TECHNICAL APPENDIX: CONSIDERATIONS FOR MEASURING T/TA PERFORMANCE

By Jesse Chandler, Mathematica
ASPE Project Officers: Ali Abazeed, Amanda Benton, and Lindsey Hutchison

This document serves as a technical appendix to the toolkit, "Is Your Training and Technical Assistance (T/TA) Effective? Considerations for Measuring T/TA Performance." It is part of a project on Measuring T/TA Effectiveness. Visit https://aspe.hhs.gov/measuring-tta-effectiveness for more resources.

Introduction

The federal government invests substantial resources in state and local programs to address poverty and child well-being. Many of these grants include training or technical assistance (T/TA) components to increase the ability and capacity of organizations and communities to improve the circumstances of children, families, and communities.

Program T/TA is assistance designed to transfer knowledge, skills, or other capacities that will help recipients identify and address solutions to challenges they face. T/TA could be narrowly focused on solving immediate identifiable challenges or more broadly focused on increasing resources such as skills or relationships that are available to face future challenges. T/TA provided to human services programs typically comes in many forms, which vary in intensity and the level of engagement between the provider and recipient.[1] T/TA providers may develop briefs and other products designed for passive consumption or they may actively engage with T/TA recipients, either in person or through virtual means. Providers may either target groups with similar interests or tailor their services to the specific needs of individuals. While many T/TA programs focus on transferring knowledge from the T/TA provider to the recipients, other models focus on facilitating the transfer of skills and knowledge between peer groups with similar challenges.

T/TA has the potential to improve programs and services, if providers appropriately design and deliver it and it matches the recipient's readiness for T/TA. In particular, effective T/TA must:

(1) Target skill or knowledge gaps that affect the desired outcome;
(2) Be delivered to programs or organizations with a culture supporting knowledge transfer (though T/TA can also address organizational culture);
(3) Be delivered to an audience within the organization that can use the knowledge to achieve the desired goals; and
(4) Be of sufficient quantity ("dose") and quality to enact longstanding change[2].

[1] For a detailed discussion of the design choices that vary across T/TA offerings see Baumgartner, Cohen, & Meckstroth (2018).

[2] Mitchell, Florin, & Stevenson (2002) provide an in-depth review of the necessary characteristics of effective T/TA.

Tracking T/TA activities and linking them to learners' outcomes and program performance can support several interrelated objectives. T/TA providers can use that information to improve the **quality** of T/TA. From a **program management** perspective, organizations charged with overseeing T/TA can better understand their activities and the effect these activities are having on recipients and the organizations they work for. From an **accountability** perspective, agencies devoting resources to T/TA can assess what the resources expended on T/TA are producing. Individual T/TA providers can also use information about T/TA delivery and outcomes to improve the **effectiveness** of their offerings.

During the design and implementation of T/TA, measurement can help individuals who evaluate, design, oversee, or fund T/TA identify and address problems with planned T/TA delivery. For example, materials can be pilot tested and revised based on feedback and surveys collected during or after training sessions. This would allow T/TA providers to **adjust the content**, **delivery mechanisms, quality, or dosage of the training** to better meet recipients' needs. Evaluations of completed T/TA efforts can be used to iteratively improve individual efforts or inform future T/TA offerings.

There are advantages to using a set of **shared measures** to achieve all of these purposes. Shared measures allow the cost of measure testing to be shared by multiple programs. Shared measures can also provide a benchmark that can contextualize the scores of any individual effort. If common measures are used to assess different T/TA offerings, funders of T/TA (including HHS) can have a better understanding of the overall effectiveness of their T/TA efforts. In addition, for evaluation purposes, responses can be pooled into a single, large sample across multiple T/TA efforts to evaluate more general questions such as the effectiveness of particular modes of delivery, or the impact of T/TA services on otherwise rare subgroups. Despite these advantages, shared measures will not always be appropriate, particularly if the measures are not aligned with the expected outcomes of a T/TA effort.

To better understand the available options for measuring T/TA delivery, the Office of the Assistant Secretary for Planning and Evaluation in the U.S. Department of Health and Human Services commissioned Mathematica Policy Research to (1) conduct a scan of the academic and gray literature on T/TA measurement and (2) describe typical and best practice methods for measuring program T/TA. Based on this scan, we provide a summary of the design choices that are important when developing a T/TA measurement strategy and present examples of these practices.

Framework for Measuring T/TA

When measuring the performance of a program or intervention, the standard approach is to develop a logic model.[3] A logic model provides the theory of how the program will achieve the desired outcome by linking the inputs and activities of a program to its outputs and outcomes. Likewise, a logic model can also govern T/TA delivery by specifying how technical assistance is expected to affect staff behavior, and in turn how this will affect organizational or customer outcomes. Measuring each link in the T/TA logic model can provide a sense of the overall

[3] See Hatry (2006) and W. K. Kellogg Foundation (2004).

performance of the program. The pathway by which T/TA achieves its goals involves several steps and several sets of stakeholders. Although these steps overlap, they are approximately linear. We refer to them as "stages" for convenience.

- The providers of the T/TA must develop a sufficient amount of high quality training content (an input) and deliver this content effectively (an activity) to learners.
- Learners (an input) must engage with materials and retain the information they learn (an activity).
- Finally, learners must transfer their skills to their jobs, and organizational outputs and outcomes must improve (an outcome).

Each of these stages of T/TA delivery provides an opportunity to evaluate the effectiveness of T/TA as a part of formative or summative evaluation. In practice, earlier stages are of greater interest for formative evaluations because they can be collected quickly; whereas, later stages are of greater interest for summative evaluations because they more accurately capture intended outcomes.

