Ed Tech Rapid Cycle Evaluation Coach

Guide: Choosing an Outcome and Measurement Instrument

INTRODUCTION

An outcome is knowledge, skills, attitudes, or other benefits that are attained as a result of an activity (for example, engaging with an educational technology). In other words, the outcome is what you are trying to achieve with an intervention. Depending on the purpose of the technology, your outcome could be students' knowledge of a subject, students' behavior, or a teacher practice such as communication. This is distinct from the measurement instrument, which is the specific assessment, survey, or method that you will use to assess the outcome.

This guide describes different types of outcomes and measures that can be used for rapid cycle evaluation of educational technology. It walks you through the process of deciding what works best for your situation and provides tips on finding measurement instruments for cases in which outcome data are not readily available.

STEPS TO SELECTING AN OUTCOME, MEASURE, AND INSTRUMENT

Exhibit 1 outlines three steps you can take to identify an outcome and source of data for your rapid cycle evaluation. Along with each step are guiding questions to help you think through the process.

Step 1: Identify Your Possible Outcomes

Selecting an outcome that can be quantified can be one of the trickiest steps in an evaluation. Some educational technologies, such as content databases or authoring and collaboration tools, do not produce easily measurable benefits. And information about student progress in the areas you care about may not be readily available within your time frame—or at all—if your district doesn't currently collect this type of data.

Moreover, a growing number of educational technologies claim to benefit students' non-academic traits, such as grit and self-efficacy, but these traits are difficult to assess. In such cases, you may need to administer an assessment or survey specifically for the purpose of your evaluation.

It's important to clearly identify your desired outcomes. But the reality is that deciding on your outcome is often an iterative process informed by your ability to measure the change you want to effect. If you end up using an instrument that is not perfectly aligned with your outcomes of interest, it is worth knowing that up front and documenting it as part of your evaluation results.

Exhibit 1. Identifying an outcome, measure, and instrument

1. Identify your possible outcomes

It is helpful to focus on your potential outcomes before choosing a measure. There may be multiple outcomes that you believe may be affected by the intervention. List them all out and for the moment do not worry about how you will measure them. Keep this list in mind as you move forward in selecting the measure you will use.

Relevant Questions

What do you want to change?

Is the intervention aimed at students, teachers, or other users?

What specific aspect of their skills, knowledge, behavior, attitudes, etc. is the intervention targeting?

Over what time frame do you expect to see changes?

2. Figure out how you can measure your outcome

Sometimes this step is straightforward, such as deciding to use an achievement test when you are targeting student knowledge of a particular topic. However, a growing number of technologies claim to benefit students' non-academic traits, such as grit and self-efficacy, or 21st century skills, such as creativity and critical thinking. These outcomes are more difficult to measure directly. Depending on the characteristics of the outcome you are interested in, some types of measures may be more useful than others. Exhibit 2 below provides suggestions for how to measure different types of outcomes.

Relevant Questions

Does your school or district already measure the outcome?

Can you think of ways to quantify progress in the desired area?

If the outcome can't be measured directly, are there proxies you could look at?

What type of assessment or survey would work in your context?

3. Select your instrument

There are numerous off-the-shelf instruments available (our instrument list can help you find publicly available instruments). Using off-the-shelf instruments can save resources, and they have often been externally validated. However, off-the-shelf instruments were not necessarily designed to measure exactly what you are interested in, nor were they designed with your context in mind. All of these considerations are important when selecting an instrument.

Relevant Questions

When will you need the data to be available?

Is the instrument aligned with your outcome?

Is the instrument valid and reliable?

Is the instrument usable in your context?

Step 2: Decide How to Measure Your Outcome

When evaluating an education intervention, it can be difficult to assess the extent to which students have complex attributes or skills such as motivation, self-esteem, or critical thinking. Sometimes it's not possible to directly measure the qualities we care about.

But there may be indirect ways to get the desired information, such as by asking a combination of questions that together provide insight into an underlying quality. Depending on the characteristics of the outcome you are interested in, some types of measures may be more useful than others.

EXAMPLE: Asking students to rate their own level of engagement is problematic because each student may have a different frame of reference and interpretation of "engagement." Instead, a survey can generate insight into students' level of engagement by asking a combination of questions about whether they are excited to come to class, how often they feel distracted, or whether they think about schoolwork in their free time.

Step 3: Select Your Instrument

Deciding what survey or other assessment to use when measuring complex qualities is a vital part of your evaluation. Choosing an invalid or unreliable instrument can lead to false conclusions.

It's often tempting to develop your own survey or assessment. But this is a major task because questions must be carefully designed to read clearly and elicit the desired information. Additionally, the instrument should be tested to ensure that it works as intended. When possible, it's best to use instruments that have been developed and proven effective by survey design or assessment specialists.

Here are some tips for selecting an instrument:

A. Look for evidence that the instrument measures the skills or attributes you care about (validity).

- ☐ **Ensure that the instrument measures what it is designed to measure and that groups of questions work together well to shed light on the underlying qualities you care about.** Documentation on these topics will often be provided by the instrument developer, in a technical manual, or in related academic articles, and it typically contains statistical checks performed by a specialist.
- ☐ **Ensure that the instrument has been used in similar contexts and for similar purposes.** An instrument that worked well with one type of student or setting may need modifications for

other types of students or settings (such as primary vs. secondary schools or urban vs. rural schools).

