Program Evaluation Plan: Assessment Literacy Professional Development Program

Diana Theisinger, M.Ed.

College of William & Mary

Program Evaluation Plan: Assessment Literacy Professional Development Program

In educational contexts, program evaluation is the process of determining the worth and quality of existing school programs in order to gauge which programs merit ongoing support and funding (Sanders & Sullins, 2006). In this program evaluation plan, I have laid out the steps necessary to evaluate a professional development program in a K-12 environment. My plan includes a logic model showing how the program is expected to work, evaluation questions meant to assess whether the program is working in those intended ways, and conclusions related to how this plan fits within existing program evaluation standards.

Context & Criteria of the Program

This program evaluation plan focuses on an assessment literacy professional development program. Faced with declining test scores and concerns over maintaining accreditation, a small school division in Southeastern Virginia hired a team of outside consultants to work with teachers and administrators to improve their collective knowledge of best practices in student assessment. The hope was that improved assessment literacy would lead to improved student performance on classroom and state assessments. The logic model presented in Figure 1 represents the inputs, processes, and outputs associated with this program.

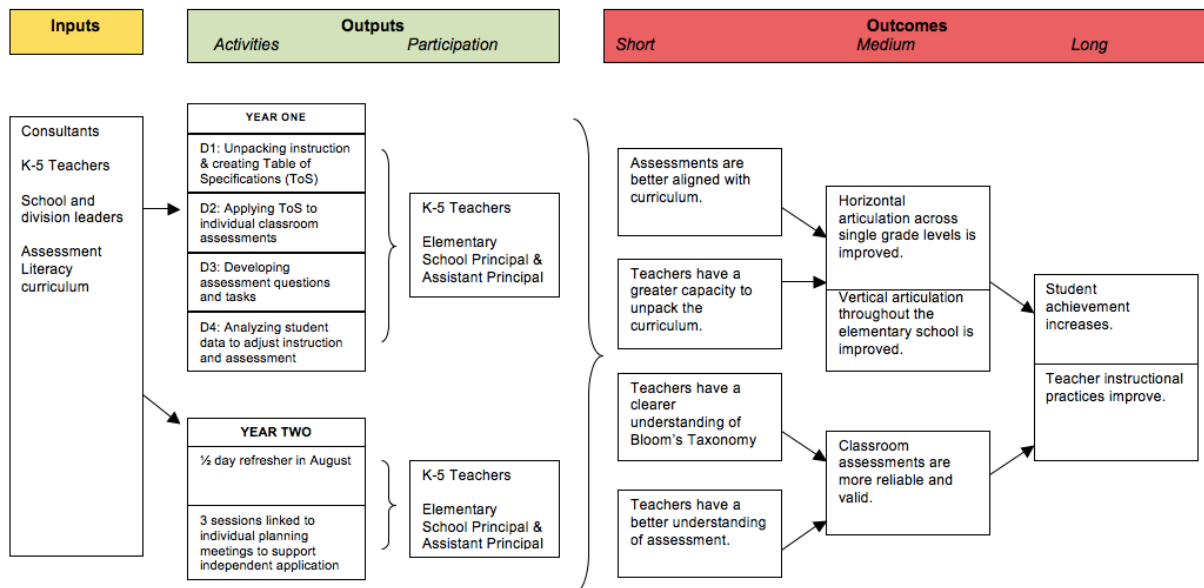Logic Model: Assessment Literacy Professional Development Program

Inputs	Outputs		Outcomes		
	Activities	Participation	Short	Medium	Long

Inputs

Consultants

K-5 Teachers

School and division leaders

Assessment Literacy curriculum

Outputs — Activities / Participation

YEAR ONE

- D1: Unpacking instruction & creating Table of Specifications (ToS)
- D2: Applying ToS to individual classroom assessments
- D3: Developing assessment questions and tasks
- D4: Analyzing student data to adjust instruction and assessment

K-5 Teachers

Elementary School Principal & Assistant Principal

YEAR TWO

- ½ day refresher in August
- 3 sessions linked to individual planning meetings to support independent application

K-5 Teachers

Elementary School Principal & Assistant Principal

Outcomes

Short:
- Assessments are better aligned with curriculum.
- Teachers have a greater capacity to unpack the curriculum.
- Teachers have a clearer understanding of Bloom's Taxonomy
- Teachers have a better understanding of assessment.

