**Review Evaluation Tool**

Overview

The following overview is designed to orient stakeholders to the tool used by the Department of Elementary and Secondary Education (ESE) during the evaluation of educator preparation programs in Massachusetts. This overview is essential in understanding the organization, structure and purpose of this tool. Both the format and construct of content within this tool reflect a newly articulated approach to state reviews and ultimately, ESE’s emphasis on building an evidence-based evaluation system that effectively assesses the quality of educator preparation in Massachusetts. ESE believes that this tool is uniquely tailored to meet the needs of the Commonwealth and is aligned with the vision of effective educator preparation set forth in the [603 CMR 7.00](http://www.doe.mass.edu/lawsregs/603cmr7.html?section=03).

In the overview that follows, stakeholders will be able to learn more about the development, structure and use of the tool in Massachusetts’ preparation program reviews. Sections include:

* [Evaluation Tool Purpose](#_Evaluation_Tool_Purpose)
* [Review Context](#_Review_Context)
* [Development Phases & Key Decisions](#_Development_Phases_and)
* [Organization & Structure](#_Organization_and_Structure)
* [Evidence-Based Decision-Making](#_Evidence-Based_Decision_Making)
* [Future Work](#_Future_Work)

## Evaluation Tool Purpose

Goals

The evaluation tool was designed to:

1. Provide explicit criteria that make clear the expectations set forth in the [Program Approval Standards](http://www.doe.mass.edu/lawsregs/603cmr7.html?section=03);
2. Triangulate data and build an evidentiary base to inform decision-making;
3. Conduct a comprehensive review that values both inputs and outputs;
4. Build a structure that supports (and can articulate rationales for) complex judgments;
5. Target and streamline the submission process.

What it IS/IS NOT

As its name implies, the tool’s primary function is to organize criteria and supporting evidence for reviewers to evaluate. It was, however, designed and intended to be used with a greater sense of nuance than a simple good/bad or yes/no determination. While it may be tempting to compare this tool to a rubric, it is decidedly not a rubric – at least not in the traditional sense of the word.

Here are a few other things this tool is **NOT**:

* A development progression that describes stages of proficiency (aka: rubric).
* A note-taking tool for reviewers.
* Private. ESE values transparency in this process. The tool is shared with the SO upon request and with the Commissioner in considering the approval status recommendation.
* An action plan for improvement. Although it is our strong hope that it is used by an SO to generate one.
* Entirely objective. Inherent in evaluation is human judgment. ESE recognizes this and has intentionally built a system that both values and structures that judgment.

On the other hand, here are a few things that this tool **IS**:

* Iterative. ESE has designed the tool in such a way that additional measures and sources of evidence may be added over time, while the structure is maintained. As a result, ESE issues new Eval Tools for each review year.
* A mechanism for differentiated evaluations. For years, SO’s have asked for models of best practice. Without a system to identify the exemplars in the state with strong supporting evidence, we’ve been left largely with “a sense of things.” This tool will begin to create an evidence base on which we can point to exemplary practices.
* The basis for all judgments and evidence included in the review report. The information contained within a completed evaluation tool directly populates the summary report provided to Sponsoring Organizations. In this way, the report serves as the best reference point for the evidence-base behind all decision making.
* Influenced by evaluation work in the state and nationally. ESE has learned from the many reviews that have been conducted in the state over recent years. Additionally, this tool aligns closely with other evaluation systems in place elsewhere in Massachusetts; including the Educator Evaluation Framework and the district/school accountability system. Finally, ESE has worked closely with national entities (i.e., NCATE, TEAC and now CAEP) and has valued their contributions and input to this tool.

Audience

The evaluation tool has three separate, yet connected, audiences: reviewers, ESE staff, and Sponsoring Organizations.

**Reviewers** are the primary users. They will be the ones who will analyze and collect evidence and evaluate it to determine a summative rating for each criterion. Their summative ratings and pertinent evidence points will be recorded in this tool.

**ESE staff** will use this tool to develop training and toolkit resources; to support reviewers in the examination of evidence; and most importantly, in reaching rating determinations and an approval status recommendation.

**Sponsoring Organizations** are provided this tool as a reference. ESE believes that SOs (and the review process overall) benefit tremendously when there is a shared awareness and understanding of the criteria being used in the evaluation. In addition to valuing transparency, it will be critical for SOs to become familiar with the criteria to inform and further target their submissions and in their drive towards continuous improvement.

## Review Context

The Review Evaluation Tool is one component of a comprehensive review system designed to be effective, efficient and consistent. As the centerpiece of this review system, the evaluation tool is designed to complement and strengthen all other aspects of the review including: the function and effectiveness of reviewers; the preparation of submission materials by the Sponsoring Organization (SO); and ultimately, the recommendation of an approval status made to the Commissioner. For a detailed overview of the review processes please see the [Guidelines for Program Approval](http://www.doe.mass.edu/edprep/ProgramApproval.pdf). The following descriptions identify the key aspects of the review process relative to the evaluation tool.

