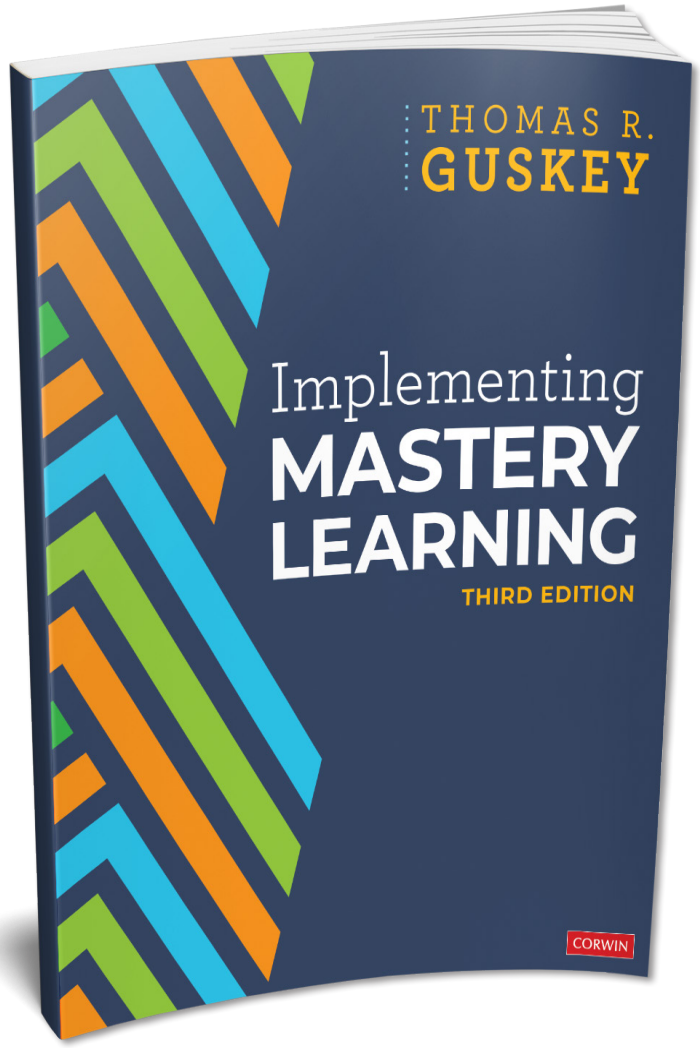

Thank you

FOR YOUR
INTEREST IN
CORWIN

Please enjoy this complimentary excerpt from Implementing Mastery Learning.

LEARN MORE about this title!

CORWIN

CHAPTER 10

EVALUATING MASTERY LEARNING

Having described the major elements of planning for mastery learning and the procedures involved in managing its implementation, we're ready to consider evaluation. To determine if implementing mastery learning has led to meaningful change and positive improvements requires some form of evaluation. We need to verify whether or not intended outcomes were achieved and also if there were positive or negative unintended consequences that resulted from implementation. This chapter focuses on how to gather that evidence and how to use it in answering these important questions.

THE PURPOSES OF EVALUATION

Evaluation in education takes place on many different levels and serves a variety of different purposes. In evaluating students' learning progress, for example, we want to determine the degree of change that has occurred in individual students. But in education, we also evaluate teachers, administrators, schools, curricula, instructional materials, policies, and numerous types of educational programs. The purpose of evaluation at each of these levels is quite different.

In addition to differences in level, evaluations at a single level can serve many different purposes. Teacher evaluations, for instance, can serve "formative" purposes to provide teachers with feedback on ways to improve their interactions with students, classroom management techniques, or instructional strategies. However, teacher evaluations can also serve "summative" purposes to inform decisions about promotion, tenure, teaching assignments, and salary (Darling-Hammond et al., 2012; Millman & Darling-Hammond, 1990).

Our primary focus in evaluating mastery learning is at the classroom and school levels. At these levels we want to gain information about the value and worth of mastery learning in a particular context. In essence, we want to determine (1) if the introduction of mastery learning made a difference, (2) what intended and unintended outcomes occurred, and (3) how our implementation of mastery learning can be improved. To accomplish these purposes, we need to decide (1) what evidence best addresses these issues, (2) how to gather that evidence, and (3) how to analyze that evidence in meaningful and purposeful ways. Assuming the evidence we gather is both reliable and valid, we will then be able to make reasonable judgments about mastery learning's value and worth. We will

also be able to make better decisions about continuation, maintenance, revisions, and further applications.

