Evaluation Planning: Models and Matrixes

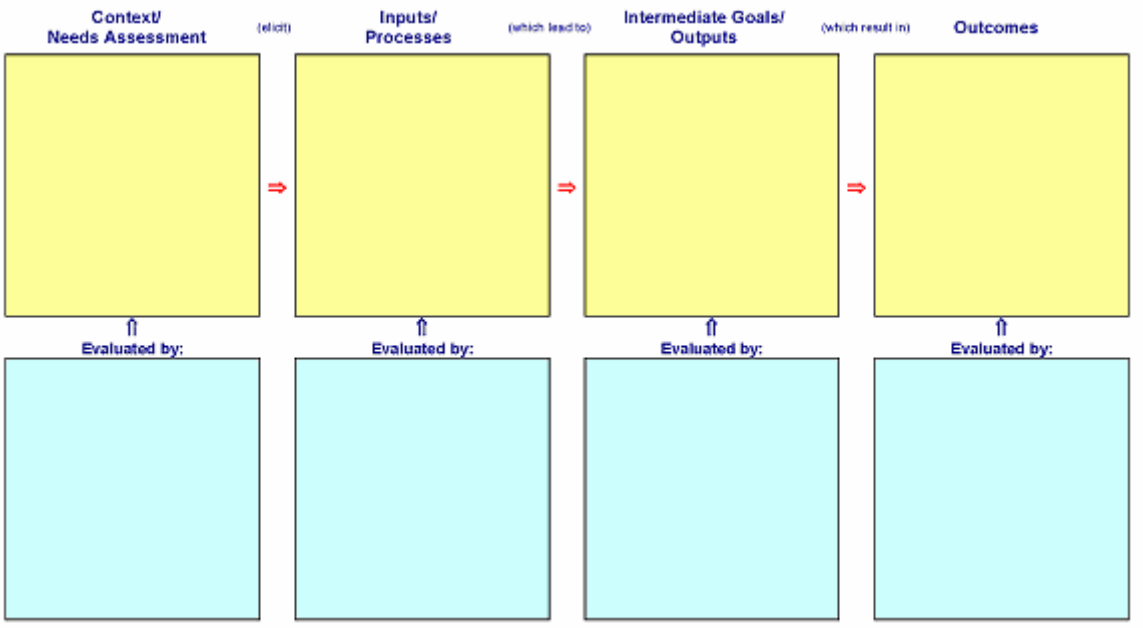
Talbot Bielefeldt (talbot@iste.org)
ISTE Research & Evaluation

In working with schools, universities, education agencies, and foundations on program evaluations, ISTE has found two common tools to be particularly useful in evaluation planning. One is the logic model, a diagram that lays out the rationale for how a program can achieve its goals and the points at which evaluation is necessary. The other is a goals/evaluation matrix, which details the activities that relate to a program's goals and serves as a checklist for the evaluation data that needs to be collected. Used together, these two tools help focus an evaluation to provide useful guidelines for program improvement and compelling evidence for program effectiveness.

If your program approaches ISTE to participate in an evaluation, we will typically ask you to help us complete one or both of these documents for your project:

The Logic Model

This can take a number of forms, but the central idea is to illustrate how proposed activities arise from conditions and lead to desired outcomes. The example below incorporates several frameworks (e.g., Goals/Intermediate Outcomes/Final Outcomes; Input/Output/Outcomes). All of these frameworks ask the project and evaluator to specify what they want to accomplish, how they want to accomplish it, and how they expect to evaluate results.



The Goals/Evaluation Matrix

The matrix starts with the project description and aligns the activities with desired outcomes and the necessary evaluation approaches. It fleshes out the logic model. At the same time, the logic model provides a standard against which to check the specific program and evaluation activities. Are they really moving toward the goals? Will the evaluation data and analysis really tell us if the program has reached the next stage in the model?

Goals/Objectives	Activities	Outcomes	Data	Analysis

The remainder of this short guide discusses some common questions that come up when using these tools

Which comes first?

