Appendix D. ESE Pilot Survey Project Summary

Purpose

Beginning in the 2014–2015 school year, districts will start collecting feedback from students and staff for use in educator evaluation. ESE is charged with recommending and supporting a feasible, sustainable way for districts to collect and report back feedback to educators in a manner that can inform and improve educator effectiveness. To that end, ESE developed model student and staff feedback surveys for optional use[1] by MA districts. The ESE model surveys are designed in accordance with the following criteria:

- Items are aligned to observable practices within the [Standards and Indicators of Effective Teaching Practice](#) (student surveys) and [Effective Administrative Leadership Practice](#) (staff survey).

- Survey information provides educators with actionable information to improve their practice.

- Items are developmentally appropriate and accessible to students across grades levels.

Pilot Sample & Implementation Schedule

Eight districts and one collaborative volunteered to participate in the ESE Model Survey Pilot Project: six districts and one collaborative piloted student surveys (n=19,255 across Pilots 1 and 2) and seven districts piloted the staff survey (n=1,559 across Pilots 1 and 2). Data indicate that samples were reasonably representative of the student and staff population in Massachusetts at each grade span.

Pilot Sample:

- Auburn Public Schools
- Boston Public Schools
- Greater Lawrence Technical School
- Lincoln Public Schools
- Malden Public Schools
- Norwell Public Schools
- Quaboag Regional SD
- South Coast Education Collaborative
- Westport Public Schools

Implementation Schedule:

- Pilot Administration #1: February 3rd – 14th

- Item Analysis & Refinement: February/March

- Pilot Administration #2: March 31st – April 11th

- Item Analysis & Refinement: April—June

---

[1] Districts are not required to use ESE model feedback surveys.

Stakeholder Engagement

Educator input has been a driving force throughout ESE's development of the educator evaluation framework, the [ESE Model System](), and, most recently, the ESE Model Student and Staff Surveys. Students and educators were involved from the early planning stages, to the development and implementation of pilot items, to the final item selection and publication of the model surveys (see next section for survey development details). Districts of varying sizes, locations, and student populations were recruited to participate in the pilot project and contribute their unique perspectives and recommendations to the final survey instruments.

ESE is indebted to the 10,000 students and 1,500 staff who piloted survey items during the 2013-14 school year, and to the more than 2,200 students, parents, teachers, and school and district administrators who provided input along the way.

Planning

Student Advisory Councils. The State Student Advisory Council, five Regional Student Advisory Councils, and the Boston Student Advisory Council provided feedback during the 2012-13 school year on the value of student feedback, how they envisioned it being used in educator evaluation, and the most important Standards and Indicators of Effective Teaching Practice that are observable by students.

Presentations. In 2013, ESE presented to 100 educators at the Educator Evaluation Spring Convening. Additionally, ESE solicited input from educators at a variety of state organization conferences to identify and prioritize content aligned to the Standards and Indicators of Effective Teaching and Administrative Practice, including the Massachusetts Association of School Superintendents Summer Conference and Executive Committee, and the Massachusetts Administrators of Special Education.

Statewide Survey. Distributed through the August, September, and October 2013 Educator Evaluation Newsletters, 150 educators provided input through an online survey on which Standards, Indicators and elements of the teacher and administrator rubrics should be reflected in feedback surveys.

District Consultations and Observations. ESE's survey developer met with local superintendents and observed K-2 classrooms to help with the development of items for younger grades.

Piloting

Expert Review Sessions. Teams of educators, including K-12 teachers, school administrators, and district leaders were convened seven times during the 2013–14 school year to review and provide feedback on pilot survey items and survey reports.

Pilot District Educator Feedback. Educators in eight pilot districts and one collaborative provided feedback on survey forms, individual items, and overall guidance through ESE site visits (75 educators attended) and post-administration feedback forms (over 400 responses).

Cognitive Interviews & Focus Groups. ESE completed cognitive interviews and focus groups with 52 students across grades K-12 to ensure that students understood the meaning and purpose of survey items for their grade levels.

Student Advisory Councils. Students continued to provide input during the 2013-14 school year through the State Student Advisory Council and Boston Student Advisory Council.

Publishing

ESE's Teacher Advisory Cabinet & Principal Advisory Cabinet. These two advisory cabinets, comprising 24 teachers and 21 school administrators, provided feedback on ESE's guidance document and suggestions for alternate methods for collecting and incorporating feedback in evaluation.

Expert Review Webinars. K-12 teachers, school administrators, and district leaders were convened for three interactive webinars to provide feedback on ESE's guidance document and provide suggestions for alternate feedback instruments.

Online Item Feedback Survey. In May and June 2014, over 500 educators and 40 students evaluated individual items in an online item feedback survey to identify items that were most useful and important to them for improving educator practice.

Educator Evaluation Spring Convening. In May 2014, educators and ESE team members presented to almost 1,000 educators about the value of student feedback in evaluation, districts' flexibilities for collecting and using student feedback, ESE's pilot administrations, and implementation resources to be published in July 2014.

Final Selection Committee. ESE staff and Massachusetts educators were convened in June 2014 to identify the final items for inclusion in the ESE model student surveys for Grades 3-5 and 6-12, as well as the ESE model staff survey for school administrators.

Presentation to the Board of Elementary and Secondary Education. In June 2014, two superintendents from pilot districts presented with Educator Evaluation team members on their positive involvement in the survey development process and plans to use ESE's model survey instruments.

Ongoing Engagement

During the 2014-15 school year, ESE will be partnering with districts to implement and refine the model feedback instruments. We look forward to continuing the dialogue with educators to ensure that the model surveys provide accurate and actionable feedback to inform educator practice.

