy Meaker Chervin, International Technical Advisor

Rachel Christina, Director, International Basic Education

Leslie Goodyear, Distinguished Scholar/
Principal Evaluation Director

Nevin Katz, Web and App Developer

Stephanie Knutson, International
Accreditors for Continuing Education
and Training Compliance Manager

Nora Nunn, International Technical Associate

Shelley Pasnik, Senior Vice President

Gerald Sanders, Facilities Administrative Manager

Tamara Vitolo, Research Associate, Center
for Children and Technology

Katherine Yasin, Principal International Technical
Advisor, Director of English for Latin America.

Special thanks to Bronwyn Taggart
for her careful editing.

My deepest gratitude to Mary Hooker, International
Technical Advisor, Education Development Center,
for review, feedback, and wonderful insights on
multiple chapters in Section II of this guide.

My final thanks to those who fund, develop, research, evaluate, design, teach and participate in distance learning programs for teachers across the globe. This guide draws on your work. Thank you as well to EDC's Digital Design Group.

About the Author

Mary Burns is a senior technology and teacher professional development specialist at EDC. A former 10-year teacher in the United States, México, and Jamaica, she has worked in the area of technology-enabled professional development since 1997, instructing, designing, and evaluating both distance-based and face-to-face professional development for teachers, teacher educators, and instructional coaches. She has authored peer-reviewed papers, books, articles, and blog posts about teacher professional development, distance learning, and teaching with technology. She works in Asia, Africa, the Middle East, Latin America, the Caribbean, Europe, and the United States.

Preferred Citation

Burns, M. (2023). *Distance Education for Teacher Training: Modes, Models and Methods*. (2nd Edition). Washington, DC: Education Development Center.



Section II. Chapter 18

EVALUATING DISTANCE EDUCATION PROGRAMS

Table of Contents

18.1 Overview.....	1
18.2 What is Evaluation?.....	2
18.2.1 Evaluation versus Assessment.....	2
18.2.2 Evaluation versus Research.....	2
18.2.3 Evaluation versus Quality Assurance	2
18.3 Why Evaluate?	4
18.4 Types of Evaluation	4
18.5 Evaluating Distance Education Programs: A Non-systematic Approach	4
18.5.1 Course Evaluations.....	5
18.5.2 Adapting In-Person Professional Development Evaluation Tools	6
18.6 Evaluating Distance Programs: A Systematic Approach.....	9
18.6.1 Initial Decision Points	9
18.7 Technology and Evaluation	14
18.8 Final Evaluation Considerations	15
18.8.1 Ensure Stakeholder Agreement	15
18.8.2 Evaluate Before Scaling	16
18.8.3 Follow Government and Funder Guidelines	16
18.8.4 Align Expectations, Outcomes, Activities, and Evaluation.....	17
18.8.5 Exercise Caution in Consuming Evaluations	17
18.9 Conclusion: “The Greatest of Mistakes Is to do Nothing Because You Can Only Do a Little”	17

Best Practice: Successful distance education programs are characterized by continual formative evaluation and rigorous summative evaluation.

18.1 Overview

As Chapter 7 of this guide has discussed, and Section 1 of the guide has shown, distance programs have increasingly embraced the importance of evaluation (Morris et al., 2021). Well-designed and implemented evaluations inform distance education policymakers, planners, funders, and implementers about the value and merit of distance programs.

Evaluating distance education programs is not without its challenges and evaluation has traditionally been one of the most overlooked areas in distance education. Outcomes may not be defined, the purpose may not be determined, and questions about who benefits (teacher-learner, school, or student) may not be developed. The program may have been designed with unclear goals or objectives against which it cannot be measured, or the evaluation may have been designed after the program began. The capacity and resources to conduct an evaluation may be limited or nonexistent—and worse, high attrition rates may render any evaluation unreliable¹ or invalid.² Combine these issues within the nontraditional setting of distance education, and the design and implementation of rigorous and meaningful evaluations are often severely handicapped.

Evaluations of any distance education program also confront a number of methodological problems, including the need for measures other than

standardized achievement tests; disparities among learners in opportunities to learn; and differences in starting points and program implementation. Many distance education programs that receive funding via external aid agencies may have to concern themselves only with monitoring and evaluation, which traditionally looks at inputs (number of teachers trained) versus outcomes (number of teachers who implement a strategy) or impact (how learner achievement has changed as a result of teachers' professional development). Left unanswered is the most salient questions about the worth and value of a distance course of study.

Yet, formidable although they may be, the drawbacks of evaluations are far outweighed by their benefits. To accrue their full worth, evaluations should not be simply a valedictory exercise—initiated just as a distance program is about to conclude. Rather they should be woven through program design and implementation and undertaken by skilled evaluators working closely and collaboratively with a distance education program over the course of the program's existence. This way, every distance program, regardless of size, characteristics, or purpose, will have "credible, useful evidence-based information" so problems can be fixed and the value and worth of the program determined (United Nations Evaluation Group, as cited in World Food Program Office of Evaluation, 2021, p.2).

¹An evaluation instrument is considered reliable if the instrument can be used repeatedly with different groups of similar subjects and yield consistent results.

²Validity refers to the accuracy of an assessment—whether or not it measures what it is supposed to measure.

Evaluation is a complex and comprehensive topic that cannot be adequately addressed in a single chapter. Thus, this chapter provides a general overview of evaluations within a distance education context. It outlines decision points, steps, and guidelines that distance programs may employ in consultation with evaluators to determine the worth and value of their distance education offerings.

18.2 What is Evaluation?

The United Nations Evaluation Group (2016) defines evaluation as a process which is:

conducted as systematically and impartially as possible... (analyzing) both expected and unexpected results by examining the results chain, processes, contextual factors, and causality using appropriate criteria (and) ... (It provides) credible, useful evidence-based information that enables the timely incorporation of its findings, recommendations, and lessons into the decision-making process of organizations and stakeholders. (Cited by World Food Program Office of Evaluation, 2021, p. 2)

Evaluation is typically done at the request of a client—funders, distance education programs, or policymakers—and in “collaboration with various stakeholders” who may be invested in distinguishing between what works and what doesn’t; assuring accountability; or improving a particular intervention, event, program, or activity (Bonney et al., 2011, p. 14; Rossi et al., 2004, p. 14).

As important as defining what evaluation *is*, is defining what it *is not*. Evaluation, although overlapping and sometimes conflated with *assessment*, *research*, and *quality assurance*, is distinct from all three of these activities.