When in doubt, start with the logic model. The logic model is used any time you need to demonstrate (to yourself or others) that the action you propose to take will actually solve the problems you face. It's not a bad idea to sketch out a logic model even if you are entirely confident in your own mind that the project makes sense. Think of the logic model as a visual elevator speech to explain your program. In verbal form, a speech might go something like this:

> "The problem is, we need more scientists, but a lot of kids drop out of science in their early teens. It turns out most kids don't even really know what scientists do. At the same time, we've got grad students who are eager to share their research, and a school district with connections to the university and good network bandwidth." (**Context/Needs Assessment)**

"We can use mobile computers and web conferencing to bring the students right into the lab. We'll give the teachers training and materials so that they get better at explaining science, too." **(Input/Process)**

"We think we'll see teachers doing more authentic science activities in middle grades, and we think students will have a more accurate picture of scientific research." **(Intermediate Goal)**

"In the end, we think that those students who might enjoy science careers will be more likely to enroll in science electives when they get to high school." **(Outcome)**

Isn't the logic model too simple? The world doesn't work like that.
True. To be accurate, every box would have feedback loops running to previous boxes, as well as incoming and outgoing links to supplementary boxes for unexpected input and discoveries. But those loops quickly make the diagram hard to read.

Our attitude toward logic models is that they are communication tools, intended to convey the fundamental rationale for a program and clarify any gaps in the logic. For illustrating complexity, other tools (such as concept maps, flow charts, Gantt charts, or narrative text) may be more useful.

Do we have to evaluate the stages in order?
It helps. It is hard to know what activities to design if we don't have an accurate needs assessment. Without knowing whether the activities took place as planned, it may not be useful to evaluate the results of those activities. In the example above, if the classroom experience of the students doesn't actually change as an intermediate goal, it doesn't make much sense to attribute any particular changes to the program.

Of course, the ground can shift under us. We may have to go back to reassess the context or the program process if those elements change while we are underway. In addition, there are non-linear projects that don't progress toward a specified outcome. We might hand out some new electronic device to students or teachers, just to see what happens. Goal-free evaluations, ethnographies, and case studies are some evaluation strategies that don't necessarily follow the logic model presented here. However, the programs that ISTE has been asked to evaluate under federal, state, local, and private initiatives generally require a goal-based approach that benefits from using these tools.

What would an "illogical" model look like?
Here's an example from an actual middle school program from the 1990's. The logic model was created after the fact:

Context: Minority middle school girls in a large urban district are not engaged in school. (Evaluated by teacher feedback on attitudes.)	**Input:** A special class provides technology-based projects to at-risk students. (Evaluated by syllabus.)	**Process:** Students participate in class. (Evaluated by class records.)	**Outcome:** Students' academic performance goes up. (Evaluated by school testing.)

The big gap in this plan was between the intermediate goal and the outcome. In a secondary school, where students take many classes from many teachers, it's not reasonable to assume that participation in one technology class will improve test scores, particularly when that class is not focused on the tested content.

The needs assessment cited teacher-perceived attitudes. A logical outcome to explore would have been a change in those attitudes, although the program correctly assumed that this would not be a very compelling result. When teachers in the program were interviewed they said the project was a failure, because test scores didn't go up. However, a requirement for student participation was zero unexcused absences for the year, and only a couple of students failed to meet this standard. This in a school with a high dropout rate. In other words, the program was extraordinarily effective in achieving improvement in attendance. However it had defined itself into failure by aiming at an outcome that was not logically related to the activities.

There are other links in the model that could be debated. What was it about the class that was supposed be motivating? (This program took place at a time when computer programs were criticized for being unfriendly to girls.) Something obviously worked, but the program was not set up to collect information outside its goals. However, the main gap was in the chain leading from attitudes to test scores.

When do we use the matrix?

Generally after we are satisfied that we have a basic logic model that makes sense. That is, we're pretty confident that our inputs will address the context and lead us toward our goals, and we're ready to get into the details of how to actually provide the inputs. Let's take a university example from several years ago. Here's the logic model "elevator speech"

Context:	**Input/Process:**	**Intermediate Goal:**	**Outcome:**
A lot of new teachers drop out of the profession within a few years of getting their credentials, because they don't really know what they were getting into. (Evaluated by the literature on teacher retention.)	Our university puts preservice teachers into K-12 classrooms as soon as they sign onto the program, while they are still undergraduates. (Evaluated by recruitment into the project and availability of field placements.)	Students decide early on whether the field is for them. (Evaluated by surveys and credentialing program enrollments.)	One and two years after credentialing, we will see our graduates in teaching jobs and intending to remain in the field. (Evaluated by surveys.)