1. Measuring Content Development and Delivery of T/TA

It is important to measure the content that T/TA providers deliver. Both the **quality and quantity** of content should be measured. How material is delivered can be evaluated against established best practices in adult learning.[4] If the knowledge, skills, or capacities to be conveyed through T/TA are well defined and specified, the content can be checked against the learning objectives to ensure that all necessary topics are addressed. The quality of materials can be measured against an objective standard. For example, experts can review (and potentially score) training materials or trainer notes to determine which content reflects best practices. T/TA is not always delivered to learners as designed, so evaluators may also wish to measure the **fidelity** of the training and coaching sessions.[5]

The quantity of T/TA provided can also be measured and used to evaluate the productivity of T/TA providers. **Quantity measures focus on the volume of materials produced**, such as the number or length of webinars or documents, or the frequency and intensity of coaching sessions. Quantity measures are appealing because they are easily available, quantitative (by definition) and objective. Quantity measures can be useful when it is important to measure the visibility of T/TA providers or the potential for their clients to access information. Quantity measures are related to training outcomes in that T/TA content must exist in order to be used. However, they should not be used as a substitute for measures of quality or effectiveness. More isn't always better, especially if the content that is produced is not wanted or does not change staff knowledge and behavior or organizational performance.

In systems that use networks of T/TA providers to address complex problems, content delivery is often dependent upon the ability of T/TA providers with different areas of expertise to provide

[4] For an overview, see Merriam & Bierema (2013).

[5] Carroll et al. (2007), Durlak, (2013).

integrated services, avoid duplication of services, share expertise internally, or refer those seeking T/TA to the correct provider. The effectiveness of these systems can be evaluated by examining the **extent to which individual T/TA providers coordinate with one another** to deliver services. The degree of coordination between providers can be measured subjectively by surveying providers about their interactions with each other or through network analysis of the communication traffic between providers or the flow of trainees between T/TA offerings.

2. Measuring Recipient Use of T/TA

T/TA materials that are not used cannot have an impact. In other words, only people who participate in T/TA can use the knowledge to improve outcomes. Usage statistics measure the **number of times that T/TA resources are accessed** and can provide an estimate of the maximum potential impact of T/TA. Examples of usage statistics include the number of people who attend a webinar, the number of times a resource is viewed or downloaded, or the number of unique users who viewed or downloaded a resource. Usage statistics that measure **engagement with T/TA providers** or facilitators are **easy to collect** and do not impose a burden on T/TA recipients. T/TA activities that are intended to foster interaction (such as peer-to-peer sharing of best-practices) can be evaluated according to whether they increase peer interactions, either through surveying recipients or through measures of their behavior (such as communication logs on a discussion board). These data collection measures can be used to count individuals' peer-interactions or to model changes in the connectedness of recipients to each other through a network analysis of peer interaction. Like other quantity measures, a major limitation of usage statistics is that they **do not measure whether recipients actually learn anything** from the T/TA experience. In some cases, such as coaching, frequent usage could even indicate that learners are using the sessions to support specific tasks without improving their knowledge.

Measurement is only informative when compared to some standard. Content measures have plausible standards (such as meeting all key benchmarks). For usage statistics, it can be particularly **difficult to establish appropriate standards of comparison** because different forms of T/TA are directed to audiences of different sizes. Usage statistics can be informative if audience sizes are the same or if the statistics adjust for differences in size. For example, use of the same T/TA initiative or use of different offerings targeted to the same audience over time can provide useful insight. Comparisons between offerings targeted at different audiences should be done with great caution, especially when the potential audience sizes are usually unknown.

3. Measuring Recipient Response to T/TA

Participant satisfaction with T/TA provides one indication of its quality. Responses assess T/TA **recipients' attitudes or feelings** following a training experience. They are by far the **most common** measure of learning outcomes, perhaps because they are **easy to collect**.[6] Although initially defined as an affective response to a training experience, later researchers developed more expansive and detailed definitions of responses. For instance, **enjoyment of training**, the **perceived usefulness** of training, and the **perceived difficulty** of training are all conceptually

[6] Antle, Barbee, & van Zyl, (2008); American Society for Training & Development (ASTD, 2003).

distinct.[7] Evaluators could also be interested in attitudinal changes beyond evaluations of the T/TA experience itself. For example, training can influence both motivation—believing the change is important—and self-efficacy—believing that change is possible—, both of which are related to favorable training results.[8] Evaluators typically measure responses through surveys, focus groups, and cognitive interviews; sometimes they infer responses from analysis of user engagement.

4. Measuring Recipient Learning from T/TA

Evaluators can measure changes within T/TA recipients that they believe to be necessary preconditions of organizational improvement and recipient readiness for other types of T/TA.[9] Learning consists of objective evaluations of T/TA recipients' performance. Learning outcomes include increased **knowledge or skills**. Measures of skill and knowledge are only modestly related to each other, so measurements of one are a poor indicator of the other. This distinction is important because as discussed in Section C., skill measures are predictive of results of skill transfer and organizational change but knowledge measures are not. Learning is typically measured through structured tests of T/TA recipient skills or knowledge.

5. Measuring Results of T/TA

There are several different ways to measure the results of T/TA. Which approach(es) are appropriate will depend on the specific initiative goals and activities, as well as available data sources.

