

District Guide to Blended Learning Measurement



As I talk to states and school districts across the country about how to measure blended learning's effectiveness, the question I receive most often is: “We’d like to measure, but where do we begin?” Many educators feel ill-equipped to consume, much less conduct, research. By extension, they may feel that all measurement activities are best left to others - even though educators crave the meaningful data and insights that would enable them to make the best instructional decisions for their students. The good news is that measurement need not be intimidating, expensive, or complicated for it to be useful.

The Learning Accelerator created this guide to help educators understand how to approach measurement of their own blended learning initiatives, so that the results are useful to them and their own local needs, while also contributing to a more general body of knowledge about blended learning that can be useful to others.

Researchers are working on one end of this question, conducting different types of studies to answer different questions about blended learning and its effectiveness. In fact, we already have large bodies of evidence that show personalization and learning for mastery, both core components of blended learning, are more effective than other models. For those specifically interested in the current research on blending or the evidence for instructional practices that can be scaled and sustained by blending, [TLA’s Research Clearinghouse 1.0](#) outlines this historical evidence, along with some of the blended learning studies that have already been released, and the implications of these studies for real-life classrooms. However, efficacy evidence is not limited only to researchers and organizations. Educators can and should contribute to the research and knowledge about blended learning’s effectiveness.

This guide is meant to help anyone interested in measuring blended learning implementation with a focus on the teaching and learning that occurs within a blended model. It is best suited for school- or classroom-level measurement. The steps we include can be carried out by principals, district administrators, coaches, and teachers. The guide begins with a general discussion about measurement as it relates to you, and is followed by five steps for developing a measurement plan for your blended learning initiative.

We [welcome your feedback](#) on this guide and any of the included resources. Let us know what works, what doesn’t, what needs further clarification, and what’s missing, so that we can continue to improve our resources to be useful to you.

Forging Ahead With Measurement: A Step-by-Step Guide

Your first step should be to **consider how measurement can help you provide the best supports for your students**. If transformative teaching and learning is a high priority for you, then measurement is too. If it is a high priority for everyone you work with, then engaging in measurement will help all of you achieve that goal.

If you have measurement staff in your district, or access to researchers, evaluators, or others who can support your measurement plan (e.g., district assessment personnel, professional service providers, consultants, technical assistance providers), be sure to include these individuals when developing your measurement plan.

The following five steps will help you create a measurement plan that provides the data and insights you need to support your students. This measurement plan will also be appropriate for your needs and available resources.

Step 1 / Research and Evaluation: Two Sides of the Same Coin

Step 2 / When to Measure

Step 3 / What to Measure

Step 4 / Whom to Measure

Step 5 / How to Measure

STEP 1 / Research and Evaluation: Two Sides of the Same Coin

Measurement work done in schools is often referred to as either research or evaluation. What is the difference between these terms? As you will see below, there are subtle differences between them, but research is not “better” than evaluation, nor do you have to choose between them in order to contribute to and learn from measurement of blended learning.

Measurement activities can be carried out by various individuals. One does not have to be a researcher in order to design and implement a study, nor does one have to be an evaluator (although if educators have access to researchers and/or evaluators, we encourage them to take advantage of these resources).

Measurement can take different forms (such as exploratory research, causal research, formative evaluation, and summative evaluation, just to name a few). In turn, these different forms of measurement activities have different implications for researchers and educators. **What is important is that you select a measurement approach and design that matches the questions you want to answer and the resources you have.** Sharing your results is equally as important for others to be able to learn from your work.

While the objectives and elements of research and evaluation have some key differences, they are both useful and important parts of measurement, and both are necessary to answer the question:

Will blended learning work for my students, teachers, and community?

- Research will answer this question more broadly and with more confidence, but will require more resources, like time.
- Evaluation will answer this question more immediately and specifically for your context, but will not be as generalizable to different settings and individuals.

Table 1 below outlines the similarities and differences between “research” and “evaluation.” This table illustrates their usefulness for your own situation, and their applicability to others who wish to learn from your work.