Medium:
- Horizontal articulation across single grade levels is improved.
- Vertical articulation throughout the elementary school is improved.
- Classroom assessments are more reliable and valid.

Long:
- Student achievement increases.
- Teacher instructional practices improve.

Figure 1. This logic model displays the key elements of the assessment literacy professional development program. The program is intended to function in the ways depicted in the model, with activities and processes leading to the achievement of short, medium, and long-term outcomes.

Purpose of evaluation. The purpose of evaluating this program is to determine: (a) whether teachers and students have begun to show progress meeting the short, medium, and long-term outcomes of the program; and (b) whether the program should be continued. If outcomes are being met, the program is on track and future professional development may be limited to periodic check-ins or ongoing discussions led by building administrators. If teachers and students have not made progress toward meeting their intended outcomes, it is possible that new or different professional development may be required in subsequent years.

Evaluation questions. This evaluation will focus on the central assumption that if teachers have a better understanding of sound assessment practices, then their classroom assessments will become more reliable and valid, their instructional practices will improve, and

student achievement will improve. These assumptions are embedded in the outcome level of this

program (see Figure 1). The specific evaluation questions are:

      1. To what extent has teachers' understanding of assessment improved?

      2. To what extent are teachers' classroom assessments reliable and valid?

      3. To what extent have teachers changed instructional practices based on what they have

      learned about assessment?

      4. To what extent has student achievement increased?

The purpose of Question 1 is to ensure that teachers have acquired knowledge and skills related

to best practices in assessment. The purpose of Question 2 is to determine whether the

knowledge and skills teachers have acquired are being applied in their classroom practice.

Simply knowing more about assessment without applying that knowledge in the classroom will

have little effect on student achievement. The purpose of Question 3 is to examine whether

teachers' overall instructional practices have been improved through their training in assessment

literacy. Because this question addresses a long-term outcome of the program, it is possible that

an evaluation conducted now may not reveal much evidence of this. However, it is important to

incorporate a plan for evaluating this outcome so that baseline data can be collected and course

corrections, if needed, can be made. Finally, Question 4 relates to the other intended long-term

outcome of the program. To answer this question, the evaluator will seek evidence that teachers'

increased assessment literacy has contributed to improved student achievement.

      Audience. The primary audience for this report is the teachers, administrators, and

outside consultants affiliated with this program. A secondary audience consists of parents, school

board members, and other concerned stakeholders who are invested in this school's success. The

teachers and administrators who have participated in these professional development sessions

have a clear interest in knowing whether their time and effort has been productive in terms of intended program outcomes. Likewise, the consultants who have developed and shared these sessions with the school division will benefit from knowing whether the content they are teaching is making its way into teachers' practices. If the evaluation shows they have been successful in helping teachers meet their intended outcomes, the consultants may be able to use the evidence from this evaluation to recruit future clients. If the evaluation shows they have not been successful, they might consider making changes to the way they deliver this program of professional development.

Evaluator. An outside evaluator will conduct this evaluation. The evaluator is a graduate student from the university with which the consultants are affiliated. She has taken several graduate level courses in both program evaluation and assessment literacy. Although she has some bias related to her university affiliation with the consultants, she is committed to conducting a transparent and objective evaluation of this program.

Collection & Comparison of Data

For this evaluation to be effective, it will need to take place in two phases. The first phase will last 2-3 weeks, and will focus on collecting data related to two of the school's short and medium-term outcomes, as well as baseline data related to one of the school's long-term outcomes. The second phase will be longer, lasting 4-6 weeks, and will not take place until the second semester of the following school year (the 3rd year since program implementation). The evaluator will need the support and assistance of the administrative team as well as classroom teachers who participated in the professional development program. Aside from the evaluator's time, this evaluation will also require access to teachers' classroom assessments and the data

from students' performance on those assessments. Figure 2 provides an overview of the

evaluation information that will be collected throughout this process.