Transfer consists of skills or knowledge transferred to T/TA recipients' job performance. Measures of transfer are conceptually distinct from learning measures such as skill acquisition because they concern **behavior change in the intended context** and outside of an explicit evaluation. Transfer can be measured through surveys of T/TA recipients and peers, and third-party observation of job performance. Measures of transfer **must be customized** to align not only with the contents of T/TA materials but also with learners' particular roles and responsibilities.

Organizational outcomes include changes in performance of overall measures of success for the organization that T/TA is intended to achieve. Training consultants and researchers typically recommend that T/TA providers design offerings with clear outcomes that support the logic model of a particular service or program. Measures of results follow from this outcome. Ideally,

[7] Warr & Bunce (1995).

[8] Colquitt, LePine, & Noe (2000).

[9] The Kirkpatrick (1975, 1996) training evaluation model is one of the oldest and most widely used training taxonomies. As an indication of its influence, Google Scholar reports that researchers have cited the book describing this method over 10,000 times. The taxonomy of learning outcomes that it describes is widely accepted and is a useful starting point for describing T/TA intervention outcomes. It identifies three categories of learning outcomes: reactions (responses), learning, and transfer. We categorize these three separately and include transfer as one way to measure T/TA results.

organization-level outcomes are operationalized as measureable changes in performance metrics, such as the speed, amount, cost, or quality of an output, or in better outcomes for its client base.

These organizational outcomes are distinct from transfer as a way to measure results because, if the skills that T/TA providers address are not sufficient to achieve an organization's goals, workers' performance can change without affecting the organization's results. There are many reasons that skill transfer can be insufficient to change organizational outcomes. To provide a few examples:

- Transferred skills may be dependent upon other skills in order to cause change. For example, improving decisions about who should receive services requires that staff are aware of how to connect clients to the services they need.
- T/TA might not transfer skills to everyone necessary to cause change. For example, the work of staff members who follow best practices may be inadvertently undermined by other staff following a business as usual approach
- Transferred skills may have positive impacts, but not on the outcomes measured. For example, improving staff performance may improve the efficiency of outcomes or reduce the need for supervisor involvement, without changing the ultimate quality of service delivery.
- Resource limitations may constrain organizational performance despite improvements in a particular skill or practice. For example, improving decisions might not improve organizational performance if there are few resources to deliver a service that most clients need.

In all of these cases, the implication is that the logic model that specified how T/TA will impact organizational results is incorrectly specified or incomplete in some way.

Return on Investment (ROI) measures the **results in light of the cost** of T/TA. A positive ROI means that the organization is earning (or saving) more than it spent on the T/TA, while a negative ROI means that the organization is earning less than it spent on the T/TA. Evaluating ROI requires identifying tangible and intangible program benefits and converting them to a monetary value, calculating the cost of T/TA, and comparing the net benefits to the costs.[10] Evaluating ROI can be a useful exercise for several reasons: it discourages a focus on only the costs or benefits of T/TA, allows different inputs and outputs to be compared using a common metric and can help with decisions about how to allocate limited resources.

The calculation of ROI can be as simple as evaluating the cost of training and the resulting improvements in productivity. More elaborate ROI models can attempt to capture intangible, future, or second-order costs and benefits. For example, training could also improve staff retention. In turn, retaining staff could lead to savings on future hiring and training costs. Retaining staff also leads to a more experienced workforce, though the benefits of experience may be difficult to quantify. At the same time, retaining staff longer may also increase labor

[10] For an overview of ROI measures see P. Phillips & J. Phillips (2004) and J. Phillips & P. Phillips (2008).

costs. In some cases, estimates of ROI require assumptions about the costs of these hard to quantify outcomes.

An important consideration for calculating ROI is **determining whose costs and benefits should be included** in the equation. What counts as a cost or benefit could differ for different stakeholders. For example, T/TA recipients could be narrowly interested in the costs and benefits to their organization or clients. In contrast, some T/TA providers or federal agencies may take a more holistic view. Depending on the measurement objective, different analyses could plausibly include costs or savings realized by clients, partners, regulatory agencies, or society as a whole.

Choosing What to Measure

Stages to Measure

T/TA outcomes should be measured to evaluate the effectiveness of the T/TA services provided. However, evaluators must consider specific measurement goals and available resources. Measures are **useful if they directly measure important outcomes, or if they measure processes that are antecedent to important outcomes**. Performance at each stage of T/TA delivery may be of direct interest to evaluators. The availability of T/TA materials is a component of capacity building and thus an important outcome.[11] Responses to training material may be related to whether people enroll in, complete, or recommend training to others.

Evaluators sometimes measure **early stage outcomes** when the questions of interest focus on later stage outcomes, because they are **faster, cheaper, and easier** to measure, and can lead to **course corrections**. Measuring **response, learning, and results together** can **provide diagnostic** information about why T/TA is not performing as expected. For example, skills that are successfully learned but do not result in transferred application on the job may suggest confusion about when to properly apply learned skills or that there are barriers in job roles or workplace culture that prevent behavior change. Skills that are transferred but do not impact an organization's outcomes could suggest that assumptions about the causes of a challenge need to be reassessed.

Inferences about later stage outcomes based on early stage outcomes should be made with caution because the correlation between different outcomes is moderate at best.[12] For example, two meta-analyses have indicated:

- **Declarative knowledge** (such as scores on test questions) and measures of **training utility** are **less strongly related to** skills transfer than might be expected (r = 0.2);

[11] Children's Bureau (2015).