Function and Effectiveness of Reviewers

Reviewers are responsible for reviewing, analyzing and evaluating evidence of program effectiveness. In tandem with the development of this evaluation tool, ESE has re-envisioned the ways in which reviewers are recruited, selected, and trained. Reviewers are recruited from all education-related affiliates including institutions of higher education, alternative education providers, PK-12 educators, district and LEA central administrators, professional associations, etc. Selection screens for a strong education background, the ability to defer bias in favor of judgments based on evidence and effective communication skills. Reviewers specialize in 1-2 areas of the review and are assigned based on their individual expertise. Trainings include review simulations, calibration activities, and performance checkpoints. The evaluation tool will be the mechanism through which reviewers make judgments about the sufficiency of evidence in support of criteria. Materials relevant to the reviewer recruitment, selection and training a can be found at [www.doe.mass.edu/edprep/reviewers](http://www.doe.mass.edu/edprep/reviewers)

Preparation of Submission Materials

SOs are required to prepare a structured offsite review folio. These documents are reviewed in advance of the onsite visit and provide one source of information for ESE and the reviewers. ESE’s requirements encourage a streamlined submission by asking targeted (rather than open-ended) questions that require responses to be driven by supporting evidence. All submission materials, as outlined in the [Review Toolkit](http://www.doe.mass.edu/edprep/toolkit/), are organized to parallel the structure of this evaluation tool. Organized this way, reviewers have all the information needed to complete a consistent, efficient review and SOs are able to present evidence directly to the criteria on which the programs are being evaluated.

Recommendation of Approval Status

A team of reviewers is convened to evaluate the evidence put forth by a Sponsoring Organization. The ultimate approval decision rests with the Commissioner of Elementary and Secondary Education. This evaluation tool serves to capture summative ratings and supporting evidence. It does not in-and-of-itself yield approval. There are processes and checkpoints built into the [review process](http://prezi.com/9bmmeccwx53s/formal-review/) to ensure that all recommendations regarding approval are grounded in evidence identified over the course of the review. This evaluation tool provides the evidence base for decision-making; it is not the decision-maker. See section below, Evidence-based Decision Making & Recommendations, for additional information.

## Development Phases and Key Decisions

ESE has taken a strategic approach to the development of this tool. Each phase of development has resulted in key decisions that have significantly shaped the evaluation tool as it currently exists.

Experiential Phase

During the 2009/2011 formal reviews ESE and the field piloted the standards and indicators that would later become the 2012 Program Approval Standards. SO representatives and reviewers provided substantial feedback both about the standards themselves and about the process used to evaluate programs. These pilots laid important groundwork for the vision of an evidence-based review. Key decisions in this phase include:

* Choice to review Sponsoring Organizations on a systemic-level. While individual programs are still reviewed and evaluated, much of the formal review and approval rests on the SO’s ability to demonstrate a systematic, organizational approach to preparing educators.
* Provide feedback and evaluative judgments at a smaller (and more action-oriented) unit than the overarching standards.
* Commitment to triangulate sources of evidence so that the review generates a holistic picture of an organization

Initial Development Phase

After the Board of Elementary and Secondary Education passed new program approval standards in 2012, ESE engaged the American Institutes for Research (AIR) to support the development of this tool. AIR provided essential national and international perspective as well as a solid research-base for articulating in greater depth what it meant to be meeting a standard. Key decisions in this phase include:

* Choice to move away from the traditional rubric format and deconstruct the descriptions used to describe “meeting standard” into a set of discrete, concrete criteria against which evidence can be collected and evaluated.
* Reorganize the Program Approval Standards & Indicators into thematic categories called Domains (see Organization section below for full details).
* Focus the rating on the sufficiency of evidence found in support of individual criteria being met.
* Embed output measures as an explicit source of evidence that reviewers will consider along with offsite and onsite evidence in the overall evaluation of the criteria

Feedback/Pilot Phase

In early 2014, ESE released a draft of the evaluation tool and this overview to a representative group of external stakeholders. In addition to a range of representatives from educator preparation organizations, a set of national experts also weighed in on these early versions. Simultaneously, the evaluation tool was piloted by ESE and a team of reviewers for organizations under review in the spring of 2014. Key decisions in this phase include:

* Significant refinement and revision to the criteria for added clarity and assured coverage of the standards and indicators.
* Decision to modify the format of the tool to the structure for increased functionality and clearer links to evidence sources.
* The identification of linkages and implications relative to reviewer training and support.