The evaluation questions we could ask about the implementation of mastery learning are undoubtedly limitless. Among these questions, however, the most important relate to *student learning*. Regardless of the context, mastery learning's primary goal is to improve student learning. For this reason, student learning outcomes are the main focus in evaluating any mastery learning program. They also are the criteria by which most teachers judge their own effectiveness. This is not to imply that other questions about program planning and implementation, teachers' attitudes and perceptions, administrative involvement and support, cost-effectiveness, or return on investment are insignificant. These are vitally important issues that relate directly to program success. Nevertheless, they lie beyond the scope of our discussion here. More detailed information about these other areas of evaluation can be found in *Evaluating Professional Development* (Guskey, 2000).

We define student learning in this context to include a wide range of outcomes. We certainly don't want to limit our evaluation efforts to a single standardized measure of student achievement that may or may not be well aligned with teachers' learning goals. Instead, we need to consider a variety of student achievement measures (Guskey, 2007a). We also may want to consider student affective measures, such as attitudes toward school, particular subjects, or learning in general. We may even want to explore mastery learning's impact on students' confidence in themselves as learners or their sense of self-efficacy. Measures of students' learning also can include school-based measures such as enrollment in advanced classes, attendance, dropout rates, and disciplinary actions.

Much of what we know about the impact of mastery learning on student learning outcomes has come from the results of evaluation studies. And, admittedly, generalizations from these investigations are sometimes limited because most evaluation studies don't employ rigorous experimental designs (S. Anderson, 1994; Guskey & Pigott, 1988). Random assignment of teachers to treatments or students to classes is difficult to achieve in most modern school settings. The majority of mastery learning studies have been conducted in real school contexts and within the constraints of existing classroom arrangements. But while they may not provide irrefutable proof of mastery learning's effects, evaluation studies offer valuable insights into how well mastery learning can work, the conditions under which it is likely to work best, and when adaptations are needed in order to attain the desired results.

GATHERING DATA ON FIDELITY OF IMPLEMENTATION

The first step in evaluating any innovation is to determine the degree to which it was implemented. This requires gathering evidence on both the breadth and

fidelity of implementation. In other words, we must determine who used the innovation, to what extent they used it, and if their implementation included all critical elements.

Innovations like mastery learning are never implemented identically in every classroom or school. To be effective, mastery learning's essential elements must be adapted to a multitude of situational and contextual factors (Elmore, 1997). This requires an appropriate balance be struck between program fidelity and contextual conditions. Researchers refer to this balance as "mutual adaptation" (McLaughlin, 1976; McLaughlin & Marsh, 1978). It means that when implementing innovations like mastery learning, change takes place in two directions. First, individuals must adapt in order to implement the essential elements of mastery learning. But second, those elements also must be adapted to fit the unique characteristics of the context.

Too much change in either direction can mean disaster. If implementing mastery learning requires too much adaptation from teachers and departs significantly from their current practice, then implementation is likely to be mechanical and ineffective. But too much adaptation of mastery learning may result in the loss of elements essential to program impact. Therefore, teachers need to be supported in their efforts to adapt mastery learning to the unique characteristics of their students, classrooms, schools, and communities while maintaining the elements most vital to mastery learning's success (Guskey, 2021).

To determine fidelity of mastery learning implementation requires constructing an *innovation configuration* (Hord et al., 2006). Much like a performance rubric, an innovation configuration describes the key components of a program or practice, along with variations for each component regarding actions that are ideal, acceptable, and unacceptable. It describes the *degree* and *quality* of implementation rather than judging it.

Figure 10.1 shows an innovation configuration for implementing mastery learning. The rows include the critical components for implementation, while the columns list the degree or level to which each component has been implemented in a classroom or course, ranging from "Not Implemented" to "Excellent" implementation. Innovation configurations like this are distributed and explained to teachers during professional learning activities on mastery learning both to guide their implementation procedures and to familiarize them with evaluation expectations.