[12] A correlation expresses the association between two variables on a scale ranging from -1.0 to +1.0. A correlation of -1.0 or 1.0 means that the two variables are perfectly negatively or positively associated with each other. By convention, an effect size is considered small if the value of r is 0.1, moderate if r is 0.3, and large if r is 0.5. To provide some context for these values, for US adults, the association between weight and height is r = .44 and the association between gender and height is r = .67 (see U.S. Department of Health and Human Services National Center for Health Statistics (1996) as cited by Meyer et al. (2001)).

- While **enjoyment** of T/TA is one of the more common measures currently used in many contexts, it is virtually unrelated to skill transfer (r = .03).[13]

It is difficult to predict whether the impact of T/TA will be smaller (or larger) than its ultimate impact on skill performance. Demonstrated skill acquisition in a testing context correlates between r = 0.2 and r = 0.6 with skill performance in an employment context. Imperfect correlations suggest that knowledge measures over or under-estimate the degree of skill-transfer that results from T/TA. Although these correlations are high relative to those typically observed across the social sciences (Richard et al., 2003), they may be lower than anticipated by providers given that skill demonstrations are intended to be direct analogues to the performance issues that T/TA is intended to address.

Overall, these modest relationships limit the potential association between measures of the content provided and organizational outcomes. Moreover, within T/TA measures surveyed in this memo—and in public and private sector training evaluation more generally—measurement efforts tend to focus on the measures that are least likely to predict the desired results of T/TA such as the content that is delivered, and how T/TA recipients feel about it.[14]

Other Measurement Considerations

Effective T/TA delivery can be measured with different degrees of rigor and sophistication. The information provided by **some form of measurement is usually better than no measurement**, and many existing measures can be adapted to measure other T/TA efforts. Deciding on whether to adopt more customized or sophisticated measures of T/TA delivery requires considering the potential costs, benefits and available resources to develop, collect and analyze these data. Evaluators must also make several additional decisions about how outcomes should be measured.

A primary consideration is the **cost and availability of information**. Web statistics and administrative data are relatively cost-efficient sources that may or may not be available and may or may not provide direct answers to the questions in which evaluators are interested. Data collection methods such as focus groups, observation, surveys, or tests can be more expensive, but also offer more flexibility in measurement.

Some **methods may be more naturally suited for measuring specific stages** than others. For example, trainees' responses and learning are easily measured through new data collection, while results may be easiest to detect in organizational performance data. However, each of these stages can be measured by a variety of different methods. For example, responses can be measured through surveys or by content-coding existing content of discussion boards or reports. Results such as skill transfer can be measured from administrative records of changes in service delivery or by interviewing an organization's leaders.

---

[13] Effect size measures are drawn from meta-analyses conducted by Alliger, Tannenbaum, Bennett, Traver, & Shotland (1997) and Colquitt et al. (2000).

[14] For other surveys of T/TA measures see Antle et al. (2008), ASTD (2003).

T/TA providers can **collect measures at different times**. Where possible, collecting baseline measures of outcomes can provide a clearer indication of how learners and organizations have changed because of T/TA. Post-T/TA measures of different stages may be best collected in different periods following the delivery of T/TA. Responses and learning information collected immediately after (or even during) T/TA can help T/TA providers adjust in response to audience needs. Long-term learning (retention) of information is often lower, but retention rates are a better predictor of skill transfer. Conversely, some changes in results—particularly skill transfer and organizational outcomes—may increase over time as staff gain opportunities to practice and apply the skills they learn, or change other elements of workplace culture, practice, or policy to support their training.

Measures will differ in their **ability to capture the construct that they are intended** to represent. Measurement error will place limits on the sensitivity of measurement instruments to detect change.[15] High quality measures are designed to minimize measurement error, and ideally they test and correct for error rates. For example, survey instruments should be pilot tested for respondent comprehension and evaluated for their reliability before use.[16] Evaluators should verify, rather than assume, the quality of administrative data because records collected for reasons other than impact analysis may have quality issues that may escape the notice of evaluators.[17]

Finally, it might be important to **compare outcomes across different T/TA programs**. Comparing outcomes is difficult if different programs use different measures. The benefits of using a common set of measures must be weighed against the potential loss of flexibility to measure program-specific outcomes. The Early Childhood T/TA System has developed a common item bank consisting of measures of mandatory and optional topics that is used to assess all T/TA efforts.

Examples of T/TA Measurement

Table 1 highlights different kinds of outcome measures that evaluators have used to assess T/TA delivery and effectiveness across the stages discussed earlier. For each outcome, we provide citations that are representative of that approach. Results are based on articles discovered through searches of Google Scholar and the websites of federal agencies, states, foundations, and public policy organizations, as well as interviews with four stakeholders.[18]

[15] For a discussion of the relationship between measure quality and measure sensitivity, see Aguinis (1995).
[16] For an overview of instrument testing, see Collins (2003).
[17] For an example from Medicaid, see Peabody, Luck, Jain, Bertenthal, & Glassman (2004).
[18] The stakeholders were from the Administration for Children and Families (ACF) and the National Opinion Research Center (NORC).