First Implementation Phase

The evaluation tool was used fully for the first time beginning in the 2014-2015 review year. During this review year, more than 50 reviewers interacted with the tool over the course of evaluating 12 Sponsoring Organizations. ESE collected feedback from providers and reviewers and found that the tool proved to be an effective evidence collection mechanism and a useful device in supporting calibrated judgments. Key decisions in this phase include:

* An affirmation on the basic structure and format of the tool.
* Decision to change the formatting of the tool from a horizontal to a vertical page layout.
* Updates associated with the review criteria and inclusion of additional outcomes as they became available.

## Organization and Structure

Organization

The essential organizing agents of the evaluation tool are the Domains, the thematic categories under which similar criteria are grouped. Basic organization structure is depicted below.



The Program Approval Standards were reorganized into domains and criteria for the purposes of this evaluation tool to ensure that submissions are prepared, reviewed and evaluated efficiently without duplication of efforts and findings. With these domains, criteria that would otherwise fall into two separate standards (i.e., Standard A – Continuous Improvement and Standard G – Program Impact) are now listed (and therefore evaluated) only once.

**Criteria**: The evaluation tool is focused on individual criterion. Each criterion is written as the expectation against which evidence is evaluated. Criteria are derived directly from the [Program Approval Standards](http://www.doe.mass.edu/edprep/ProgramApproval.pdf) and are designed to distill high-level concepts into a set of concrete, actionable criteria. Criteria were intentionally developed to be descriptive of expectations, not prescriptive of approaches or strategies. A full list of criteria can be found at [doe.mass.edu/edprep/pr.html](http://www.doe.mass.edu/edprep/pr.html).

**Domains:** The six domains are listed below. Each domain covers a set of criteria that help to address an essential question.

* The Organization (ORG): *Is the organization set up to support and sustain effective preparation?*
* Partnerships (PAR): *Is the organization meeting the needs of the PK-12 system?*
* Continuous Improvement (CI): *Is the organization engaging in continuous improvement efforts that result in better prepared educators?*
* The Candidate (CAN): *Is the candidate’s experience in the program contributing to effective preparation?*
* Field-Based Experiences (FBE): *Do candidates have the necessary experiences in the field to be ready for the licensure role?*
* Instruction (INS): *Do candidates have the necessary knowledge and skills to be effective?*

All domains, with the exception of the Instruction Domain, are evaluated at the organizational level. Sponsoring Organizations, as the educator preparation provider seeking approval, will submit one set of evidence for each of these domains. In the case of the Instruction Domain, however, SOs will submit one set of evidence per program grouping[[1]](#footnote-1). For example, Madeup University will submit a single set of evidence in support of The Candidate Domain, whereas for the Instruction Domain Madeup University would submit an evidence set for their initial teacher-baccalaureate programs separately from their administrative leadership programs.

Structure

This tool is designed to triangulate evidence from three different sources; the offsite review, the onsite review, and output measures. During a review, the offsite serves to provide initial and foundational evidence; the onsite verifies that what was described in the offsite is indeed happening; and, the output measures confirm that what’s happening onsite is indeed positively impacting the preparation of future educators. The structure of the tool allows all aspects of the review to be present in one document and visually represents how all the pieces fit together.

The diagram below is labeled to orient the reviewer to the tool’s structure.



As indicated in the diagram above, each criteria receives a rating for the level of evidence reviewed both offsite and onsite. The rating system is:

|  |  |  |
| --- | --- | --- |
| Rating | Evidence Label | Evidence Description |
| 4 | Compelling | Irrefutable evidence that criteria is being met consistently; or, *sufficient* evidence that while criteria is being met throughout organization, one or more areas (i.e., programs) presents evidence above and beyond criteria |
| 3 | Sufficient | Clear, convincing evidence demonstrating criteria is being met |
| 2 | Limited | Evidence inconsistently supports criteria; gaps within evidence exist; evidence is weakly linked to criteria |
| 1 | Insufficient | Inadequate evidence was found in support of the criteria |

In this approach to evaluation, reviewers are asked to rate the quality/sufficiency of evidence provided in support of a particular criteria. This is in contrast to other models which require reviewers to pass judgment on the quality of the criteria itself. Rather than asking whether the advising structures are good or not, reviewers ask instead whether there is evidence to substantiate that advising is in fact effective for candidates. This is a nuanced difference, but an important one as it reinforces the shift to outcomes and evidence of impact outlined in the 2012 Program Approval Standards. The burden of evidence is on the Sponsoring Organization. ESE will, however, actively collect evidence for purposes of triangulation throughout the process (particularly during the onsite phase).

## Evidence-Based Decision Making

Triangulating Evidence

Reviewers will examine and evaluate all available evidence offsite, onsite and output measures for each criteria. Output measures are used to corroborate offsite and onsite ratings. Taken together, this triangulation of evidence helps the review team to determine what becomes a finding, a professional suggestion, or a commendation in the final report.