Program Evaluation Plan: Evaluation Questions & Data Sources

Evaluation Question	Importance of Question	Information Needed	When/How Information Will Be Collected	Data Analysis and Interpretation Procedure
1. To what extent has teachers' understanding of assessment improved?	A primary outcome of this program is to improve teachers' knowledge of assessment; it is essential to determine whether this has actually happened.	Some type of assessment of teachers' knowledge about assessment. Teachers' report of what they have learned.	PHASE ONE: During the second semester of Year Two, teachers will complete a brief post-test with items related to assessment best practices. They will also complete a brief narrative reflection related to their own mastery of the assessment literacy content.	The test will be scored and the data analyzed for descriptive statistics. The reflections will be coded and emerging themes will be used to triangulate the quantitative data collected elsewhere during the evaluation.
2. To what extent are teachers' classroom assessments reliable and valid?	It will be important to determine whether teachers' presumed knowledge increases have translated to improved classroom assessments. This speaks to teachers' application of knowledge.	Evidence of teachers' application of best practices for assessment as demonstrated through changes in their teacher-developed classroom assessments.	PHASE ONE: During the first and second semester of Year Two, random audits of classroom assessments (using a rubric) will be conducted.	Classroom assessments will be graded using the rubric and the scores will be analyzed using descriptive statistics. The evaluator will also note the frequency and type of any common errors found on the assessments.
3. To what extent have teachers changed instructional practices based on what they have learned about assessment?	The plan assumes instruction will improve as a result of teachers' increased knowledge about assessment practices. It's important to know to what extent this is actually happening.	Observation of teachers' instructional practices; survey of teachers' self-report instructional practices.	PHASE TWO: Classroom observations before* and after implementation of the program. Survey asking teachers to reflect on how their own practices have/have not changed. *data gleaned from extant observation forms	•Classroom observations will provide qualitative data •The survey will use Likert-scale items and will provide quantitative data •These two forms of data will be triangulated
4. To what extent has student achievement increased?	The second long-term outcome of this program is to improve student achievement. It will be important to assess whether this has happened.	Student testing data.	PHASE TWO: At the conclusion of the 1st, 2nd, and 3rd years of the program, SOL data will be collected and compared to SOL data from the year before the program.	Data for teachers who participated in the program will be analyzed using descriptive statistics. Depending on teacher turnover, these data could be compared with that of teachers who did not participate in the program or who participated for only a short time.

Figure 2. This chart displays the evaluation questions that will guide data collection during the program evaluation process. Each question is displayed alongside a justification for its inclusion in the plan as well as a breakdown of the relevant data sources and data analysis methods that will be used to answer the question.

Phase One data sources. Ideally, a pre-test/post-test model would be used to assess teachers' assessment knowledge. However, the evaluator did not have access to the teachers in the time frame during which a pre-test would have been administered. Instead, the evaluator will have to rely on teachers' self-reported reflections of what they have learned as a result of this program. Self-report data may not be as reliable as other forms of data, since they are subject to factors such as the reporter's mood, self-esteem, and focus on the task (see Sanders & Sullins, 2006, p. 37). Combining information from an objective test of teachers' assessment literacy— quantitative data—with teachers' self-reported reflections—qualitative data—will allow the evaluator to make connections between what teachers think they know and what they can demonstrate on the test. This is a form of triangulation and may allow the evaluator and those who read the final evaluation report to gain insight about any needed changes to the assessment literacy professional development program. For example, if there is a high level of agreement between teachers' performance on the objective test and what teachers report in their reflections, this is a strong indicator that the data are reliable and the school division can make informed decisions based on what is reported. To help the evaluator construct the most in-depth report possible, all teachers who participated in the professional development program will participate in the test and reflection tasks. To ensure that teachers have protected time to complete the brief post-test and reflection, both will be issued electronically during the week following the final professional development session with the outside consultants. The test will be scored electronically; the reflections will be coded and analyzed for themes. For purposes of feasibility, reflections will be limited to 1-2 paragraphs in length.