Table 1. Methods of Measuring Training and Technical Assistance

Construct or topic	Example Data collection method or source	Example Outcome measurement or analysis	Example project or agency
Measuring Content and Delivery of T/TA			
Purpose of T/TA activity	T/TA documents	Qualitative description of training materials and their objectives	• The National Early Childhood Technical Assistance Center (Kahn et al., 2009)
Content of T/TA activity	T/TA records	Cross-tabulation of time spent delivering T/TA by topic	• Children's Bureau Training and Technical Assistance (Sun, Griffith, Randhawa, & DeSantis, 2017)
Quality of T/TA materials	Focus group of trainees	Qualitative summary of trainee experience	• Scaling Up a Place-Based Employment Program (Tessler et al., 2017)
Time spent delivering T/TA activity	T/TA records	Count of time spent delivering T/TA	• The National Early Childhood Technical Assistance Center (Kahn et al., 2009) • Children's Bureau Training and Technical Assistance (Sanclimenti & Caceda-Castro, 2017) • Community-based prevention programs (Hunter et al., 2009)
Time to fulfill T/TA request	T/TA records	Processing time between request submission and work plan approval	• Children's Bureau Training and Technical Assistance (Sun et al., 2017)
Mechanism through which T/TA was delivered	T/TA records	Cross-tabulation of time spent delivering T/TA by method (coaching, training, etc.)	• Children's Bureau Training and Technical Assistance (Sun et al., 2017)
	T/TA provider notes	Cross-tabulation of time spent delivering T/TA by mode (in person or by telephone or email)	• Children's Bureau Training and Technical Assistance (Sun et al., 2017) • Community-based prevention programs (Hunter et al., 2009)
Fidelity of training delivery	Survey of T/TA providers	Checklist of content to be taught in each session	• Community Youth Development Study (Fagan, Hanson, Hawkins, & Arthur, 2008)
	Observational coding	Observer assessment of T/TA fidelity	• Community Youth Development Study (Fagan et al., 2008)
Coordination between organizations providing T/TA	Survey of site administrators	Social network analysis of reported contacts with other agencies	• Children's Bureau Training and Technical Assistance (James Bell Associates & ICF International, 2015)
	Administrative records of staff interactions	Social network analysis of interactions between staff members	• Feasibility study of Educational Process Mining (Cairns et al., 2014)

Construct or topic	Example Data collection method or source	Example Outcome measurement or analysis	Example project or agency
Referral of T/TA recipients to other providers	Program data	Social network analysis of handoffs of T/TA recipients between providers	• Flow of trainees between training providers (Cairns et al., 2014)
Measuring Use of T/TA			
Number of organizations or groups that received T/TA	T/TA records	Count of sites that received T/TA	• Maternal and Child Health Services (Texas Department of State Health Services, 2016)
Number of individual T/TA recipients	T/TA records	Count of trainees that received T/TA	• Children's Bureau Training and Technical Assistance (Sun et al., 2017) • The National Early Childhood Technical Assistance Center (Kahn et al., 2009) • Community-based prevention programs (Hunter et al., 2009) • Promoting Science-Based Approaches to Teen Pregnancy Prevention (PSBA) Project (Ray, Wilson, Wandersman, Meyers, & Katz, 2012)
Use of online resources	Analysis of website user traffic	Count of number of times resources were accessed	• Private sector child care training provider (Ackerman, 2017)
Flow of T/TA recipients between T/TA offerings or efforts	Analysis of website user traffic, administrative data	Network analysis of course enrollment patterns	• Sequence of course enrollments by trainees (Cairns et al., 2014)
Frequency of TA requests	T/TA records	Count of trainee requests	• Community-university teen substance use prevention partnership (Spoth, Claire, Greenberg, Redmont, & Shin, 2007).
Demographics of T/TA users	Administrative data	Tabulation of demographic characteristics	• Children's Bureau Training and Technical Assistance (Sanclimenti & Caceda-Castro, 2017)
Measuring Responses to T/TA			
Responses to T/TA Activities—Perceived Enjoyment			
Satisfaction with T/TA provider	Survey of organizational leaders	Single-item measures of training and T/TA provider satisfaction	• Children's Bureau Training and Technical Assistance (Barbee, DeWolfe, & DeSantis, 2017)
	Survey of staff T/TA recipients	Single-item measures of training and T/TA provider satisfaction	• Healthy Relationship and Marriage Education Training (HRMET) Project (Futris, Schramm, Lee, Thurston, & Barton, 2014)
		Mean reported on an adapted consumer satisfaction measure	• Community-based prevention programs (Hunter et al., 2009)
		Mean reported on course evaluation form	• A community-based substance abuse coalition (Chinman, Hunter, & Ebener, 2012)

Construct or topic	Example Data collection method or source	Example Outcome measurement or analysis	Example project or agency
		Mean reported on Kirkpatrick Level One Training Evaluation Scale (a widely used measure of satisfaction)	• Kentucky Child and Family Services (Antle, Barbee, & van Zyl, 2008; Johnson, Antle, & Barbee, 2009) • Texas Department of Family and Protective Services (Dettlaff & Rycraft, 2009)
Responses to T/TA activities—Perceived Utility			
Quality of T/TA provider	Survey of organizational leaders	Single-item measures of trainer effectiveness, knowledge of local systems, and subject matter expertise	• Children's Bureau Training and Technical Assistance (Barbee et al., 2017) • Supporting change in child welfare: An evaluation of training and technical assistance (Children's Bureau, 2015)
	Survey of T/TA recipients	One question measuring T/TA provider satisfaction	• Children's Bureau Training and Technical Assistance (Barbee et al., 2017) • Supporting change in child welfare: An evaluation of training and technical assistance (Children's Bureau, 2015)
Usefulness of training	Survey of organization leaders	One question measuring usefulness of training	• Children's Bureau Training and Technical Assistance (Barbee et al., 2017)
	Survey of T/TA recipients	Perceived appropriateness and overlap between T/TA materials and job	• HRMET Project (Futris et al., 2014)
		Mean score on the Kirkpatrick Level One Training Utility Scale (a widely used measure of usefulness)	• Kentucky Child and Family Services (Antle et al., 2008; Johnson et al., 2009) • Texas Department of Family and Protective Services (Dettlaff & Rycraft, 2009) • North Carolina Center for Public Health Quality (Cornett et al., 2012)
Facilitating factors and barriers to using T/TA	Semi-structured interviews with T/TA recipients	Qualitative analysis of topics that surfaced in the interview	• Two community-based substance abuse coalitions (Chinman et al., 2008)
Anticipated difficulty in implementing lessons learned from T/TA	Survey of T/TA recipients	Mean score on a multi-item scale	• Community-based prevention programs (Hunter et al., 2009)
Responses to T/TA activities—Perceived Difficulty			
Perceived difficulty of T/TA content	Survey of T/TA recipients	Mean score on a multi-item scale that measured preparing for, participating in, and completing assignments related to T/TA	• Community-based prevention programs (Hunter et al., 2009)

