Educator Preparation
Stakeholder Surveys
Development & Validation

Overview

As outlined in 603 CMR 7.03 (5), The Massachusetts Department of Elementary and Secondary Education (ESE) is required to administer and publicly report survey data about the quality of educator preparation in the state. The intent of this requirement is that preparation providers, as well as the general public, have access to important information about perceptions of educator readiness in the Commonwealth. ESE has invested significantly in the development of statistically valid instruments as these surveys will be used for various purposes, including programs' continuous improvement, public transparency, and as one source of evidence in program evaluations.

The information that follows details the multi-year development process as well as the validation results for four surveys associated with *teacher* preparation[1]:

1. **Candidate survey**, issued at the point of program completion
2. **Completer survey**, issued to individuals employed in an Massachusetts public school one year after program completion
3. **Supervising Practitioner survey**, issued to individuals who served as a supervisor to a candidate during the practicum experience
4. **Hiring principal survey**, issued one year after program completion to principals who hired a teacher completer

ESE has developed this suite of surveys in order to triangulate perceptions across different perspectives and at varying points in time. For instance, perceptions of readiness for a teacher candidate (survey #1) will be interesting to compare relative to the perceptions of readiness of that same individual the following year as a teacher completer (survey #2). Similarly, examining the hiring principal's (survey #4) perceptions with those of the completer's self-perception (survey #2) may also be illuminating.

Development Process

ESE developed these surveys over the last three years. The overall development timeline is below:
- Fall 2013: Initial item development
- June 2014: Pilot survey of teacher candidates
- Fall 2014: Pilot results released, item refinement and development
- June 2015: Pilot surveys of teacher candidates, completers, supervising practitioners and hiring principals
- Fall 2015: Results released, item selection completed

Throughout these phases of work, several guiding principles remained constant:

- ESE leveraged work done during the development of the Staff & Student Feedback surveys for the Massachusetts Educator Evaluation Framework as a starting point for the Stakeholder Surveys. The same Staff & Student Feedback surveys are also a component of the Candidate Assessment for Performance (CAP). In this way, preparation providers will be able to crosswalk items from both sets of surveys.

- ESE created survey items that fell into one of two categories:

---

[1] ESE prioritized the development and administration of surveys associated with *initial teacher licensure* as this the largest group of program completers in Massachusetts every year (approximately 4,500 completers annually). ESE is also working to develop a set of corresponding surveys that can be used in association with administrative programs. There are no current plans to develop surveys for professional support personnel licenses.

- o Items that align to observable outcomes relative to the Program Approval Standards and Review Criteria, which set forth expectations for providers.

- o Items align to observable practices within the Professional Standards for Teachers (PSTs), which define the pedagogical and other professional knowledge and skills required of all teachers. Particular emphasis was placed on items that correspond to one of the 6 essential elements identified through the Candidate Assessment of Performance.

- ESE developed the surveys in conjunction with one another. Items from the first development and pilot of the teacher candidate survey served as a foundation for all other surveys. The psychometric analyses informed the final form of multiple surveys.

National & State Item Review

ESE consulted other national surveys including the Measures of Effective Teaching Project (MET) and the Danielson Framework survey instruments. A crosswalk was created to map survey items in national surveys to each other and to the Massachusetts survey items. Although item development for the final surveys is original, the crosswalk was used as a resource to identify gaps and areas of alignment.

Content and Stakeholder Engagement

ESE developed survey items for each of the four PSTs and several Review Criteria. Expert review panels (Educators from preparation programs, classroom teachers, and ESE program staff) reviewed these items check for alignment and clarity. Specifically, the panels reviewed items for:

- item representativeness (did they measure the PST it was designed to measure?);
- accessibility (would respondents be able to read the item and understand it?);
- actionability (would programs be able to use the information?); *and*
- responsiveness (would survey items elicit a range of responses?)

The expert panels also helped prioritize items in order to reduce the length of the survey. They prioritized items based on whether they were critical to the work associated with educator preparation, whether there were other ways to get this information that is already available to programs, and the extent to which the item could trigger action within a provider.

Pilot Administrations

In 2015-2016, ESE administered pilot surveys for each stakeholder group. Surveys were sent to everyone that fell under a given category (i.e., the candidate survey was sent to all candidates in the state at the point of their program completion). The only exception was the Supervising Practitioner survey, which was sent only to Supervising Practitioners who work with Sponsoring Organizations that opted in to piloting the survey. Table 1 provides the number of respondents and pilot implementation schedule for the four surveys.

Table 1: Implementation Schedule and Number of Respondents

	Teacher Candidate	Teacher Completer	Supervising Practitioner	Hiring Principal
June 2014	721	-----------	-----------	-----------
February 2015	81	-----------	92	-----------
June 2015	321	208	287	450
Total:	1,123	208	379	450

With the exception of the hiring principal survey, ESE used a traditional Likert scale with five response options to rate respondents' perceptions of the preparation programs. Coding for all items dictated that a response of "0" (*strongly disagree*) would be indicative of the lowest level of agreement with a "4" (*strongly agree)* denoting the highest level of agreement. Response categories scored "1", "2" and "3" corresponded to *disagree, neither agree or disagree,* and *agree*, respectively.