Table 1. Differences and similarities between research and evaluation

		RESEARCH	EVALUATION
DIFFERENCES	Goal	To uncover and document an existing, underlying “truth”	To understand how specific practices are related to intended outcomes
	Sample	Most often broad, “representative” settings and individuals	Most often specific, local settings and individuals
	Implementation flexibility and setting	Long, carefully controlled, inflexible implementation, often in ideal settings	Short, nimble, flexible implementation in real life settings
	Design	Full range of research designs and rigor, from descriptive case studies through statistical syntheses of multiple studies (meta-analyses)	Broad range of designs, but infrequently includes summaries or syntheses of multiple studies
	Generalizability	Depending on research design and rigor, findings can be assumed to apply in a range of settings and with a variety of individuals	Depending on evaluation design, findings can be assumed to apply to similar settings and with similar individuals
	Publication review	Findings are vetted by a peer review process	Findings are often reviewed internally
	Separation of interests	Researcher, developer, and implementer are often the same entity (or work closely together), while the funder of the research is usually a separate entity	Developer, implementer, and the funder of the evaluation could be the same entity, while the evaluator must be a separate entity for external evaluations
SIMILARITIES	Role of theory	Existing research theory and evidence is used, where available, to determine which practices are most likely to cause intended outcomes	
	Fidelity	Measures of practices and processes (sometimes called fidelity measures), are included ideally both in treatment and comparison groups in order to link specific practices with results	
	Implications for practice	Results can be applicable to others (to different extents, see “generalizability” above) interested in achieving specific outcomes and/or understanding the relationships between specific practices and outcomes	

FURTHER READING:
[Stop \(Just\) Measuring Impact, Start Evaluating](#)

STEP 2 / When to Measure

It is never too early to start thinking about and doing measurement, especially if your measurement approach is one of continuous improvement. In most cases, if measurement work feels like it's happening “too soon” then it is being used as a compliance tool and outcomes or impacts are being measured out of order.

So what's the right order for measuring? The same order in which you expect change to occur. This order is detailed in [TLA's Blended Learning Measurement Framework](#), but can also be documented in any theory of action, theory of change, or strategic plan, which you may already have developed. Table 2 shows a simplified version of our blended measurement framework.

Many times, there are existing measures of the **inputs** that affect your blended learning initiative. By continuing to track these, you can later draw connections between different groups of students and teachers, or different amounts of resources and the outcomes that you are seeing.

In general, your **activities** (the practices and processes that make up your blended learning initiative) are implemented and should be measured first. For example, activities include the type of professional development sessions and other teacher supports you have put in place for your blended learning initiative, or the extent to which personalized learning is occurring in blended classrooms.

After you are satisfied that you are doing things the way you intended, measure the **outputs** of your activities. Outputs are the numerical results of processes, likely the people or things that are directly affected by your practices (e.g., how many classrooms or students experience blended learning?, how many teachers have been trained?). Outputs are not about the quality of what is being done, nor are they about the outcomes or changes you intended to make - they are simply a way of documenting that you are “reaching” or “touching” your intended audience with your practices. Your outputs consist of the individuals and settings through which change can occur.

When you are satisfied that you're doing the activities you intended, and the pipeline for change is in place, then change itself is likely to begin in the form of shorter-term **outcomes**, so these should be measured next. Outcomes of interest might include student engagement, or school climate.

Finally, change will eventually propagate in the form of long-term **impacts**. Impacts include things that take a long time to occur, like teachers' professional engagement, or students' academic proficiency (e.g., graduation rate), or college and career readiness, so these might not be measurable until years after you have begun implementation. (Common student-level impacts are discussed in detail in [Measuring Quality From Inputs to Outcomes: Creating Student Learning Performance Metrics and Quality Assurance for Online Schools](#) from iNACOL.)

Table 2. Summary of TLA’s Blended Learning Measurement Framework

INPUTS	ACTIVITIES	OUTPUT	OUTCOMES	IMPACTS
Students	Integrate in-person instruction and technology	Number of blended classrooms	Student engaged learning	Highly effective teaching and learning in the form of:
Teachers			Teacher scaffolded, individualized learning	
Technology infrastructure			Highly effective instructional actions	
Administrators /Leaders	Personalize learning	Number of students learning in a blended environment	Positive school climate	College and career readiness for all students (academic achievement, emotional well-being, and cognitive skills & behavioral habits of success), and Teacher professional engagement
Other infrastructure	Enable competency-based progression	Number of teachers blending		
	Use real-time data to make instructional decisions			

Your long-term impacts may have been your original goals or objectives for implementation - but remember that change cannot occur in your impacts unless you:

1. understand the context in which you are working (your **inputs**)
2. do something differently (your **activities**)
3. reach those you intended to reach - the individuals and settings in which you would like for change to occur (your **outputs**)
4. achieve the short-term or initial changes you intended (your **outcomes**)
5. combine or sustain these initial changes over a period of time to eventually achieve your goals (your **impacts**)

FURTHER READING:
[Measuring Quality From Inputs to Outcomes: Creating Student Learning Performance Metrics](#)

STEP 3 / What to Measure

In education, our responsibility is to support students' academic achievement, so you may immediately think that's what you should measure. You should measure students' academic outcomes, when the time is right (as discussed above) but academics are not the only skills and habits that lead to college, career, and civic success. In addition, there are teacher outcomes that also support student success that may be affected by your blended learning initiative.