Next, the audit of classroom assessments will provide another layer of quantitative data related to whether teachers are able to apply what they have learned in practice. Working with

the outside consultants, the evaluator will develop a rubric (see Appendix A) for assessing

classroom assessments according to best practices. Noting the type and frequency of errors made

may also shed light on any areas where teachers are in need of clarification or re-teaching. These

data are likely to be more reliable and valid than teachers' self-reported data in that a single

evaluator will score each assessment using a detailed rubric. Many features of creating sound

assessments are objective measures, such as stating the stem of the question in positive language

(see Waugh & Gronlund, 2013). The process of scoring each teacher assessment according to

these relatively objective measures should be fairly straightforward. However, because this task

is more in-depth and time-intensive than other forms of data collection, only a small, purposive

sampling of classroom assessments will be audited. The evaluator will take care to ensure that all

grade levels, content areas, and ability levels are represented in the assessment audit. Further, the

audit will be conducted once in the fall semester and again in the spring to maximize the

reliability of the information collected.

Phase Two data sources. In Phase Two, classroom observations using an observation

protocol (see Appendix B) will reveal whether teachers' instructional practices have changed as

a result of their participation in the assessment literacy professional development. The evaluator

will analyze these data qualitatively, with comparisons made to relevant notes in existing

observations of the participating teacher (to be provided, with the teacher's permission) from the

teacher or previous observer. Due to the time-intensive nature of these observations, the

evaluator will once again use a small, purposive sampling of teachers representing multiple grade

levels. However, all teachers who participated in the professional development program will be

asked to complete a final reflection, using Likert-scale items to allow for quantitative analysis, to

rate how they perceive their practices have changed based on what they learned. Once again, by

combining teachers' self-reported data with classroom observation data, more robust conclusions may emerge. If what the majority of teachers perceive aligns with what the evaluator observes in the purposive sample of teachers, this is a good indication that the data are reliable. If the data do not align, the evaluator will have to investigate further to determine the source of this misalignment.

The final set of data to be collected and analyzed is student SOL data in Reading and Math. Since data exist for the years preceding the implementation of the assessment literacy professional development program, these can be used as baseline performance data. The evaluator will compare baseline SOL data in Reading and Math to data from Years 1, 2, and 3 of the professional development program. These quantitative measures will be included in the final report and will provide a final source for triangulation of data alongside the other measures that are presented.

Conclusions & Communication

Program evaluation standards. Sanders and Sullins (2006) present four standards for program evaluation attributed to the Joint Committee on Standards for Educational Evaluation: utility standards, feasibility standards, propriety standards, and accuracy standards. Utility standards "are intended to ensure that an evaluation will serve the information needs of intended users" (p. 75). I have aligned my evaluation plan with the utility standards by ensuring that each evaluation question is closely matched with an intended outcome of the program being evaluated. I have also identified the key stakeholders in the program—the school administrators and teachers who participated and the outside consultants who delivered the professional development—as my intended audience. My early identification of stakeholders and audience

further ensures that the data I collect and the conclusions I share will be presented in a way that is relevant to my readers.

The standards related to feasibility "are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal" (Sanders & Sullins, 2006, p. 76). I have made every attempt to limit the scope of data collection to those measures that can be feasibly collected in a realistic time frame by a single, volunteer evaluator. However, this is probably the weakest area of my plan. Teachers are extremely busy, so expecting their cooperation and participation in any task that is not entirely essential (e.g., a reflection on learning related to assessment literacy) may be unrealistic. It is possible that the plan I have developed will result in less participation than I am hoping for.