Survey Development

Pilot student surveys about classroom teacher practice were developed for students in three grade spans: grades K-2, grades 3-5, and grades 6-12. Student items were based on observable teacher practices related to *Curriculum, Planning & Assessment* (Standard I) and *Teaching All Students* (Standard II). Items were first developed for the G6-G12 survey; K-G2 and G3-G5 items were derived from G6-G12 items.

The pilot staff survey was developed for school staff to provide feedback to school leaders on observable administrative leadership practices related to *Instructional Leadership* (Standard I), *Management & Operations* (Standard II), and *Professional Culture* (Standard VI). Publicly available items were considered for inclusion in both surveys.

Through two pilot administrations and extensive stakeholder engagement (see previous section for details), the ESE model instruments published in Appendix A were developed. Stakeholder engagement focused on assessing survey items for their Content Validity, Utility and Accessibility. Reliability and validity analyses were performed to support the development of the model instruments. The survey development activities are briefly summarized here.

Response Options

Each pilot survey used a Likert-scale to measure student and staff perceptions of teacher and administrator practice respectively. In Pilot 1, two different response option conditions were tested in the G3-G5, G6-G12 administrations, and in the staff administration. Psychometric analyses revealed that student responses were more reliable when using a 4-point rating scale of "strongly disagree" to "strongly agree" (as opposed to "in no lessons" to "in nearly all lessons"). For the staff survey, a five option response rating scale proved more reliable than a six response option response rating scale; both rating scales were on a "strongly disagree" to "strongly agree" continuum.

Item Development

A Rasch model was used to place ordinal Likert responses on to an interval-level, linear scale. In Pilot 1, to maximize the number of items tested, four forms of 45 and 60 items were administered to G3-G5 and G6-G12 students respectively; three forms of 80 items were administered to teachers. Each form was representative of the Massachusetts Standards and Indicators of effective teaching and administrative leadership practice, with each form designed to represent a "model" survey. Using classical methodology, the reliability (Cronbach Alpha) of each form from the Pilot 1 surveys is shown in Table 1 (reliabilities are broken down by Standard). The data indicate that across all survey forms there is high degree of internal consistency, and, similarly, across Standards and Indicators within each form. Cronbach alphas ranged from 0.78 (Form D, G3-G5) to 0.94 (Form B, G6-G12) for Standard I, with the reliability of Standard II items ranging from 0.89 (Form B, G3-G5) to 0.95 (all forms, G6-G12). Reliabilities across staff forms were all above 0.94 for each Standard measured.

Table 1. Reliability: Consistency Across Forms in Pilot 1

	G3 – G5 (4 Forms: N=1,225)		G6 – G12 (4 Forms: N=2,916)		Staff (3 Forms: N=554)	
	# Items	α	# Items	α	# Items	α
Std. I	12-14	0.78-0.85	20-21	0.91-0.94	32	0.97
Std. II	30-33	0.89-0.92	35-40	0.92-0.95	22-24	0.95-0.96
Std. IV	NA	NA	NA	NA	24-26	0.97

Instrument Development

Instrument development was primarily conducted using a Rasch model. Wolfe and Smith's (2007a, 2007b)[2] construct validity framework for Rasch-based survey development was used to guide psychometric analyses. A subset of common items (present on each form) that represented the breadth of the teacher/administrator effectiveness constructs was used to anchor items from each form onto the same metric, enabling the psychometric properties of all items to be examined together (e.g., 16 common items anchored all 192 G6-G12 Pilot 1 items tested using the four forms). Content, substantive, structural and generalizability validity evidence was assessed to determine which items were psychometrically sound and supported the validity evidence provided by our stakeholder engagement activities. For example, Differential Item Functioning analyses were performed to ensure that items were interpreted similarly across subgroups (e.g., gender, low income etc.).

Using the feedback from stakeholders, combined with the psychometric analyses, items were developed for Pilot 2, with the goal of further improving the content representativeness and psychometric properties of each of the surveys. In order to ensure items from Pilot 2 could be placed on the same metric as items from Pilot 1, items common to Pilot 1 were embedded within Pilot 2's survey forms. This helped to ensure the identification of new items that improved the psychometric properties of the surveys. The reliabilities of each form used in Pilot 2 are provided in Table 2. Cronbach alphas ranged from 0.82 (Form B, G3-G5) to 0.92 (Form A, G6-G12) for Standard I, with the reliability of Standard II items ranging from 0.90 (Form A, G3-G5) to 0.97 (Form B, G6-G12). Reliabilities across staff forms ranged from 0.81 to 0.97 across all Standards measured.

Table 2. Reliability: Consistency Across Forms in Pilot 2

	G3 – G5 (2 Forms: N=2,265)		G6 – G12 (2 Forms: N=3,694)		Staff (2 Forms: N=462)	
	# Items	α	# Items	α	# Items	α
Std. I	12	0.82-0.84	16-17	0.91-0.92	24-26	0.90-0.93
Std. II	24	0.90-0.91	33-34	0.96-0.97	12-13	0.81-0.83
Std. IV	NA	NA	NA	NA	22-23	0.94-0.95

Final Forms

In conjunction with psychometric data from both pilots, items for the final model instruments were chosen by expert review panels; the instruments are presented in Appendix A.

Survey	Long Form	Short Form
Student Survey, Grades 3-5	45 items	25 items
Student Survey, Grades 6-12	55 items	25 items
Staff Survey	65 items	35 items

---

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