The best way to be sure you're measuring the whole picture is to reflect on your goals or objectives for implementing the program or intervention in the first place. Some questions that might help you think about this include: Why did you select the specific practices and processes that you implemented? What was the specific problem you were trying to solve? Were there certain students or teachers you were intending to help?

Answering these types of questions will help you align your measurement work with all of the teaching and learning metrics you need to include, beyond student academic outcomes. You may not even need to add many measures if you notice that you're already measuring some of the outcomes or impacts that you're interested in. For example, your district is probably already tracking graduation rates, and may even be measuring student engagement or school climate, for all schools across your district.

Measuring your specific activities or your implementation fidelity helps illustrate what is working and what isn't. Blended learning is still new enough that there isn't a clear, single definition of what it is, so it is especially important to measure implementation, not just outcomes. Specific activities - practices and processes - are what make your blended learning initiative different from what was done before, or what others are typically doing. In other words:

- Measure what you are doing, in detail
- Compare what you are doing to:
 - > what you intended to do
 - > what was done before, and/or
 - > what is typically done

This ensures that there are specific practices to which your results can be linked, whether positive or negative. **If there's no tracking of what is being implemented, then there's no way to know if it is or isn't working.**

FURTHER READING:

[CCI/Evergreen Proof Points](#)

(Multiple examples of using measurable student outcomes already collected by districts to determine blended learning success)

STEP 4 / Whom to Measure

This question is intertwined with our “what to measure?” question above, and again, the answer often seems immediately obvious. Of course, you measure those who are participating in your blended learning initiative. Often these are students, but measuring educators (teachers and others), administrators, families, and community members may also align with your original objectives for implementation.

Another group to consider measuring is a comparison group. This group is often overlooked in current blended learning measurement work, but it can provide valuable insights into what is and isn’t working. This non-participating group will help give a sense of what might have happened to your participants if they had not taken part in your initiative. Measuring the practices that are happening in this group is especially important in understanding any connections between what you’re doing, and the outcomes you’re finding.

How this comparison group is selected affects the certainty with which you can know if your initiative is causing the results that you find. Different selection methods have different potentials for bias, which is the likelihood that some other cause (not your initiative) is actually responsible for the results you find. However, minimizing bias often requires more resources, like time and money, so you will have to balance your need for minimizing bias with the resources you have available to you. Regardless of how your comparison group is selected, having one will let you compare outcomes from different situations, and give you clues about relationships between what you’re doing, and any changes you find in outcomes.

Your comparison group may not actually be a different group of participants - the simplest comparison group is often a baseline or pretest measure that can let you know what the outcomes were for your participants before you implemented your program. Other comparison groups could include:

- participants that did not qualify for your initiative, based on a cutoff of some kind
- matched participants that are similar somehow to those who are participating in your blended learning initiative
- the least-biased method for selecting a comparison group: a group that is randomly selected from the same pool of individuals as the participants

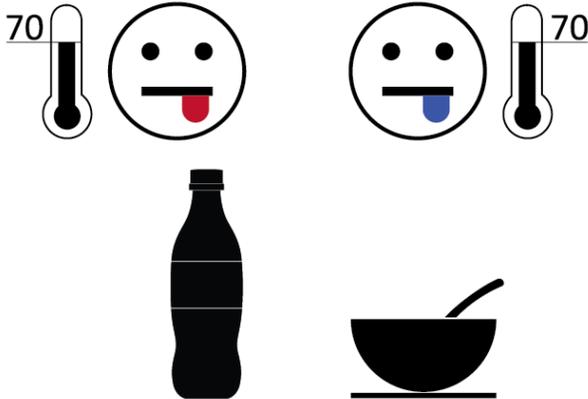
In addition, your comparison group may not even be a group of individuals that you measure yourself. This is related to the “What to measure?” question above, as you can construct **comparison groups from data that are already being gathered within your district or state**. Using groups that are measured by assessment developers, or other groups that have taken the same assessments has the potential to create very biased comparisons, but again, can give you some indication of how your participants might have performed had they not taken part in your blended learning initiative. Some comparisons from these types of assessments might include:

- mean scores from standardized tests (“norm” comparison group)
- percentile-equivalent scores from standardized tests
- expected or predicted growth scores from benchmark tests (“virtual” comparison group)
- proficiency ranks, or proficiency rates, from annual state tests
- other scores from district- or state-wide tests and measures

STEP 5 / How to Measure

Selecting the tools you use to measure all of the things mentioned above is not an easy task. Some of the metrics used might be simple counts or percentages, while others will be scores from rigorously-developed standardized tests. Still others may be self-report surveys. Across these different types of tools, **there are two characteristics that should be considered when deciding which tools to use to measure your blended learning initiative: reliability and validity**.