Propriety standards refer to the legal and ethical aspects of an evaluation, specifically with respect to "regard for the welfare of those involved in the evaluation as well as those affected by its results" (Sanders & Sullins, 2006, p. 76). Since participation in this evaluation is entirely voluntary for teachers, and since the results are unlikely to result in anything more harmful than additional professional learning opportunities, my concerns about the propriety of this plan are minimal. The conclusions of this report will not be used to evaluate individual teachers or the outside consultants. Rather, it will be used to guide future professional learning opportunities for the school division, targeting those areas where teachers may need additional instruction.

Finally, the standards related to accuracy "are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth [or] merit of the program being evaluated" (Sanders & Sullins, 2006, p. 77). The logic model of the assessment literacy professional development program (see Figure 1) guided my development of

the evaluation questions for this plan. The logic model was developed through collaboration with one of the outside consultants responsible for delivering the program. It is reasonable to assume, then, that information collected and shared through the investigation of those evaluation questions will be directly tied to the short, medium, and long-term outcomes of the program, thus ensuring accuracy.

Communication. The results of this program evaluation will be communicated in a variety of ways. First, the full report will be made available to the school's administrators and the outside consultants. The school division may choose to post the evaluation report on the school or division website or to make copies available in a teacher or parent resource center. Next, the evaluator will draft a short executive summary of the primary findings detailed in the report to share with teachers who participated in the program and other interested parties. This will allow those who are interested in the findings, but not the depth of detail provided in the full report, to view a brief version of the main points. Those who are interested in the more thorough analysis can seek out the full report.

Leadership considerations. There are three likely conclusions that might be reached from this program evaluation: (1) The assessment literacy professional development program is effective at helping teachers and the school reach the intended short, medium, and long-term outcomes; (2) The assessment literacy program is ineffective; or (3) The assessment literacy program is somewhat effective and somewhat ineffective. Administrators at the participating school should be prepared for what do in each scenario. For example, if the program is found to be effective, administrators should be prepared to share this news with a broad audience, including parents and the school board. If the program is ineffective, administrators might be less eager to share the news, but they should be prepared to read and understand the evaluation report

and quickly draft a plan to shift how the school addresses faltering student performance. In the

event of an unfavorable evaluation report, administrators should also be prepared to confront

negative attention from parents and the school board about the money spent on an ineffective

professional development program. However, by cooperating with the program evaluation

process, the administrators have already taken a very proactive stance toward ensuring that only

worthwhile programs continue to be funded.

References

Sanders, J. R., & Sullins, C. D. (2006). *Evaluating school programs* (3rd ed.). Thousand Oaks,

   CA: Corwin.

Waugh, C. K., & Gronlund, N. E. (2013). *Assessment of student achievement* (10th ed.). Boston,

   MA: Pearson.

Appendix A
Sample Assessment Evaluation Rubric Items

Item type	1: Unacceptable	2: Needs Improvement	3: Acceptable	4: Exemplary
Multiple Choice: Stem Language	Many item stems (>3) are not clear and concise or are stated in negative language.	A few item stems (1-3) are not clear and concise or are stated in negative language.	Item stems are stated in clear, positive language but could be more concise or could contain more of the wording.	Item stems are stated in clear, concise, positive language; the bulk of the wording is in the stem.
Essay Questions	Essay items are unnecessary or are completely unrelated to intended learning outcomes.	Essay items do not assess complex learning outcomes or are not clearly related to intended learning outcomes.	Essay items assess complex intended learning outcomes but could be more clearly stated.	Essay items assess complex intended learning outcomes and are clearly stated.

Note: developed using guidelines presented in Waugh, C. K., & Gronlund, N. E. (2013). *Assessment of student achievement* (10th ed.). Boston, MA: Pearson.

Appendix B
Sample Observation Protocol Items

The evaluator observed:

___ clear evidence of intended learning outcomes

___ alignment between intended learning outcomes and taught curriculum

___ alignment between taught curriculum and assessment items (either formal or informal)

___alignment between planned instruction (i.e., lesson plans) and delivered instruction