In a matrix, this might look like:

Goal: Retain early-career teachers	Activities	Outcomes	Data	Analysis
Objective: Reduce withdrawals from credentialing program.	Establish early preservice field placement opportunities in K12 schools. Provide faculty mentoring for early preservice teachers. Provide faculty mentoring and graduate credit for supervising teachers.	Increased percentage of students completing credentialing program. More positive attitudes toward the program and teaching profession.	Annual student and teacher surveys, student progress in program.	Frequency analysis of completions; content- and statistical analysis of survey comments and scales (pre-program/in-program/out-of-program).
Objective: Reduce the number of graduates leaving the teaching profession.		Increased percentage of students taking up teaching positions after graduation. Increased percentage of graduates remaining in teaching after one and two years.	Annual graduate surveys.	Content and statistical analysis of survey comments and scales (pre-program/in-program/out-of-program).

What do we look for in analyzing gaps in a matrix?
The matrix is organized around goals and activities. The matrix is more fine-grained than the logic model, and more specific to a project. We are looking for program or evaluation activities that might not be feasible. The matrix above, for instance, reveals certain gaps that could make this program difficult to implement.

- Where are all these field placements coming from? Perhaps we need **another objective**: "Establish a working relationship with school districts."
- Who is going to do the field placements and the mentoring? Are the faculty and support staff on board? Maybe we need **other activities**, such as: "Work with the Dean to free up mentoring FTE," or "Hire a field placement coordinator"?
- How will we know if the mentoring for the supervising teachers is effective? How do they provide formative feedback to the program? Do we need **another evaluation instrument**?
- How are we going to keep track of students after graduation so we can survey them about their employment and intentions? Do we need **other stakeholders**, such as the alumni office?
- Will the university require this project to pass human-subjects protection review? Do we need a **process or policy** to ensure that a student isn't jeopardized by sharing doubts about the program?
- The analysis assumes we have good data on pre-program conditions that we can compare to future conditions. Is this wishful thinking or **do data actually exist**?

Do we need to have these exact categories, with unique cells for each goal/objective?
Only if they are useful. As in the example, you might have one survey that assesses more than one activity or objective. And often the data and analysis categories are combined into one column called "evaluation." The important thing is to not bury an important issue by lumping it into a more general category. If our only data collection instrument is a survey, and our goal is student achievement, how exactly are we going to assess achievement with a survey? If our

objective is a statistically significant increase in learning, what kind of comparative analysis can we do with the available data that will give us a meaningful result?

How do we know changes in our outcomes are related to our program?
Attribution is the main technical question in any program evaluation. It has two facets:

1. **Accurately measuring change,** usually by measuring the same variable over time. Some variants include:
    - **Pre/post testing of the same individuals on the same measure** (e.g., math placement on the ACT Compass before and after online math tutorials). This is the most straightforward approach. Questions to ask include: Does the test actually measure what we are interested in? Can we get the same individuals to test pre and post?
    - **Comparing performance in the program to some pre-existing baseline** (e.g., percentage of students enrolling in advanced science courses). Questions to ask include: Is the baseline measure really comparable to the outcome measure? Is the baseline population comparable to the program population? (e.g., were last year's enrollees in advanced courses different from, similar to, or the same as this year's enrollees).

2. **Ruling out other reasons for change**, such as other program influences, individuals moving between groups, and random unmeasured differences that might influence the effects of the program. Common ways to avoid or account for alternative explanations include:
    - **Minimizing differences** between groups by randomly assigning individuals to treatment and control groups. This is the ideal for controlled experiments, but often hard to do in the field. The alternative is to match the groups as closely as possible on existing attributes (e.g., math students who are all freshmen and all permitted to enroll in different sections of the same Algebra class on the basis of testing.)
    - **Accounting for differences** between groups by keeping track of them. This involves noting the effects of related variables, or covariates, such as pretest scores, race, gender, overall academic performance, attitude toward the program, observed engagement in the activities, and so on.

There are tradeoffs in any of these approaches. Strictly-controlled random-assignment studies (besides being difficult to impose) sometimes create an artificial environment that no longer resembles the educational setting. Matching group characteristics and accounting for covariates is always approximate, and as the number of interactions between variables increases, so does the complexity of the mathematics needed to analyze and express the relationships.

In making these tradeoffs, consider common standards for evaluation. Will the data collection and analysis we propose be feasible, accurate, and appropriate? Most of all will it be useful in improving or communicating about the program? The real sign that a project's logic, goals, activities, and evaluation have been thought through is that we are able to implement a program with only a few crises, and the findings are useful for guiding our work in the future.