Construct or topic	Example Data collection method or source	Example Outcome measurement or analysis	Example project or agency
Measuring Learning from T/TA			
Learning Outcomes—Declarative Knowledge after T/TA			
Knowledge test	Immediate post-test of T/TA recipients	Mean score on a multiple-choice test	• Healthy Marriage Initiative (Antle, Frey, Sar, Barbee, & van Zyl, 2010) • Kentucky Child and Family Services (Antle et al., 2008; Johnson et al., 2009) • Texas Department of Family and Protective Services (Dettlaff & Rycraft, 2009) • Local Safeguarding Children Boards—United Kingdom (Szilassy, Carpenter, Patsios, & Hackett, 2013)
	Delayed test of T/TA recipients (one month post-TA)	Mean score on a multiple-choice test	• Kentucky Child and Family Services (Antle et al., 2008; Johnson et al., 2009).
Attitudes	Survey of T/TA recipients	Change in mean attitudinal score (e.g., racial attitudes after sensitivity training)	• Kentucky Child and Family Services (Johnson et al., 2009) • University-based training of child welfare workers (Jones, Packard, & Nahrstedt, 2002). • Local Safeguarding Children Boards—United Kingdom (Szilassy et al., 2013)
Learning Outcomes—Application of Knowledge			
Application of knowledge	Survey of T/TA recipients	Reported competency in training relevant domains	• Healthy Marriage Initiative (Antle et al., 2010)
			• PSBA Project (Ray et al., 2012)
			• University-based training of child welfare workers (Jones et al., 2002)
		Self-reported comfort and ability in applying skills	• HRMET Project (Futris et al., 2014; Futris, Schramm, Richardson, & Lee, 2015)
Response to a vignette	Immediate post-test of T/TA recipients	Mean scores on measures of different outcomes (e.g. case assessment, intervention, and case worker attitudes)	• Michigan Family Independence Agency (Saunders & Anderson, 2000)

Construct or topic	Example Data collection method or source	Example Outcome measurement or analysis	Example project or agency
Measuring Results of T/TA			
Skill Transfer Outcomes—Application of T/TA in the Workplace			
Frequency of skill use at work	Survey of T/TA recipients	Frequency with which trainees reported using skills	• HRMET Project (Futris et al., 2014)
Described application of skills in the workplace	Survey of T/TA recipients	Content analysis of descriptions of how skills were applied	• HRMET Project (Futris et al., 2014)
	Semi-structured interviews of T/TA recipients	Qualitative analysis of topics surfaced in the interview	• A community-based substance abuse coalition (Chinman et al., 2012) • Texas Department of Family and Protective Services (Dettlaff & Rycraft, 2009)
Achievement of project aims at end of T/TA	Survey of T/TA recipients	Description of reported improvements in processes and outcomes (each T/TA recipient specified a unique goal)	• North Carolina Center for Public Health Quality (Cornett et al., 2012)
Peer performance rating	Survey of T/TA recipient supervisees	Means reported on two subscales of the Supervisory Behavior Description Questionnaire (a widely used measure of performance assessment)	• Kentucky Child and Family Services (Antle et al., 2008)
Third-part performance rating	Observational coding of trainee performance	Percentage correspondence between planned and actual performance	• Youth Development Program—Hong Kong (Law & Shek, 2011)
		Observer ratings of quality of training implementation	• PSBA Project (Ray et al., 2012)
		Observer ratings of whether trainee objectives were achieved	• Middle school drug prevention program (Bishop et al., 2014)
Organizational Change Outcomes			
Changes in staff performance	Survey of organization leaders	Perceived change in staff capacity	• Children's Bureau Training and Technical Assistance (Sun et al., 2017)
	Survey of T/TA recipients	Frequency of sharing T/TA materials with peers	• North Carolina Center for Public Health Quality (Cornett et al., 2012)
Improvement in organizational capacity	Survey of T/TA recipients	Self-reported degree to which T/TA aided the organization and the changes made as a result of T/TA	• California Department of Social Services (2013)

Construct or topic	Example Data collection method or source	Example Outcome measurement or analysis	Example project or agency
	Evaluation by independent evaluators	Ratings of progress on key implementation drivers (e.g., leadership, cultural responsiveness, and stakeholder engagement)	• Children's Bureau Training and Technical Assistance (Armstrong et al., 2014; Sanclimenti & Caceda-Castro, 2017)
	Focus group of local stakeholders	Qualitative analysis of how T/TA contributed to capacity	• Children's Bureau Training and Technical Assistance (Sanclimenti & Caceda-Castro, 2017)
Improvements in measured outcomes	Administrative data	Caseload size, expenditures	• Child welfare (Jones & Biesecker, 1980)
		Client outcomes (e.g., maltreatment reports, face-to-face visits with parents)	• Texas Department of Family and Protective Services (Dettlaff & Rycraft, 2009)
Return on Investment (ROI) Outcomes			
ROI	Administrative data	Cost of T/TA relative to the benefit of reduced service use costs for clients	• Community-based behavioral health providers (Chung et al., 2018)
		Cost of providing T/TA relative to the benefit of reduced worker turnover	• California Department of Children and Family Services (Nguyen, 2012)