In all cases, evidence of impact will be valued to a greater extent than descriptions of inputs. For instance, if the offsite review of a policy is rated as sufficient but candidate interviews onsite indicate otherwise, reviewers will be advised to weight evidence of the impact (in this case the impact on candidates as evidenced via interviews) more heavily in their overall recommendation for that criteria.

There are instances where all three sources of evidence (offsite, onsite, output measures) are neither available nor included as part of the review. In these cases, information and evidence that is available will be relied upon for decision-making purposes. Regardless, the expectations described in criteria remain unchanged. Three such scenarios are outlined below.

**Formal v. Informal Reviews**

An informal review is conducted entirely offsite through a review of documents. In this instance, all approval status recommendations will be based on the evaluation ratings in the offsite category. If output data is available for similar or connected programs within the Sponsoring Organization it may also be used to inform the process.

**Individual Criteria**

Some criteria lend themselves more effectively to a specific form of evidence collection. Therefore, in order to ensure a streamlined and efficient system, ESE may decide that based on the nature of individual criterion, certain evidence be reviewed only offsite, or only onsite. Where available, output measures will still be used to corroborate ratings. This is noted in the evaluation tool sources of evidence box.

**Output Measures**

ESE is committed to developing and implementing a review process that values impact. ESE and reviewers look closely at all available information and use it to inform the review and decisions. Given that the availability of state-administered and collected output measures are rolling out over a period of years, each review year will include an increasing number of outputs (e.g., evaluation ratings of completers, survey data from stakeholders, etc.). Some criteria are not yet linked to specific output measures and therefore determinations in these cases will rest solely on offsite and onsite ratings.

Decision-Making

Decisions and recommendations occur at several different levels within the process.

**Criteria Ratings**: During the review, an individual reviewer’s summative criteria ratings are challenged and corroborated by the entire review team. The review team, under the guidance of the ESE Ed Prep Specialist, must work towards agreement for each finding and/or commendation cited in the report. ESE reserves the right to change a criterion rating based on an in-depth understanding of regulatory requirements or in order to maintain consistency across reviews. Criteria recommendations result in:

* *Commendation*: Commendations are reserved for truly exceptional, innovative or outstanding practices.
* *Criteria Met:* evidence sufficiently demonstrates that criteria is being met as expected.
* *Finding*: Findings are areas of concern that require action or additional evidence to substantiate that the criteria is being met.

**Domain Recommendations**: Once the review team has rated all criteria in a domain, the team will make an overall recommendation weighing the cumulative impact and significance of the findings and commendations within that domain. Domain recommendations result in one of the following descriptions:

* *Exemplary*: The Exemplary level represents the highest level of performance. It exceeds the already high standard of Proficient. A rating of Exemplary is reserved for performance on a domain that is of such a high level that it could serve as a model for other providers in the organization, state, or nation.
* *Proficient*: Proficient is the expected, rigorous level of performance for SOs. It is a demanding but attainable level of performance for most.
* *Needs Improvement*: SOs whose performance on a domain is rated as Needs Improvement may demonstrate inconsistencies in implementation or weaknesses in a few key areas. They may not yet have fully developed systems to provide preparation in an effective way.
* *Unsatisfactory*: SOs whose performance on a domain is rated as Unsatisfactory is significantly underperforming as compared to the expectations.

**Approval Determinations**: Once all domain recommendations have been determined, the review team again weighs the cumulative impact and significance of all the domain ratings on an organization’s ability to effectively prepare educators and recommends one of the following approval determinations:

* Approved with Distinction
* Approved
* Approved with Conditions
* Probationary Approval
* Not Approved

**Commissioner Determination**: The review team’s recommendations and determinations are assessed by an internal ESE panel to ensure that the proper review protocols were followed and that sufficient evidence exists in support of all judgements. Upon the panel’s agreement, the approval status recommendation goes to the Commissioner for a final determination with a report detailing the evidence collected during the review.

*For full information about the process of review and approval, including more about the potential outcomes associated with each Approval Determination, see the Guidelines for Program Approval.*

## Future Work

With the support of external stakeholders and Sponsoring Organizations, it is our intention that the work done with this tool contribute to a growing knowledge base around educator preparation and undergo ongoing refinement as we learn more about what works and what does not. Additionally, ESE plans to complete the following:

* Refinement of the criteria as the research-base grows
* Inclusion of benchmarks, particularly for output measures, to make expectations even more clear
* The release of exemplars and models of best practice for each domain and strand to better support continuous improvement within organizations
1. ESE does, however, reserve the right to review other domains outside of the Instruction Domain at the program level if SO data and/or evidence warrant further inquiry. [↑](#footnote-ref-1)