18.2.1 Evaluation versus Assessment

Evaluation shares similarities with assessment, discussed in Chapter 17. Like assessment, it has multiple models, and as with assessment,

evaluations can be front-end (sometimes diagnostic), formative, and summative (Rossi et al., 2004). However, while evaluations ascertain whether a set of standards have been met and render value judgments about the work or value of a *program*, assessment is a process for gathering information that is used to make decisions about *people*—it provides feedback on learner performance and ways to enhance that performance in the future (Brookhart & Nitko, 2011). Thus, within this guide, assessment focuses on *people* (specifically teacher-learners) while evaluation focuses on *programs*.

18.2.2 Evaluation versus Research

Evaluation is also similar to, but distinct from, *research*. Both evaluations and research have different foci and purposes, and there are lively debates as to how to define each. For the purposes of this guide, evaluations are focused on judging and improving the merit, worth, value, or effectiveness of a particular program, while educational research studies a particular phenomenon, often within an academic discipline or a specific theoretical framework. Unlike evaluations, which are client-focused, undertaken for a specific purpose and often are not published, research has no particular client and is undertaken for purposes of knowledge generation, with the ultimate goal of publication in a peer-reviewed or other respected education journal (Bonney et al., 2011, p. 15).

18.2.3 Evaluation versus Quality Assurance

Finally, evaluation is a major part of *quality assurance*, and both program evaluation and quality assurance are part of continuous improvement cycles. Yet, again, these two terms are distinct. If evaluation is focused on programs and assessment on people, quality assurance is focused on processes and outcomes. And while evaluation is about discernment and judgment, quality assurance is about elimination of defects and alignment to standard). Evaluation is one—extremely critical—component of a quality assurance system designed to identify, analyze, and eliminate defects in

Figure 18.1
The Three Types of Evaluation (Bonney et al., 2011, pp. 16, 26, 49)

Type	When It Generally Occurs	Focus	Purpose
Front-End	Before a distance education course or program begins	Design: Plan and shape the content and instruction that distance learners receive	<ul style="list-style-type: none"> • Audience research: For example, current teacher practice and teacher needs • Market research: For example, current distance learning opportunities • Contextual information: National educational priorities and socioeconomic conditions of teacher-learners
Formative	Throughout the life of the distance education course or program	Improvement: Identify areas of improvement	<ul style="list-style-type: none"> • Gather data about a project's strengths and weaknesses for purposes of revision and improvement • Monitor a project on an ongoing basis through regular data collection • Describe how a project functions • Provide recommendations to improve project functionality • Clarify program purpose or theory
Summative	At the end of a distance education course or at certain intervals in a distance program (e.g., end of the year or semester)	Judgment: Make decisions about continuing, replicating, or terminating a program	<ul style="list-style-type: none"> • Determine a program's overall effectiveness and value • Gauge whether targeted outcomes have been achieved • Summarize learning from the evaluation and any unintended effects that were documented • Identify project strengths and weaknesses • Determine overall value or worth of a project • Determine cause-and-effect relationships between an intervention and outcomes

processes and outcomes (Donabedian, 1988, as cited in Leahy et al., 2009, p. 70).

18.3 Why Evaluate?

Continual monitoring and rigorous, well-designed and implemented evaluations are critical to the success of any distance education program. They inform distance education policymakers, planners, funders, and instructors about the value and merit of distance programs and indicate what assumptions, inputs, and activities should change and how change should be accomplished. As discussed in the previous chapter, evaluation has teacher licensing implications; as discussed in the next chapter, it also has accreditation implications.

Thus, evaluation processes and results can improve programs and determine which ones should be maintained, changed, or closed. They provide insights into a particular program, generate knowledge, and help educators generalize or predict future behaviors or outcomes in similar situations in order to scale up innovations (Patton, 2008).

Simply put, without well-designed and rigorous evaluations, we cannot make claims about the effectiveness or ineffectiveness of an intervention. Without evaluation, we have no idea whether a distance education program really works. And if a program does fail, a good evaluation can help planners and designers understand and learn from the failure.

18.4 Types of Evaluation

Evaluations generally fall into one of three types: front-end (sometimes referred to as diagnostic), formative, or summative.

A well-known simile for understanding the above types of evaluation is to compare evaluation to soup. When the cook asks the customer what they want to eat (something hot and filling—soup!), this is *front-end* evaluation. As the cook tastes the soup in its preparation, this is *formative* evaluation. When the customer makes

a pronouncement on the soup—delicious or needing more salt—this is *summative* evaluation (Scriven, 1991, p. 63).

All three of these types of evaluation should be part of any distance education program. Moon et al. (2005) suggest early and ongoing formative evaluation during course development and during the pilot phase to ensure that courses are effective and achieving their stated objectives. This process might include the following:

- A review of course prototypes by content experts, distance education experts, and instructional design experts
- A pilot study tracking learner usage, along with instructor and learner surveys and focus groups
- Interviews and focus groups with learners (pre-service teacher candidates and in-service teachers) on questions of pace, workload, responsiveness of instructor/facilitator/tutor, levels and types of support, student learning, learner satisfaction, and ease of technology use
- A final pilot evaluation report (Moon et al., 2005)

All of this information should then be used to inform future planning, make midcourse corrections and revisions, and curtail any projects that are not succeeding before more time, energy, and funding are devoted to them (Gaible & Burns, 2007). The distinct types of evaluations—front-end (diagnostic), formative, and summative—can accommodate distinctive designs depending on the questions being asked and the different measures and outcomes used to provide information to answer these questions (see Figure 18.5).

18.5 Evaluating Distance Education Programs: A Non-systematic Approach

Many distance education programs may not be ready to embark on a systematic or full-scale evaluation, but they still can gather information from teacher-learners. They can use the course survey features embedded in a learning management system (LMS), design their own

surveys or interview questions for teachers, or adapt previously developed teacher professional development evaluation tools for an online or distance-based environment. This section explores these three potential options for collecting basic data from teachers for formative purposes.