REFERENCES

Ackerman, D. J. (2017). Online child care training in the United States: A preliminary investigation of who participates, what is offered, and on which topics the workforce is focusing. International Journal of Child Care and Education Policy, 11. doi:10.1186/s40723-017-0037-7

Aguinis, H. (1995). Statistical power with moderated multiple regression in management research. Journal of Management, 21(6), 1141-1158.

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. Personnel Psychology, 50(2), 341-358.

American Society for Training & Development (ASTD). (2003). Annual report of the industry. Alexandria, VA: ASTD.

Antle, B. F., Barbee, A. P., & van Zyl, M. A. (2008). A comprehensive model for child welfare training evaluation. Children and Youth Services Review, 30, 1063-1080.

Antle, B. F., Frey, S. E., Sar, B. K., Barbee, A. P., & van Zyl, M. A. (2010). Training the child welfare workforce in healthy couple relationships: An examination of attitudes and outcomes. Children and Youth Services Review, 32(2), 223-230.

Armstrong, M. I., McCrae, J. S., Graef, M. I., Richards, T., Lambert, D., Bright, C. L., & Sowell, C. (2014). Development and initial findings of an implementation process measure for child welfare system change. Journal of Public Child Welfare, 8(1), 94-117.

Barbee, A. P., DeWolfe, J., & DeSantis, J. (2017). Key findings of a cross-site evaluation of the Children's Bureau Training and Technical Assistance Network and methodological issues faced. Human Services, 9, 22-33.

Baumgartner, S., Cohen, A., & Meckstroth, A. (2018). Providing TA to local programs and communities: Lessons from a scan of initiatives offering TA to human services programs. Washington, DC: U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation.

Bishop, D. C., Pankratz, M. M., Hansen, W. B., Albritton, J., Albritton, L., & Strack, J. (2014). Measuring fidelity and adaptation: Reliability of an instrument for school-based prevention programs. Evaluation & the Health Professions, 37(2), 231-257.

Cairns, A. H., Gueni, B., Fhima, M., Cairns, A., David, S., & Khelifa, N. (2014). Towards custom-designed professional training contents and curriculums through educational process mining. The Fourth International Conference on Advances in Information Mining and Management, 53-58.

California Department of Social Services (2013). Children and Family Services Plan 2010-2014. Retrieved from http://www.cdss.ca.gov/ccr/res/2013%20APSR%20-%20v%2005.30.2013%20SH.docx

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. Implementation Science, 2(1), 40.

Children's Bureau. (2015). Supporting change in child welfare: An evaluation of training and technical assistance. Washington, DC: Author.

Chinman, M., Hunter, S., & Ebener, P. (2012). Employing continuous quality improvement in community-based substance abuse programs. International Journal of Health Care Quality Assurance, 25, 606-617.

Chinman, M., Hunter, S. B., Ebener, P., Paddock, S. M., Stillman, L., Imm, P., & Wandersman, A. (2008). The getting to outcomes demonstration and evaluation: An illustration of the prevention support system. American Journal of Community Psychology, 41(3-4), 206-224.

Chung, B., Ong, M., Ettner, S. L., McCreary, M., Jones, F., Gilmore, J., & Koegel, P. (2018). 12-month cost outcomes of community engagement versus technical assistance for depression quality improvement: A partnered, cluster randomized, comparative-effectiveness trial. Ethnicity & Disease, 28(Supp), 349-356.

Coetsee, W. J. (1998). An evaluation model for human resources interventions (Unpublished dissertation). Rand Afrikaans University, Johannesburg.

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. Quality of Life Research, 12(3), 229-238.

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. Journal of Applied Psychology, 85(5), 678.

Cornett, A., Thomas, M., Davis, M., Mahanna, E., Cordova, A., Herring, C., & Randolph, G. D. (2012). Early evaluation results from a statewide quality improvement training program for local public health departments in North Carolina. Journal of Public Health Management and Practice, 18, 43-51.

Dettlaff, A. J., & Rycraft, J. R. (2009). Culturally competent systems of care with Latino children and families. Child Welfare, 88(6), 109-126.

Durlak, J. (2013). The importance of quality implementation for research, practice, and policy. ASPE Research Brief. Washington, DC: U.S. Department of Health and Human Services.

Fagan, A. A., Hanson, K., Hawkins, J. D., & Arthur, M. W. (2008). Bridging science to practice: Achieving prevention program implementation fidelity in the Community Youth Development Study. American Journal of Community Psychology, 41(3-4), 235.

Futris, T. G., Schramm, D. G., Lee, T. K., Thurston, W. D., & Barton, A. W. (2014). Training child welfare professionals to support healthy couple relationships: Examining the link to training transfer. Journal of Public Child Welfare, 8(5), 560-583.

Futris, T. G., Schramm, D. G., Richardson, E. W., & Lee, T. K. (2015). The impact of organizational support on the transfer of learning to practice. Children and Youth Services Review, 51, 36-43.

Hatry, H. P. (2006). Performance measurement: Getting results (2nd ed.). Washington, DC: The Urban Institute.