Interestingly, after the first round of pilots in June of 2014, ESE found that the items included in the survey did not vary enough in complexity. ESE recognized the need to add more survey items to assess specifically why survey respondents who were more favorable about their program rated as highly as they did. As a result, ESE added more challenging questions to the survey to help unpack this. Collecting this information will help glean what about these programs are making survey respondents respond in this way.

The hiring principal survey used a 10 point slider scale, from *bottom 50%* (scored 1) to in the *top 1%* (scored 10). This scale allows hiring principals to compare the performance of the new teacher(s) from a given preparation program to all other teachers in their school. The survey is bifurcated such that principals who assess the performance of a recent program completer who *was already teacher of record in their school* can evaluate the performance growth of that teacher after they complete their preparation program. Principals who assess the performance of a recent program completer who *was not already a teacher of record in the school* evaluate the completer's performance relative to the other teachers in the school. ESE found in pilot administrations that respondents did not use all the response categories. As a result, in the future, the hiring principal survey will use a five point rating scale to increase comparability to the other stakeholder surveys.

Survey Reliability and Construct Validity

ESE analyzed surveys using a rating scale Rasch model (Rasch, 1960; Andrich, 1978). Wolfe and Smith's (2007a, 2007b) construct validity framework for Rasch-based survey development was used to guide psychometric analyses. When possible, ESE administered two pilots to respondent groups; the second pilot helped to assess whether the new items improved the psychometric properties of the surveys. Content, substantive, structural and generalizability validity evidence was assessed to determine the final forms of the surveys. However, conducting external and consequential validity analyses was beyond the scope of the survey development project. Table 2 provides some descriptive psychometric data for the final forms of each survey.

Table 2: Descriptive Data and Person Separation Reliabilities (PSR)

	Teacher Candidate Long Form	Teacher Completer New Teachers[1]	Teacher Completer Teacher of Record[1]	Supervising Practitioner[2]	Hiring Principal New Teacher[3]	Hiring Principal Teacher of Record[3]
Number of Respondents[4]	394	100	116	352	255	154
Number of Items	42	27	27	41	6	6
Mean	1.92	1.55	1.56	2.36	2.92	5.36
Standard Deviation	1.52	1.29	1.81	1.73	3.15	3.86
Variance Explained	49.9%	41.1%	60.7%	53.1%	86.6%	83.3%
Real (Model) PSR	0.93 (0.94)	0.92 (0.93)	0.94 (0.95)	0.93 (0.94)	0.95 (0.96)	0.95 (0.96)
Responsiveness	5 person strata	5 person strata	6 person strata	5-6 person strata	6-7 person strata	6-7 person strata

[1]ESE performed only one pilot for the teacher completer respondent group. As a result, there were only 27 items that were in common with the final forms of the teacher candidate and supervising practitioner surveys. The teacher completer survey was analyzed separately depending if the teacher completer was a new teacher during their time in the teacher preparation program or a teacher of record.

[2]One item was not tested in the supervising practitioner pilot surveys but will be used in the final form.

[3]A 10 point rating scale was tested out in the HP survey. The HP survey was analyzed separately depending if the candidate was a new teacher to the school or a teacher of record.

[4]Number of respondents is based on those who did not exclusively pick only the lowest or highest level of agreement.

Overall, ESE is confident that the development and validation work of these stakeholder surveys will produce valid data that will aid Sponsoring Organizations' continuous improvement efforts. Despite this, it is important to note that the number of respondents used to assess the reliability and validity of the surveys was small for each respondent group; as a result, ESE will continue to assess the reliability and validity of the surveys using data from future administrations. This will also include a differential item functioning analyses to assess the extent to which different respondents differ in their responses to survey items. This could not be conducted with the pilot data due to low sample sizes. ESE will use data from the 2016 survey administration to assess again whether the items of each survey are valid and that the instruments overall are reliable and generalizable.

Additional information about the surveys as well as all current versions of the surveys can be found at:
http://www.doe.mass.edu/edprep/surveys/.

References

Andrich, D. (1978). Rating formulation for ordered response categories. *Psychometrika*, *43* (4), 561-573.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research. (Expanded edition, 1980. Chicago: University of Chicago Press).

Wolfe, E. W., & Smith, E. V. Jr. (2007a). Instrument development tools and activities for measure validation using Rasch models: Part I – Instrument development tools. *Journal of Applied Measurement*, *8*(1), 97-123.

Wolfe, E. W. & Smith Jr., E. V. (2007b). Instrument development tools and activities for measure validation using Rasch models: Part II – Validation activities. *Journal of Applied Measurement*, *8* (2), 204-234.