18.5.1 Course Evaluations

Course evaluations can be designed within the online course platform itself as part of the LMS or as a webinar or online seminar, or they can be sent via a mobile device. Course evaluations provide useful summative information as well as formative information on how to improve the course offering. Figure 18.2 provides an overview of a potential

Figure 18.2
Questionnaire/Survey

Purpose
Assess teachers' perceptions of the distance program—its benefits, weaknesses, changes on their practice, and suggested improvements
Appropriateness
<ul style="list-style-type: none"> • This is a predetermined list of questions that can consist of structured or unstructured responses. • The format can be print or digital. • It can be mailed or dropped off to teachers and collected for completion.
Strengths
<ul style="list-style-type: none"> • A large sample size can be accommodated. • It works well if the sample is geographically dispersed. • The sample is useful if it is certain to be completed and returned. • It allows for easy data analysis.
Weaknesses
<ul style="list-style-type: none"> • It is more difficult to differentiate among levels of response (e.g., on scale from 1–5, is there an incremental and discernible difference between 3 and 4?). • Surveys generally have a low return rate. • If done on paper, there can be lengthy delays in completion and the return of results. • The information yielded is more superficial and narrower. • Surveys suffer from “desirability bias.” Respondents often select the response they believe the surveyor wants to see.
Cost Considerations
<ul style="list-style-type: none"> • Comparatively low: It can be conducted online via mobile phones (e.g., text) or interactive voice response (IVR). • It may involve printing, mailing (or personal delivery), and collection. • It will involve data entry, cleaning (spreadsheet), and data analysis.

course survey. Note that the “Purpose” column provides an example, as opposed to a recommendation.³ The advantage of designing a survey in an LMS⁴ is that questions can be stored in a Question Bank or Item Bank and then reused and repurposed as needed for future courses.

18.5.2 Adapting In-Person Professional Development Evaluation Tools

Online and blended distance programs, interactive audio instruction (IAI), print-based, video-based, and mobile learning programs may, and often do, adapt, or use in their entirety existing evaluation frameworks created for in-person environments. The following three such frameworks are discussed: Kirkpatrick, Guskey, and Scriven.

Kirkpatrick’s Four Levels of Training Evaluation

Internationally, one of the best-known frameworks for evaluating professional development has been Kirkpatrick’s model, developed in 1959 to evaluate trainings for Heifer International. This model was updated continuously until 1993. It comprises four levels, each of which builds stepwise on the previous level:

- Level I evaluates teachers’ *reactions* to the professional development.
- Level II evaluates teachers’ *learning*.
- Level III evaluates teachers’ *behavior*.
- Level IV evaluates professional development *results* in the classroom (Mindtools, n.d.).

Not surprisingly for a framework that is so long-lived, the four-level model is not without criticism. There are suggestions that it be implemented in reverse order, and there are questions about its utility in an age of so much informal learning (Mindtools, n.d.). Despite the critiques, the framework is still popular.

Guskey’s Five Levels of Evaluating Professional Development

A similar, but more comprehensive, professional development evaluation framework is that of Thomas Guskey (2000, 2016), whose five-level framework for evaluating professional development is outlined in Figure 18.3 (next page). These levels range from the lowest level of evaluation—assessing teachers’ reactions to the professional development—to the highest—determining whether the professional development for teachers had any impact on *student* learning.

These five levels reflect the complexity of evaluating professional development, but they also serve as a good model for evaluating professional development—whether in-person, blended or via distance. As Figure 18.3 implies, multiple types of evaluations can be created to measure different outcomes while many levels of the evaluation may also use many of the same instruments (e.g., interviews and teacher portfolios).

Scriven’s Evaluation of Training

A third model is Scriven’s Evaluation of Training (2009), a training or professional development evaluation checklist that can be used for formative and summative evaluations, monitoring professional development, and even conducting meta-evaluations. As will be seen, it combines elements of Kirkpatrick’s four levels and Guskey’s five levels of evaluating professional development. The checklist consists of 11 questions, listed in Figure 18.4.

Programs have at least two options for using these data. They can use the information they’ve gathered to inform future iterations of an online course, or they can then hire an external evaluator to explore larger questions and issues that emerge as part of a more systematic evaluation (L. Goodyear, personal communication, September 16, 2022). This more systematic approach is discussed after Figure 18.3.

³The University of Wisconsin offers comprehensive information on designing end-of-course surveys, including questions, considerations, and uses of the survey. See <https://assessment.provost.wisc.edu/best-practices-and-sample-questions-for-course-evaluation-surveys/>

⁴Another resource for survey design can be found at Tools4Dev at <https://tinyurl.com/4xad58f4>

Figure 18.3
Five Levels of Evaluating Professional Development (Guskey, 2000, 2016)

Evaluation Level	What Questions Are Addressed?	How Will Information Be Gathered?	What Is Measured or Assessed?	How Can This Information Be Used?
Level 1: Teachers' reactions	<ul style="list-style-type: none"> • Did teachers like it? • Was their time well spent? • Did the materials make sense? • Was the instructor knowledgeable and helpful? 	<ul style="list-style-type: none"> • Questionnaires administered at the end of the session 	<ul style="list-style-type: none"> • Initial satisfaction with the experience 	<ul style="list-style-type: none"> • To improve program design and delivery
Level 2: Teachers' learning	<ul style="list-style-type: none"> • Did teacher-learners acquire the intended knowledge and skills? 	<ul style="list-style-type: none"> • Paper-based/digital instruments • Simulations • Demonstrations • Participant reflection • Participant portfolios 	<ul style="list-style-type: none"> • New knowledge and skills of teacher-learners 	<ul style="list-style-type: none"> • To improve program content, format, and organization
Level 3: Organization support and change	<ul style="list-style-type: none"> • What was the impact on the organization? • Did it affect organizational climate and procedures? • Was implementation advocated, facilitated, and supported? • Were problems addressed quickly and efficiently? 	<ul style="list-style-type: none"> • District and school records • Minutes from follow-up meetings • Questionnaires • Structured interviews with participants or administrators • Participant portfolios 	<ul style="list-style-type: none"> • Organization's advocacy, support, accommodation, facilitation, and recognition 	<ul style="list-style-type: none"> • To document and improve organizational support • To inform future change efforts
Level 4: Teacher-learners' use of new knowledge and skills	<ul style="list-style-type: none"> • Did teachers effectively apply new knowledge and skills? 	<ul style="list-style-type: none"> • Questionnaires • Structured interviews with teachers and administrators • Teacher portfolios • Teacher reflections • Direct or videotaped classroom observations 	<ul style="list-style-type: none"> • Degree and quality of implementation 	<ul style="list-style-type: none"> • To document and improve implementation of program content