Hunter, S. B., Chinman, M., Ebener, P., Imm, P., Wandersman, A., & Ryan, G. (2009). Technical assistance as a prevention capacity-building tool: A demonstration using the Getting To Outcomes framework. Health Education and Behavior, 36, 810-828.

James Bell Associates & ICF International (2015). Collaboration among T/TA providers. Retrieved from https://www.acf.hhs.gov/sites/default/files/cb/brief_collaboration.pdf

Johnson, L. M., Antle, B. F., & Barbee, A. P. (2009). Addressing disproportionality and disparity in child welfare: Evaluation of an anti-racism training for community service providers. Children and Youth Services Review, 31(6), 688-696.

Jones, L., Packard, T., & Nahrstedt, K. (2002). Evaluation of a training curriculum for inter-agency collaboration. Journal of Community Practice, 10(3), 23-39.

Jones, M. L., & Biesecker, J. L. (1980). Training in permanency planning: Using what is known. Child Welfare, 59, 481-489.

Kahn, L., Hurth, J., Kasprzak, C. M., Diefendorf, M. J., Goode, S. E., & Ringwalt, S. S. (2009). The National Early Childhood Technical Assistance Center model for long-term systems change. Topics in Early Childhood Special Education, 29(1), 24-39.

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. Journal of the American Society for Training and Development, 11, 1–13.

Kirkpatrick, D. L. (1994). Evaluating training programs: the four levels. San Francisco: Berrett-Koehler.

Law, B. M., & Shek, D. T. (2011). Process evaluation of a positive youth development program: Project PATHS. Research on Social Work Practice, 21(5), 539-548.

Merriam, S. B., & Bierema, L. L. (2013). Adult learning: Linking theory and practice. San Francisco: JosseyBass.

Mitchell, R. E., Florin, P., & Stevenson, J. F. (2002). Supporting community-based prevention and health promotion initiatives: Developing effective technical assistance systems. Health Education & Behavior, 29(5), 620-639.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., ... & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. American psychologist, 56(2), 128.

Nguyen, L. H. (2012). Using return on investment to evaluate child welfare training programs. Social Work, 58(1), 75-79.

Peabody, J. W., Luck, J., Jain, S., Bertenthal, D., & Glassman, P. (2004). Assessing the accuracy of administrative data in health information systems. Medical Care, 42, 1066-1072.

Phillips, J. J., & Phillips, P. P. (2008). Distinguishing ROI myths from realities. Performance Improvement, 47(6), 12-17.

Phillips, P. P., & Phillips, J. J. (2004). ROI in the public sector: Myths and realities. Public Personnel Management, 33, 139-149.

Ray, M. L., Wilson, M. M., Wandersman, A., Meyers, D. C., & Katz, J. (2012). Using a training-of-trainers approach and proactive technical assistance to bring evidence based programs to scale: An operationalization of the interactive systems framework's support system. American Journal of Community Psychology, 50(3-4), 415-427.

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. Review of General Psychology, 7(4), 331.

Sanclimenti, J., & Caceda-Castro, L. E. (2017). Children's Bureau training and technical assistance: Synthesis of lessons learned from child welfare implementation projects. Human Services, 9, 34-40.

Saunders, D. G., & Anderson, D. (2000). Evaluation of a domestic violence training for child protection workers and supervisors: Initial results. Children and Youth Services Review, 22(5), 373-395. doi: 10.1016/S0190-7409(00)00086-4

Spoth, R., Claire, C., Greenberg, M., Redmont, C., & Shin, C. (2007). Toward dissemination of evidence-based family interventions: Maintenance of community-based recruitment results and associated factors. Journal of Family Psychology, 21(2), 137-146.

Sun, J., Griffith, J. D., Randhawa, R., & DeSantis, J. (2017). Describing the T/TA services data collected through the T/TA tracking system with implications for child welfare training units. Human Services, 9, 15-22.

Szilassy, E., Carpenter, J., Patsios, D., & Hackett, S. (2013). Outcomes of short course interprofessional training in domestic violence and child protection. Violence Against Women, 19(11), 1370-1383.

Tessler, B., Verma, N., Bigelow, J., Quiroz-Becerra, M. V., Frescoln, K. P., Rohe, W. M., … Miller, E. K. (2017). Scaling Up a Place-Based Employment Program: Highlights from the Jobs Plus Pilot Program Evaluation. New York: MDRC. Retrieved from https://www.mdrc.org/sites/default/files/JobsPlus_InterimRpt-MDRC.pdf.

Texas Department of State Health Services (2016). Maternal and Child Health Services Title V Block Grant, Texas; FY 2017 application/FY 2015 annual report. Retrieved from https://mchb.tvisdata.hrsa.gov/uploadedfiles/StateSubmittedFiles/2017/TX/TX_TitleV_PrintVersion.pdf

U.S. Department of Health and Human Services National Center for Health Statistics. (1996). Third National Health and Nutrition Examination Survey, 1988-1994. Hyattsville, MD: Center for Disease Control and Prevention. Retrieved from NHANES III Laboratory data file (CD-ROM, No. 76200).

Wandersman, A., Imm, P., Chinman, M., & Kaftarian, S. (2000). Getting to outcomes: A results-based approach to accountability. Evaluation and Program Planning, 23, 389-395.

Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. Personnel Psychology, 48, 347-475.

W. K. Kellogg Foundation. (2004). W. K. Kellogg Foundation logic model development guide. Battle Creek, Michigan: W. K. Kellogg Foundation.