- ☐ **Ensure that the instrument has been shown to predict other outcomes that you would expect to be related.** For example, you might expect that a measure of student engagement for 12th graders is positively correlated with attendance rates.
- ☐ **Ensure that the instrument is not overly aligned with the intervention you are evaluating.** Try to use an outcome measure that is not tailored specifically to the intervention. For example, if you are testing the impact of a video-editing platform where some students use the technology and others do not, an assessment that measures students' comfort with that specific video-editing platform may be overly aligned.

NOTE: Surveys can sometimes be adapted to serve your specific purpose, but it's best to avoid pulling out individual questions from a group of questions that address a similar topic. These groups of questions are often designed to shed light on related aspects of a complex quality (such as student engagement) that cannot be measured effectively by individual questions.

B. Look for evidence that the instrument produces stable and consistent results (reliability).

- ☐ **Ensure that the instrument yields similar results if participants take it at different times.** For example, a measure of fairly stable traits, such as motivation, should not change much from one day to the next.
- ☐ **Ensure that different administrators of the assessment obtain similar results.** For example, if you are using a rubric, it's important that the results are consistent regardless of who does the grading.

C. Look for evidence that the instrument is sensitive enough to detect impacts.

- ☐ **Ensure that the instrument has been used to measure changes in the skills or attitudes you care about over a comparable time period.** Outcomes that tend to remain stable over time, such as grit, may require a combination of more sensitive instruments than other outcomes and longer intervention periods.

D. Consider the administrative burden.

- ☐ **Ensure that the level of effort required to administer the instrument is reasonable given the importance of the decisions that will be made based on the results.** Think about how long the instrument takes to administer, whether any training is required to use the instrument, and the cognitive burden the instrument places on participants.
- ☐ **Decide whether to conduct a pre-test.** In general, it's important to have a baseline measure to establish that your treatment and comparison groups were similar before the intervention and to increase the accuracy of your results. If no baseline data are available, you can administer the instrument both before and after the intervention to ensure that the groups were balanced

at the outset. If this isn't feasible, using random assignment can increase your confidence that participants in the treatment and comparison groups were equivalent at baseline.

E. Consider what data collection entails.

- ☐ **Ensure that the instrument can be administered in the same manner for both treatment and control groups.** This means a survey or assessment should be given at a similar time and in a similar setting for both groups.
- ☐ **Decide how to collect the data.** Will the instrument be administered online or in person? How will responses be collected and assembled into a data set?
- ☐ **Decide how to process and analyze the data.** For example, how will responses from a survey be used to create an outcome measure that can be compared across groups or used to inform decisions about implementation? More information about working with survey data can be found in our Using Surveys guide.
- ☐ **Ensure that the instrument allows for the protection of student privacy and confidentiality.** Will you be able to ensure protection of sensitive data? At what point should any student names or other identifiers be removed?

Exhibit 2. Types of measures for consideration

If you want to examine…	Possible outcome measures include...	Examples	Pros	Cons
Longer-term change in a broad range of academic outcomes	Standardized assessments	• PARCC • MAP • NWEA	• Tested for validity and reliability • Widely accepted evidence of learning	• May not be as useful for measuring fine-grained or short-term impacts • Long lag time to obtain data

If you want to examine…	Possible outcome measures include...	Examples	Pros	Cons
Medium-term change in academic outcomes	Diagnostic systems	• iReady • STAR 360	• Tested for validity and reliability • Widely accepted evidence of learning • Frequent administration and web-based reports provide timely information	• Time and cost to administer
Short-term change in academic outcomes or growth in targeted skills/ knowledge	Locally created assessments	• Teacher-developed tests and quizzes	• Aligned with desired learning objective	• May not be accepted as valid and reliable evidence of outcomes outside classroom or school
21st century skills (creativity, critical thinking)	Performance task assessments	• Torrance Tests of Creative Thinking	• Can provide authentic demonstration of skills and knowledge in multiple realms	• Likely to be time and resource intensive • Typically require training to administer and score
Student non-academic outcomes such as social-emotional learning and digital literacy	Externally developed survey instruments	• Patterns of Adaptive Learning Scales (PALS) • NAEP Technology and Engineering Literacy Assessment	• Established validity and reliability	• Not necessarily aligned with desired outcomes • Administration burden may be high • May require specialized skills to administer • May require specialized skills to analyze results
Student behavior	Administrative data	• Attendance • Disciplinary incidents • Suspensions	• Readily available	• Does not capture some types of behavior, such as attentiveness during class • May not be consistently measured or applied across classrooms, schools, or districts

If you want to examine…	Possible outcome measures include...	Examples	Pros	Cons
Teacher attitudes (stress, self-efficacy)	Externally developed survey instruments	• Perceived stress scale • Quality of Work Life (QWL) Assessment	• Established validity and reliability	• Not necessarily aligned with desired outcomes • Administration burden may be high • May require specialized skills to administer or analyze results
Teacher practices	Externally developed observation rubrics	• Framework for Teaching	• Can be grounded in research on effective teaching practices	• Not necessarily aligned with desired outcomes • Administration burden may be high • May require specialized skills to administer or analyze results
Parent engagement	Externally developed survey instruments	• Parent and School Survey (PASS) • Parent Trust in School Scale	• Established validity and reliability	• Not necessarily aligned with desired outcomes • Administration burden may be high • May require specialized skills to administer and analyze results