Evaluation Level	What Questions Are Addressed?	How Will Information Be Gathered?	What Is Measured or Assessed?	How Can This Information Be Used?
Level 5: Student learning outcomes	<ul style="list-style-type: none"> •What was the impact on students? •Did it affect student performance/ achievement? •Did it influence students' emotional/ physical well-being? •Are students more confident as learners? •Is attendance increasing? 	<ul style="list-style-type: none"> •Student grades •School records •Questionnaires •Structured interviews with students, teachers, administrators, and parents •Teacher portfolios 	<ul style="list-style-type: none"> •Student learning outcomes: cognitive, affective, conative, and psychomotor 	<ul style="list-style-type: none"> •To focus and improve all aspects of program design, implementation, and follow-up •To demonstrate overall impact of professional development

Figure 18.4
Evaluation of Training Checklist (Scriven, 2009)

No.	Topic	Question
1.	Need	<ul style="list-style-type: none"> • Is this professional development the best way to address this particular need?
2.	Design	<ul style="list-style-type: none"> • Does the design of the professional development target the particular need defined above? • Does it target teachers' background and current knowledge, skills, attitudes, and values? • Does it consider existing resources?
3.	Delivery	<ul style="list-style-type: none"> • Was the professional development announced, attended, supported, and presented as proposed?
4.	Reaction	<ul style="list-style-type: none"> • Was the professional development relevant, comprehensible, and comprehensive?
5.	Learning	<ul style="list-style-type: none"> • Did teachers master intended content, acquire intended value, or modify their attitudes as a result of the professional development?
6.	Retention	<ul style="list-style-type: none"> • Did teachers retain the learning for appropriate intervals?
7.	Application	<ul style="list-style-type: none"> • Did teachers use and appropriately apply what they learned in the professional development?
8.	Extension	<ul style="list-style-type: none"> • Did teachers use what they learned at other times, in other sites, or with other subjects?

No.	Topic	Question
9.	Value	• What was the value of the professional development for teachers?
10.	Alternatives	• What alternative approaches could be used to meet the same needs?
11.	Return on Investment	• What is the value of the professional development for students, the school, the district, the region, and the educational environment?

18.6 Evaluating Distance Programs: A Systematic Approach

For many distance education programs, the information gathered from either a design-it-yourself survey or the three teacher professional development evaluation tools discussed in the above section may suffice, particularly for a formative evaluation where the purpose is to get an idea of teachers' reactions to a distance course or get a general idea of teacher enactment of a strategy, insight, or information.

But for distance programs that want deeper and broader evidence for purposes of program improvement, scaling their intervention, or determining impact, then a more systematic approach, undertaken by a professional evaluator or evaluation team, is necessary. The remainder of this chapter lays out steps and guidelines for this approach. While distance educators need not immerse themselves in the weeds of methodologies, measures, and indicators, it is helpful to understand the broad contours of the evaluation process since they may work closely and consult with evaluators.

This last suggestion emerges from two motivations. First, while evaluation and research are distinct, many evaluators publish evaluations as research. At that point, it is often *they*, not the distance program designers, instructors, and learners, who become associated with the program's success.⁵ Second, including the perspectives, expertise, and experiences of distance education practitioners and teacher-

learners suffuses the findings with a depth and texture that may be absent from a purely external process, potentially resulting in more compelling, coherent explanatory narratives (Burns, 2020).

18.6.1 Initial Decision Points

Distance education programs embarking on a more systematic evaluation will confront four immediate "decision points" (Bonney et al., 2011, p. 25):

1. What type of evaluation?
2. An internal or external evaluator?
3. An independent contractor or an evaluation firm?
4. A local or out-of-area evaluator?

The type of evaluation required will be determined by the questions the program wants answered and the evaluation design it uses to answer those questions. For decision points 2–4, decisions will be determined by considering tradeoffs: impartiality versus in-depth knowledge of a program versus perceived bias; project requirements versus budget constraints; knowledge of the local context and cultural competence versus knowledge of the global evaluation field and professional credibility (Bonney et al., p. 26). For some distance programs, the answers to decision points 2–4 may be one, another, or both (i.e., an internal *and* an external evaluator).

Once these decisions are made, evaluators, in consultation with distance education programs, can embark on the steps outlined in Figure 18.5.

⁵ Given publication bias, reflected in this document, it is typically successes rather than failures that are published.

Figure 18.5
Evaluation Steps (Led by Evaluator in Consultation with Distance Education Staff)
 (L. Goodyear, personal communication, September 16, 2022)

Steps	Think About
<p>Determine the purpose of the evaluation</p>	<ul style="list-style-type: none"> • What do funders want to know? • What does this distance program want to know? • How will this information be used (e.g., program improvement, determining whether a distance program should continue or be shut down)? • Who will use this information and for what purpose? Not who is interested in the findings, but who will actually use them (Rossi et al., 2004, p. 91)? • What will this audience want to know exactly (Rossi et al., 2004, p. 91)? Once evaluation questions have been determined, they can be ranked in order of importance. • What information is required to answer these questions (Rossi et al., 2004, p. 91)?
<p>Decide on an appropriate evaluation approach</p>	<p>Options might include the following:</p> <ul style="list-style-type: none"> • Utilization-focused evaluation. This evaluation approach has a beneficiary or an audience to whom it provides information that is useful and usable, involving members of this audience in the planning and performance measurement of the evaluation and granting them both the responsibility and the authority to make or oversee changes in the distance learning program based on the evaluation's findings (Patton, 2008). • Theory-driven evaluation. This evaluation approach focuses on the contextual or holistic assessment of a program based on the conceptual framework of program theory (SAGE, n.d.). • Developmental evaluation. This evaluation approach supports the use of evaluation tools, empirical data, and critical thinking in frequent cycles, working in close collaboration with program actors in a process of adaptive learning (United States Agency for International Development, n.d.). • Culturally responsive evaluation. This evaluation approach places culture and the community of focus at the center of the evaluation, helps to support community empowerment, and has a goal of social justice (SAGE, n.d.).
<p>Decide on and develop the evaluation questions that will guide the inquiry</p>	<p>Examples of potential questions include:</p> <ul style="list-style-type: none"> • What are teachers' perceptions of their greatest needs? • Did teachers participate in blended courses as envisioned? Why or why not? • Is there empirical evidence that teachers improved instruction in X?