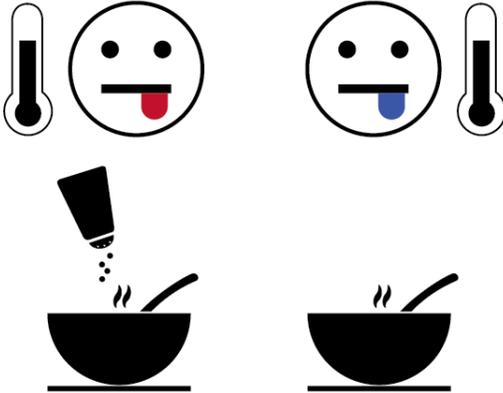
Reliability and validity go hand-in-hand, they are the statistical characteristics of a measure that tell us how consistent and how appropriate the measure is for our use. In less formal terms, **reliability** tells us if a measure is measuring something consistently. This is the likelihood that the measure gives the same score or rating when there is no change in what we’re measuring.



RELIABILITY

For example, a thermometer is a more reliable measure of temperature than your tongue because a thermometer will always give the same reading when something is 70F, while your tongue may tell you that 70F is “cool” or “warm” depending on what is being measured (70F soup may feel cool to you, while 70F soda may feel warm).

Validity tells us if the thing that the measure is measuring is what we think it is. Validity is also dependent on how we intend to use the data from the measure, because how we use the data is related to what we think we're measuring. A ruler, measuring tape, or scale may all be reliable measures of "growth" (assuming that multiple data points are gathered over time) but only the scale, for example, would be a valid measure of growth if the measurement is being done to ensure adequate supports are being provided to a growing object. The measuring tape might be the only valid measure if the goal is to ensure that clothes will fit a growing person.



The diagram consists of four icons arranged in a 2x2 grid. The top row shows a thermometer with a red liquid level next to a neutral face with a red tongue sticking out. The bottom row shows a thermometer with a blue liquid level next to a neutral face with a blue tongue sticking out. Below these are two bowls of food. The left bowl has a spoon and steam rising from it, with a small black icon above it representing a change in temperature. The right bowl also has a spoon and steam, but no change icon above it.

VALIDITY

Back to our example, we can be pretty confident that the thermometer is measuring temperature because the only time the reading changes is when the temperature has changed. We can verify this by comparing to other, known measures of temperature. We can also make sure that the reading does not change when other non-temperature changes occur. Your tongue again is a less valid measure here, because your sensation of "hot" or "cold" might change when the temperature of your food has changed, but it might also change when other things about the food (e.g., spiciness) have changed as well. In fact, our tongue cannot be a valid measure of temperature because it is not a reliable one.

A measure cannot be valid for any use, unless it is reliable - we can't know that we're measuring what we think we are measuring, if a measure isn't measuring anything consistently at all.

Data can only be as good as the measures used to generate them. If a measure is unreliable, and/or invalid, then the data generated are either inaccurate/noisy, misleading, or both. Therefore, selecting measures that are reliable and valid for the ways you wish to use them is necessary for your measurement system to provide you with useful data.

Call to Action

The Learning Accelerator is excited about the ongoing and future measurement work in blended learning. Our hope is that the guide above has demystified measurement and shown how manageable it is. There is a lot to be done, and it is hard work - but the good news is that it is achievable, not overwhelming.

The better news is that there are many resources out there that can help you do this work. The Learning Accelerator will continue to develop and share resources to help guide educators through the work. In addition, there are other organizations and groups - researchers, evaluators, professional services, technical assistance providers, youth voice and survey organizations, to name a few - dedicated to supporting educators in furthering this work. Our students are counting on all of us to roll up our sleeves and dive in.

District Guide to Blended Learning
Measurement was compiled by TLA Partner
Saro Mohammed, August 2015.

For more information, please contact Saro at
saro.mohammed@learningaccelerator.org.

For more information about The Learning
Accelerator, please visit
www.learningaccelerator.org.