In some cases, those evaluating mastery learning programs quantify implementation fidelity by assigning numerical values of 0–3 to each quality level for each component. Scores on this particular innovation configuration then could range from 0 for classrooms where there was no evidence of implementing any of the six critical components to 18 for classrooms where all six components were implemented at the "Excellent" level. This fidelity measure can then be used

Figure 10.1 • *Innovation Configuration for Implementing Mastery Learning*

Mastery Learning Implementation (Developed by T. Guskey)				
Critical Components	**Levels of Quality Implementation**			
	Not Implemented	**Partial**	**Satisfactory**	**Excellent**
1. Learning Units	Learning units cannot be recognized	Some learning units are clear	Instruction organized in meaningful learning units	Instruction organized in meaningful learning units related to specific standards
2. Tables of Specification	No tables prepared	Tables prepared but imprecise and unclear	Tables prepared for each unit but the relation to standards is unclear	Tables prepared for each unit and clearly related to learning standards
3. Formative Assessments	No formative assessments prepared	Formative assessments prepared but not parallel or not aligned to tables	Parallel formative assessments prepared and aligned to tables for each unit	Parallel formative assessments prepared, aligned to tables for each unit with 10-20% spiraling items
4. Corrective Activities	No corrective activities prepared	Corrective activities prepared but without different modes of student engagement	Corrective activities prepared with a single different mode of student engagement	Corrective activities prepared with multiple different modes of student engagement
5. Enrichment Activities	No enrichment activities prepared	A single enrichment activity planned for each unit	Multiple enrichment activities prepared for each unit	Multiple rewarding and challenging enrichment activities planned for each unit
6. Documenting Student Progress	Student progress not documented or displayed	Student progress documented but not displayed	Student progress documented and displayed at the classroom level	Student progress documented and displayed at the classroom and student levels

to examine the relationship between implementation fidelity and results attained from the program.

GATHERING DATA ON STUDENT ACHIEVEMENT

Three different types of student achievement data can be collected at different points in time during an instructional sequence. Each point provides different information and can be used to address different questions. The first is evidence gathered *before instruction begins*, referred to as preinstruction or preassessment data. The second is evidence gathered *while instruction is progressing*—that is, formative data. And the third is evidence collected *when the instruction sequence is completed*, or summative data.

Preinstruction Data

Most teachers want to get some idea of the entry-level knowledge and skills of their students before they begin teaching. With this information, they can adjust their teaching to make it more appropriate or revise their instructional format to better accommodate students' individual learning needs. To gain this sort of information, many teachers administer a *preinstruction* or *preassessment* at the beginning of the term or school year.

In Chapter 5, we described the three different forms of preassessments: *prerequisite*, *present*, and *preview*. We discussed how *prerequisite* preassessments address what students need to know and be able to do in order to begin in a particular course or instructional unit. *Present* preassessments gauge students' current knowledge, skills, dispositions, and interests. And *preview* preassessments measure the knowledge, skills, and dispositions that make up the learning goals of forthcoming units and identify for students what they will be learning. We emphasized that these three forms serve different purposes and vary in their effectiveness depending on their content, format, and use.

We further explained that Benjamin Bloom and his colleagues were dubious of the value of formal preassessments, especially in light of noted drawbacks (Bloom et al., 1981a, 1981b, 1981b). Although gathering information on students' entry-level skills through informal procedures can be useful, formal preassessments take significant time to administer, score, and analyze. In addition, there are potentially negative consequences in assessing students on concepts and skills they haven't yet been taught.

But in Chapter 5 we also described the benefits of *prerequisite* preassessments in mastery learning classes when teachers use the results to teach students the prerequisite knowledge and skills to a mastery level. We further emphasized that the results of present and preview preassessments in

mastery learning classes have yet to be thoroughly investigated but are likely to vary depending on how teachers present the assessment to students and how they use the results. In some instances they can guide teachers in planning more effective instruction, but in other instances they simply waste valuable instructional time. Successful implementation requires teachers to take advantage of the potential benefits while avoiding obvious drawbacks, keeping in mind the central purpose of helping *all* students learn well (Guskey, 2018; Guskey & McTighe, 2016).

But in Chapter 5 we also described the benefits of *prerequisite* preassessments in mastery learning classes when teachers use the results to teach students the prerequisite knowledge and skills to a mastery level. We further emphasized that the results of present and preview preassessments in mastery learning classes have yet to be thoroughly investigated but are likely to vary depending on how teachers present the assessment to students and how they use the results. In some instances they can guide teachers in planning more effective instruction, but in other instances they simply waste valuable instructional time. Successful implementation requires teachers to take advantage of the potential benefits while avoiding obvious drawbacks, keeping in mind the central purpose of helping *all* students learn well (Guskey, 2018; Guskey & McTighe, 2016).