Steps	Think About
<p>Based on the evaluation questions, determine the need for benchmarks or indicators</p>	<p>Evaluations often will develop measures, metrics, benchmarks, or indicators to be used to ground the inquiry.</p> <p>An <i>indicator</i> is a piece of information that communicates a certain state, trend, or progress to an audience. It defines the data to be collected to measure progress, so that the actual results achieved can be compared with the originally designed results. Core indicators are context-specific ways to understand inputs and outcomes of a program or project that we may or may not be able to observe directly, such as the following:</p> <ul style="list-style-type: none"> • Input indicators: For example, the type of ICT equipment and/or software and/or organizational design features of a distance education program • Outcome indicators: For example, student and teacher effects (affective, cognitive, and behavioral) • National educational and socioeconomic indicators: For example, enrollment rates, literacy, and gender equity • Cost indicators: For example, fixed and recurrent costs (Kozma & Wagner, 2006). <p><i>Criteria</i> are standards by which a distance program may be evaluated. Benchmarks serve as references against which an intervention may be compared or assessed. Examples of criteria or benchmarks may include the following:</p> <ul style="list-style-type: none"> • Reach: Access to technology (i.e., devices, software, infrastructure, programming, and content) • Engagement: The extent to which users participate as intended in the programming, including participants' views of the learning experience • Outcomes: Measured changes in learning and behavior (Morris et al., 2021)
<p>Develop a design that will answer these questions</p>	<p>There are numerous evaluation designs. Three are often used in education:</p> <ol style="list-style-type: none"> 1. Case study design: An in-depth descriptive analyses of a particular person, set of persons, or program. <ul style="list-style-type: none"> ◦ Targets a small set of learners who have performed at various levels as a result of the distance learning program and examines the factors that affected their rates of success (Bamberger & Mabry, 2019; Kratochwill et al., 2010) ◦ Pros: Particularly appropriate for generating information in applied fields. It generates rich "stories" of the characteristics, enabling factors, and interventions that contribute to change. ◦ Cons: Case study findings can't be generalized beyond that case study. If not done in depth or impartially, they become little more than public relations.

Steps	Think About
<p>Develop a design that will answer these questions (continued)</p>	<p>2. Randomized controlled trial (RCT): Randomly assigns teachers, for example, to a treatment group who participates in the distance course and to a control group that does not.</p> <ul style="list-style-type: none"> ◦ Pro: They can compare or contrast these groups, potentially answering the question, <i>Did the activities implemented lead to the outcomes documented and with what certainty?</i> (L. Goodyear, personal communication, September 16, 2022). ◦ Cons: Time, expense, and issues with external validity, and also many questions cannot be answered with an RCT. They often are restricted by how many participants researchers can manage or how long participants can be expected to operate in a controlled condition. They are particularly difficult for new educational technologies or products that are rolled out so quickly and change so rapidly, because outcomes may take a long time to appear. They emphasize mean effects and de-emphasize contextual variables, yet context matters in introducing modern technologies. With “so many variables at play, it is difficult to determine whether the findings are replicable” (Van Nostrand et al., 2022, p. 4). <p>3. Pre-test and post-test group designs: Pre-distance course and post-distance course assessments on the same group of learners. The baseline score serves as the comparison group (or counterfactual).</p> <ul style="list-style-type: none"> ◦ Pros: The advantage of such a design is that it works well in isolated areas where there’s no risk of contamination, and it can provide an approximate estimate of project impacts (Bamberger & Mabry, 2019, p. 225). ◦ Cons: No control or comparison group, so no way of judging whether the process of pre-testing actually influenced the results. Does not measure the exact magnitude of the impact of an intervention or changes over time (Bamberger & Mabry, 2019, p. 225).
<p>Decide on (1) the appropriate methods to collect data and (2) the kind of data required to answer the above questions</p>	<p>Questions to consider:</p> <ul style="list-style-type: none"> • Who is the intended audience and what specific information do you hope to get from its members? • What method of data collection is best suited for obtaining the information that you need from this audience? • When will the information be collected and by whom? (Bonney et al., 2011, p. 53.)
<p>Develop, pilot, and revise data collection instruments</p>	<ul style="list-style-type: none"> • This step can include surveys, focus group protocols, or observation tools. • Pay attention to wording and language of data collection tools. • Ensure reliability through test-retesting and other measures. • Pilot with a small group and revise accordingly. • Ensure that directions for use are clear and easy to follow.

Steps	Think About
Collect data	<ul style="list-style-type: none"> • Train data collectors in appropriate data collection methods: for example, how to conduct classroom observations, where to sit, length of observation, how to accurately complete forms, and other data-collection related activities. • Develop and implement data collection protocols for standardization, for example, the scripts for interviews and focus group protocols that discuss the purpose of data collection, how the collection of personal data will be limited, and how data will be stored, as well as asking for verbal assent and promising confidentiality. (See information on IRB in the next section.) • Make data collection easy on the audience: Be sure to use brief, clear questions; short surveys; and simple vocabulary. • Allow sufficient time for all data collection. • Avoid leading questions.
Analyze the data collected	<ul style="list-style-type: none"> • Data analysis can take many different forms and can rely on different methodologies. It may be qualitative data, that is, data derived from interviews or focus groups; it may be quantitative, gathered from surveys or a high-inference (Likert scale) classroom observation; or it may be a mix of qualitative and quantitative (Bonney et al., 2011, p. 53). • Evaluators will analyze and interpret the data in light of the evaluation questions and the outcomes (anticipated and unintended) and will make larger sense of the data (Bonney et al., 2011, p. 53).
Write up findings	<ul style="list-style-type: none"> • Pay attention to writing: Present clear, concise findings and avoid whitewashing and obfuscations; be open about failures and what can be learned from them; and unpack methodologies, explaining them in layperson terms with examples and analogies. • Consumers of the evaluation have to be able to understand the strengths and limitations of the methodologies employed. Above all, the evaluation should tell a compelling story (Burns, 2020). • Include and pay special attention to the executive summary, since it may be the only part of the evaluation that funders have time to read (Bonney et al., 2011).
Disseminate findings	<ul style="list-style-type: none"> • If evaluations are published, ensure they can be found on open access platforms and distributed to the donors and implementers who funded and did the actual work (Burns, 2020). • Evaluation budgets may have to include money for publishing the most salient findings in straightforward, digestible language accompanied by a clear explanation of methods, intuitively presented data, and usable evidence (e.g., visuals, graphics, data dashboards, explanatory briefs) (Burns, 2020).

This should be part of an overall evaluation plan the evaluator develops in consultation with distance providers.

18.7 Technology and Evaluation

Technology can play an integral role in every step of the evaluation process outlined above, and distance courses undertaking an evaluation will want to consider the types of technology tools required as part of the evaluation process.

Technology can allow for group consensus and decision-making through polling software and flow charts. It can enable “multi-modal” data collection, expanding not simply *what* types of data are collected but *how* they are collected (via distance or in-person) and *who* participates in the process (Morris et al., 2021, p. 32). Technology can generate data—for instance, LMSs create metrics, such as course completion rates, grades, or the amount of time spent on a reading or activity, to provide information on engagement and learner progress. Over time and with enough learners, this information can coalesce into large data sets that provide programs with the power to evaluate every aspect of their online courses’ quality and effectiveness and to adjust accordingly. Commercial tools such as *CourseEval HQ* can serve as a vehicle for evaluations as well as aggregating, analyzing, and reporting data. For television and radio broadcasts, evaluators can access program viewing data, and for print-based distance learning, they can find the number of learning packets delivered, whether they were delivered via the Internet or mobile phones, or the number of downloads of a teaching guide or video.

Technology may be most helpful in terms of data capture, storage, and analysis. In the case of survey design, evaluators may use the survey tools within an LMS or specific survey tools such as *SurveyMonkey* or *Qualtrics*. To access the experiences and views of learners with low

literacy abilities, or who have vision or hearing loss, evaluators can use other technology tools, such as interactive voice response, phone calls, screen readers, computer-assisted telephone interviews, and SMS or text messages (Morris et al., 2021, p. 32). For example, Worldreader uses photo messages sent through a text message or *WhatsApp* to measure whether learners participated as intended in their programs (Morris et al., 2021, p. 15). These data can be analyzed in applications such as *Excel* software, the open-source *R*, or *SPSS Statistics*.

Focus groups and observations can be recorded via webinar applications such as *Zoom*, voice-to-text applications, or audio tools, or they can be conducted via *WhatsApp*. Although limited by the screen’s field of vision, classroom observations can be conducted via *Zoom* or *Webex* (although much more efficiently via *Swivl* cameras), and teacher practices can be captured via digital images and then annotated with tools such as *ThingLink*.

Qualitative research tools such as *Dedoose* and *NVivo* can allow evaluators to code images, audio, and video recordings as well as text-based interviews, and then analyze these data. The coding schemes that drive this data analysis often can be developed in a word processor, spreadsheet, or offline. If an evaluation company has a spatial data analyst, they can use a geographic information system (GIS)—a database application with mapping capabilities—to spatially analyze data and create information-rich evaluation data. For example, they can create point data to show the schools where teachers have participated in distance trainings or, as in the case of Farm Radio international, polygon data that maps the contours of coverage zones for “Her Farm Radio” broadcasts in Ethiopia, Malawi, Tanzania, and Uganda (Morris et al., 2021, p. 32; see also Farm Radio International, 2017).

Finally, for the purposes of data reporting, evaluators can use Web-based data visualization⁶ applications and data dashboards that allow evaluators to present quantitative evaluation findings in an attractive and intuitive format. These range from free and open-source tools such as *Kobo Toolbox*, *BatchGeo*, and *Tableau Public* to fee-based *Gapminder* to Microsoft's *Power BI*, part of the Microsoft Power Platform.

18.8 Final Evaluation Considerations

Evaluations are complex and time-consuming, particularly in large-scale government or donor-funded distance programs. When done well, openly, thoroughly, in partnership between evaluators and the distance provider, and without political pressure, evaluations offer numerous benefits to distance program designers, instructors, teacher-learners, and ultimately to students. This section offers some final advice.

18.8.1 Ensure Stakeholder Agreement

As the reader has probably surmised from the various chapters of this guide, distance education programs have multiple stakeholders—ministries of education, district, or regional education offices, institutional or organizational leaders where courses are housed, program managers, funders, course designers, course instructors, teacher-learners, course coaches, students, community members, parents, taxpayers (in many cases), or educational technology company representatives. Like the proverbial blind man with the elephant, these stakeholders may hold one particular view and have one dominant priority regarding the distance program and the purpose and scope of an evaluation. For instance, course designers may be most interested in improvement-oriented formative evaluation, while funders may prioritize summative or accountability-focused evaluations (Bonney et al., 2011, p. 15).

Figure 18.6 Evaluation Standards

This guide has repeatedly emphasized the importance of standards in ensuring a certain degree of quality in all distance education-related processes. The same holds true for evaluations. There are numerous evaluation standards from which to draw. For example, the [Joint Committee on Standards for Educational Evaluations](#) outlines a set of evaluation standards that include the following:

- *Utility*: To ensure that stakeholders find evaluation processes and products valuable in meeting their needs
- *Feasibility*: To increase evaluation effectiveness and efficiency
- *Propriety*: To support what is proper, fair, legal, right, and just in evaluations
- *Accuracy*: To enhance the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgments about quality
- *Accountability*: To encourage adequate documentation of evaluations and a meta-evaluative perspective focused on improvement and accountability for evaluation processes and products (Yarbrough et al., 2011)

The American Evaluation Association offers a free [rubric of evaluator competencies](#), along with an [explanatory guide](#) (Minnesota Evaluation Studies Institute, 2018). Finally, the [United Nations Evaluation Group \(2016\) norms and standards for evaluation](#), available in multiple languages.

Therefore, stakeholders and evaluators must have open conversations to agree on the goals and intended purposes for a project's evaluation. Then the evaluator will be able to determine the best approaches and methods to conduct one or more studies. Sometimes data collected from

⁶ For some examples of data visualization in action, see <https://towardsdatascience.com/the-10-best-data-visualizations-of-2022-3e49d7ccb832>

participants can be used for multiple evaluation purposes (Bonney et al., 2011, p. 15).

18.8.2 Evaluate Before Scaling

Many government-financed distance programs feel pressure to scale, if for no other reason than to ensure equity in professional development. Yet they often fail to evaluate their models, practices, or innovations to see if they are worth continuing or scaling (Duflo, 2004). Thus, for distance education programs that wish to scale, evaluations, particularly RCTs, are extremely valuable: They determine impact, and their information can be shared with others. The benefits of knowing which programs work and which do not—and which elements of a program work, for whom, and under what conditions—are important public goods “in the sense that they can offer reliable guidance to schools, universities, teacher education programs, ministries of education, funders, and nonprofits in their ongoing search for effective programs” (Duflo, 2004, p. 342).

18.8.3 Follow Government and Funder Guidelines

Summative assessment approaches must follow established best practices in evaluation. For example, Morris et al. (2021) remind international education implementers using United States Government funds to follow the United States Agency for International Development’s Collaborating, Learning, and Adapting guidelines and guidance “whether assessments are conducted in person or remotely” (Morris et al., 2021, p. 5; see also United States Agency for International Development, n.d.).