Formative Data

Formative assessment results can be very useful in evaluating mastery learning. As we described in Chapter 7, teachers can readily identify teaching successes and dilemmas by constructing assessment response summary charts like the one shown in Figure 7.1. These summaries help teachers determine students' successes and shortcomings. But they also identify teachers' instructional strengths and areas where practices need to be revised to gain greater student success (Guskey, 2007c).

Mastery charts like the one shown in Figure 9.3 of Chapter 9 offer another form of useful feedback and evaluation information. These charts help teachers gauge students' progress across learning units and identify where problems may be evident. For example, having few students do better on formative assessment B is a clear sign of problems. It may be the correctives were ineffective, students didn't fully engage in the correctives, or formative assessment B was simply more difficult than formative assessment A. Similarly, if the number of students attaining mastery on formative assessment A doesn't increase in later units, students may not fully understand the mastery learning process, may fail to see any incentive for doing well on formative assessment A, or perhaps may perceive enrichment as simply more work rather than as a rewarding and exciting activity. Analyzing formative assessment results in these ways can guide teachers in making immediate revisions in their implementation procedures rather than having to wait for summative results at the end.

Summative Data

One of the most important sources of achievement evidence for evaluating mastery learning is summative assessment results. These results represent mastery learning's "bottom line." In most cases, students' scores on summative course assessments provide the principal source of summative data, regardless of their form. They may be projects, reports, exhibits, demonstrations, compositions, or examinations. In some cases, scores on department assessments, end-of-course exams, district- or state-wide assessments, or other standardized achievement assessments such as Advanced Placement exams are also considered for evaluation purposes.

When using data from any of these types of large-scale assessments, however, steps must be taken to ensure the concepts and skills measured are closely aligned with instructional goals and standards of the grade level or course (see Polikoff et al., 2011). If they are not well aligned or if the assessments measure things other than what was taught, they are inappropriate and invalid indicators of the success of an instructional strategy designed to improve student learning (Leinhardt & Seewald, 1981). Such assessments are referred to as "instructionally insensitive" (see Popham, 2007; Popham et al., 2014). Using an "instructionally insensitive" assessment to evaluate the effectiveness of any instructional program makes little sense.

WHAT EVIDENCE DO TEACHERS FIND MOST USEFUL?

In an exploratory, descriptive study of one school district's mastery learning program, we asked teachers about the usefulness of three types of feedback evidence on students' performance to guide instructional improvements (Guskey & Link, 2021, 2021b). The first was formative assessment error analyses, like the one shown in Figure 7.1. These analyses tabulate the number of students in a class who answer each formative assessment item incorrectly or fail to meet a particular criterion. The second was mastery charts of class progress on formative assessments similar to that displayed in Figure 9.3. This chart illustrates the percent of students in a class who reached the mastery standard over a series of formative assessments.

The third source of evidence was summative assessment results comparisons with previously taught classes. Specifically, teachers were asked to compare the average scores students attained on culminating summative assessments under mastery learning with the average scores attained by students on the same assessment in previous years when mastery learning was not used.

Survey results revealed that teachers at all levels consistently rated the tallies of student errors on individual formative assessments as the most meaningful and most useful form of feedback in planning corrective instruction and in making instructional revisions. The detail provided by these item-by-item analyses of formative assessment results provided teachers with highly specific data based

on students' performance. Teachers indicated that with these data they could determine precisely which concepts and skills had been taught and learned well, and which required a different approach. The mastery charts and summative assessment results showed students' performance on a more general basis. The teachers involved in this investigation considered that information more useful in evaluating the overall effectiveness of mastery learning and making changes in implementation procedures.

When asked about ways to improve the quality and utility of feedback from students' formative assessment results, teachers most frequently noted two factors. First was the provision of more time to develop common formative assessments both to improve the quality of their assessments and to make better use of colleagues' expertise in developing instructional alternatives for the corrective process. Second was stronger leadership, especially from building principals, to ensure greater consistency among teachers in establishing mastery-level criteria for the formative assessments. Although teachers at all levels expressed satisfaction with the improvements they saw in their students' performance as a result of implementing mastery learning, many indicated that stronger administrative support and more guidance from school leaders would help them achieve greater consistency in their implementation plans. Several noted that increased time and opportunity for collaboration with teaching colleagues would also assist in their improvement efforts.