One area that deserves particular mention here is institutional review board (IRB) approval for working with a “human subject,” which is the person from whom an evaluator obtains data. The purpose of IRB guidelines is to protect the safety, rights, and welfare of individuals participating

as subjects in research and evaluation. This is particularly germane for evaluations that might be obtaining data from “vulnerable populations,” such as teachers or students in areas of conflict and crisis, in authoritarian regimes, in areas with gang violence, and in refugee contexts, or from teachers or students who are undocumented in terms of immigration status or who may be members of persecuted religious, ethnic, or sexual minority groups. IRB also is a process that often falls through the cracks in many international development education programs. The concept of IRBs evolved from the 1964 Helsinki Declaration to protect human subjects as part of medical research. The Helsinki Declaration was influenced by the Nuremberg Trials of 1945–1946.

In the United States, IRBs are administered on a federal level by the Office for Human Research Protections (OHRP), an office within the U.S. Department of Health and Human Services. OHRP assists IRBs in their work, and it receives and investigates claims of inappropriate research practices (United States Department of Health and Human Services, 2023.). The U.S. Department of Health and Human Services offers free, self-paced IRB training. Many educational organizations and universities will have IRBs whose job it is to ensure compliance with IRB guidelines. These typically require evaluators to submit a detailed description of the evaluation project, the list of human subjects to be surveyed or interviewed, the data collection instruments to be used, strategies for recruitment of evaluation participants—including evidence of consent forms—and a plan for ensuring confidentiality and protecting data. All of this is time-consuming and should be built into an evaluation timeline. A full checklist of U.S. government-funded program IRB requirements can be found at the Department of Health and Human Services: IRB Written Procedures site.⁷

⁷ See <https://tinyurl.com/cwkm8rve>. A full list of various countries and regions and their ethical research requirements can be found here: <https://www.hhs.gov/sites/default/files/ohrp-international-compilation-2021.pdf>.

18.8.4 Align Expectations, Outcomes, Activities, and Evaluation

Evaluations often yield disappointing results, the origin of which has multiple causes: lofty goals combined with parsimonious program activities; ambiguous outcomes that are neither empirical nor measurable; the slow rate of teacher change and the problem of enactment (both discussed in Chapter 16) (Hord et al., 2006); inflated program expectations in terms of the design of the distance program; the gap between resources and expectations (Bonney et al., 2011); the fact that many intended effects are distal—measurable only in the long term—versus proximal (measurable soon after a distance course) (Gaible & Burns, 2007); and the fact that the effects of teacher professional development—student achievement and improved teaching—are incremental and are not completely visible by the end of a distance program (Kennedy, 2016, pp. 3–4).

To truly capture the effects of an intended program, distance education programs would do well to temper expectations about what a distance course can reasonably achieve in a set amount of time and develop clear, measurable learning outcomes that are reflected in course activities and in the evaluation design. Evaluators and distance educators may also have to moderate their expectations about what evaluations can reasonably measure during the lifespan of a course or program. Funders may have to help by supporting evaluations that continue “beyond the lifespan of a particular distance education program, (focusing) not just on an intervention but on the systems and stakeholders that influence learning transfer” (Kennedy, 2016, p. 4).

18.8.5 Exercise Caution in Consuming Evaluations

The above considerations deal with producing evaluations. A final consideration focuses on the consumption of evaluation information. Stakeholders, funders, and distance course developers often consult evaluations of evaluations as part of decision-making. These “meta-evaluations” often report measures

of the impact or magnitude of the effect of a program on an outcome as an “effect size.” These effect size magnitudes have traditionally been interpreted based on rules of thumb suggested by Cohen (1988), in which an effect size of approximately 0.20 is considered “small,” 0.50 is considered “medium,” and approximately 0.80 is considered “large.” This standardized form of effect sizes is useful because it allows comparison of the magnitude of effects on different outcome variables and across different studies. Reporting effect sizes also allows other researchers to conduct meta-analyses and helps funders determine whether the difference between two distance programs or educational technology products is meaningful or not (Bakker et al., 2019).

Despite the value of effect sizes, however, evaluators have long cautioned that educators and decision-makers should interpret them carefully in making determinations about the overall effectiveness of an approach or distance modality. Effect sizes are broad generalizations that may have more to do with the design of a study than with the intervention itself, and they may have less to do with whether or not these magnitudes of effects are substantively or practically important (Bakker et al., 2019; Hill et al., 2008). A “small” effect size may hold enormous practical significance—the risk is in dismissing an approach by looking at effect sizes alone (Hill et al., 2008; Kraft, 2018). Evans & Yuan (2020), in a study of 156 RCTs that measured learning outcomes, note that effect sizes for changes to learning tend to be small to medium. A number of researchers have proposed that educators and decision-makers use different guidelines and new frameworks for interpreting effect sizes (Bakker et al., 2019; Hill et al., 2008; Kraft, 2018).

18.9 Conclusion: “The Greatest of Mistakes Is to do Nothing Because You Can Only Do a Little”

Evaluation is one of the most critical factors in the success of a distance learning program. However, despite the importance of measuring

effects and uncovering evidence, distance education programs across the globe often face time, resource, access, and political constraints—realities that adversely impact the quality and utility—indeed, the ability—to conduct systematic evaluations of a distance education program.

As this chapter has detailed, evaluation is important, not just for a distance program itself, but for the fields of teacher pre-service and in-service education and distance education writ

large. More tangibly, it is critical for teaching quality, which, as discussed in Chapter 8, is the most important school-related factor in a child's education. Distance education programs that embrace evaluations may do much to ultimately improve a child's learning; programs that eschew evaluations ultimately harm the quality of that learning. As the quote⁸ that frames this conclusion admonishes, the gravest mistake a distance program can make is to fail to evaluate.

⁸ Attributed to the English cleric, the Reverend Sydney Smith (1771–1845).

Citation: Burns, M. (2023). Evaluating Distance Education Programs. In *Distance Education for Teacher Training: Modes, Models and Methods*. (2nd Edition). Washington, DC: Education Development Center.

References

- Bakker, A., Cai, J., English, L., Kaiser, G., Mesa, V., & Van Dooren, W. (2019). Beyond small, medium, or large: Points of consideration when interpreting effect sizes. *Educational Studies in Mathematics*, 102(1), 1–8. <https://doi.org/10.1007/s10649-019-09908-4>
- Bamberger, M., & Mabry, L. (2019). *Real world evaluation: Working under budget, time, data, and political constraints*. Sage Publications.
- Bonney, R., Ellenbogen, K., Goodyear, L., Hellenga, R., Phillips, T., Russell, L., . . . Yalowitz, S. (2011). *Principal investigator's guide: Managing evaluation in informal STEM education projects*. Center for the Advancement of Informal Science Education. <https://www.informalscience.org/activities/principal-investigators-guide>
- Brookhart, S., & Nitko, A. (2011). Assessment of higher order thinking skills. In G. Schraw & D. Robinson (Eds.), *Strategies for constructing assessments of higher order thinking skills* (pp. 327–359). Information Age Publishing.
- Burns, M. (2020). From theory to use: Making educational research more usable and useful for practitioners. *Annual Review of Comparative Education 2019 International Perspectives on Education and Society*, 39, 47–53. doi:<https://doi.org/10.1108/S1479-3679202039>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Academic Press.
- Duflo, E. (2004). Scaling up and evaluation. *Annual World Bank Conference on Development Economics* (pp. 341-369). World Bank.
- Evans, D. K., & Yuan, F. (2020). *How Big Are Effect Sizes in International Education Studies? CGD Working Paper 545*. Global Development. <https://www.cgdev.org/sites/default/files/how-big-are-effect-sizes-international-education-studies.pdf>
- Farm Radio International. (2017, June 30). *Final report: Her farm radio in Ethiopia, Malawi, Tanzania and Uganda*. Farm Radio International. <https://farmradio.org/wp-content/uploads/2020/09/143-Her-Farm-Radio-Final-Report-to-IFAD-Sept-2017-2.pdf>
- Gaible, E., & Burns, M. (2007). *Using Technology to train teachers: Appropriate uses of ICT for teacher professional development in developing countries*. InfoDev | World Bank. <https://www.infodiv.org/articles/using-technology-train-teachers>
- Guskey, T. (2000). *Evaluating professional development*. Corwin Press.
- Guskey, T. (2016, February). Gauge impact with 5 levels of data. *Learning Forward*. <https://tguskey.com/wp-content/uploads/Professional-Learning-1-Gauge-Impact-with-Five-Levels-of-Data.pdf>
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172–177. doi:10.1111/j.1750-8606.2008.00061.x
- Hord, S., Rutherford, W. L., Huling-Austin, L., & Hall, G. (2006). *Taking charge of change*. Southwest Educational Development Laboratory.
- Kennedy, M. (2016). How does professional development improve teaching? *Review of Educational Research*, 86(4), 945–980. doi:<https://doi.org/10.3102/0034654315626800>

Distance Education for Teacher Training: Modes, Models, and Methods

- Kozma, R., & Wagner, D. (2006). Core indicators for monitoring and evaluation studies in ICTs for education. In D. Wagner, B. Day, T. James, R. Kozma, J. Miller, & T. Unwin, *Monitoring and Evaluation of ICT in Education Projects* (pp. 21-32). World Bank.
- Kraft, M. (2018). *Interpreting effect sizes of education interventions: Brown University working paper*. Brown University. https://scholar.harvard.edu/files/mkraft/files/kraft_2018_interpreting_effect_sizes.pdf
- Kratochwill, T., Hitchcock, J., Horner, R., Levin, J., Odom, S., Rindskopf, D., & Shadish, W. (2010). *Single-case designs technical documentation*. U.S. Department of Education What Works Clearinghouse. <https://files.eric.ed.gov/fulltext/ED510743.pdf>
- Leahy, M., Thielsen, V., Millington, M., Austin, B., & Fleming, A. (2009). Quality assurance and program evaluation: Terms, models, and applications. *Journal of Rehabilitation Administration*, 33(2), 69–82. https://digitalcommons.usu.edu/sped_facpub/201/
- Mindtools. (n.d.). *Kirkpatrick's model: Four levels of training evaluation*. Mindtools. <https://www.mindtools.com/pages/article/kirkpatrick.htm>
- Minnesota Evaluation Studies Institute. (2018). *Minnesota Evaluation Studies Institute*. American Evaluation Association Evaluator Competencies. <https://www.eval.org/Portals/0/Self%20assessment%201oct.pdf>
- Moon, B., Leach, J., & Stevens, M. (2005). *Designing open and distance learning for teacher education in Sub-Saharan Africa: A toolkit for educators and planners*. World Bank. http://oro.open.ac.uk/8401/1/teacher_education_toolkit_may13.pdf
- Morris, E., Farrell, A., & Venetis, E. (2021, January). *A roadmap for for measuring distance learning: A review of evidence and emerging practices*. United States Agency for International Development: https://www.edu-links.org/sites/default/files/media/file/Measuring%20Impact%20and%20Outcomes_Final_01.20.2021-508%20%281%29.pdf
- Patton, M. (2008). *Utilization-focused evaluation* (4th ed.). Sage.
- Rossi, P., Lipsey, M., & Freeman, H. (2004). *Evaluation: A systematic approach* (7th ed.). Sage.
- SAGE. (n.d.). *The SAGE encyclopedia of educational research, measurement, and evaluation*. SAGE Research Methods. <https://methods.sagepub.com/>
- Scriven, M. (1991). *Evaluation thesaurus*. Sage.
- Scriven, M. (2009). *The evaluation of training*. The Evaluation Center Western Michigan University: <https://wmich.edu/evaluation/checklists>
- United Nations Evaluation Group. (2016). *Norms and standards for evaluation*. United Nations Evaluation Group. <http://www.unevaluation.org/document/detail/1914>
- United States Agency for International Development. (n.d.). *CLA toolkit*. Learning Lab. <https://usaidlearninglab.org/cla/cla-toolkit>
- United States Department of Health and Human Services. (2023). *Office for human research protections*. <https://www.hhs.gov/ohrp/index.html>
- Van Nostrand, P., Noakes, S., Shah, Z., & Luna, L. (2022, February). *An overlooked indicator of edtech quality: The use of learning sciences research*. Digital Promise. <https://doi.org/10.51388/20.500.12265/150>
- World Food Program Office of Evaluation. (2021, February). *Technical note principles, norms and standards for evaluations*. Evaluation Quality Assurance System (EQAS). <https://docs.wfp.org/api/documents/WFP-0000003179/download/>
- Yarbrough, D., Shulha, L., Hopson, R., & Caruthers, F. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Sage.

