Department of Research and Evaluation
AUSTIN INDEPENDENT SCHOOL DISTRICT

Student Course Feedback Pilot, 2011–2012

Administration and Instrumentation

In Spring 2012, Austin Independent School District (AISD) staff in the Department of Research and Evaluation and the Office of Educator Quality developed a student survey to obtain students' feedback about their classroom experiences. The instrument was designed to measure the instructional domains AISD educators identified as indicators of high-quality teaching for a new observational rubric developed for use by teachers who served as peer observers at 27 AISD schools in 2011–2012 and by administrators at three schools participating in a new pilot appraisal system. To obtain multiple observational measures of teaching, the student survey was created using a combination of existing items from the Tripod instrument[1] and new items. Students were asked one or two items about each of 10 domains in the observational rubric (see Table 1). The rubric included 13 domains, 10 of which students could assess.[2]

Students from three AISD schools (n = 1,531) answered the survey about their classroom experiences with specific teachers. Researchers documented the challenges and procedural issues associated with the survey instrument and its administration, analyzed results of the survey, and examined properties of the instrument. Results are described for three groups of students: kindergarten through Grade 2, Grades 3 through 5, and Grades 6 through 12.

OVERVIEW OF THE PILOT ADMINISTRATION

Kindergarten through Grade 2. The instrument for primary grades included only 11 items, a subset of those asked of older students. The response scale included faces displaying various degrees of happiness, and instructions asked students to circle the face best representing how much each statement was true for their teacher's class. Students responded about their homeroom teacher and a randomly selected special area teachers (i.e., art, music, or physical education). Two researchers facilitated the process without the presence of teachers. The researchers read instructions and items aloud and answered any questions about the meaning of items. With help, all students were able to complete the surveys. Consensus among researchers was that kindergarten students can take a survey like this in a group with assistance. A team of three (i.e., one person to read the instructions and two to assist) seemed to be the right balance. However, asking kindergarten students for feedback about two different teachers was too much. Students seemed confused when asked to complete the survey about their special area teachers, possibly because they were not in the appropriate classroom, and the abstract task of thinking about a different teacher was difficult for many. For this reason, researchers determined that the responses from kindergarten students about special area teachers were not valid. Additionally, students needed about 45 minutes to respond about two teachers, which was too long. Students could respond about their homeroom teacher in less than 30 minutes.

[1] http://tripodproject.wpengine.com/wp-content/uploads/2012/03/Flyer-Tripod.pdf

[2] For more information about the observational rubric or peer observation program, see http://archive.austinisd.org/inside/initiatives/compensation/index.phtml

DRE publication #11.76 RB
October 2012

Lisa N. T. Schmitt, Ph.D.

Many 1st-grade students could read along with the survey administrator, and administration was easier than for kindergarten students. However, as with kindergarten students, asking 1st-grade students about special area teachers created confusion and caused the session to last too long. Administration was shorter with 1st-grade students than with kindergarten students, lasting approximately 25 to 35 minutes. Second-grade students seemed to have no problem taking two surveys during a time period of 20 to 25 minutes. However, the time period before lunch was less than optimal for their full attention to the process.

The items generally were easy for students to comprehend. However, one item, "this teacher does not allow students to break the rules," was problematic for students in all grades due to the response scale. After recognizing the problem during the first administration, researchers in several classrooms changed the wording to "This teacher *allows* students to break the rules." Changing the statement from negative to positive seemed to improve comprehension. However, results for this statement were omitted. Items on the kindergarten through Grade 2 survey instrument are noted with an asterisk in Table 1.

Table 1. Observational Domains and Corresponding Survey Items

Observational domain	Survey item
1. Actively engages students during instructional activities	This teacher asks us to lead lessons. This teacher asks us to share our thoughts.*
2. Checks for student understanding and modifies instruction to address student misconceptions	This teacher answers my questions when I don't understand something.* This teacher has several good ways to explain each topic so I understand.
3. Differentiates instruction for student needs, utilizing a variety of instructional strategies	This teacher uses different ways to help us learn.* This teacher gives me a new assignment when I already understand something.
4. Develops problem solving and critical thinking skills for all students	This teacher asks us to explain our ideas to each other. This teacher asks me to explain (say) why I think what I think.*
5. Sets rigorous academic expectations for students	This teacher pushes (tells) everybody to work hard.* This teacher tells me I can do challenging work.
6. Provides relevant and useful feedback to students	This teacher helps me understand why my answers are right or wrong.* This teacher tells us how our work will be graded.
7. Sets and implements classroom routines and procedures that support student learning	Everybody knows what they should be doing and learning in this class. This class stays busy and does not waste time.*
8. Establishes and maintains standards for student behavior	Students in this class make sure everyone follows the rules. This teacher does not allow students to break the rules.*
9. Creates a safe and secure classroom environment that is organized and engages students	This teacher shows us examples of high quality work (work that is very good).* Students in this class celebrate when someone does a good job.
10. Establishes a climate that promotes fairness, respect, and diversity	This teacher shows respect to all students. I feel like this teacher really cares about me.*
n/a	I would choose to have this teacher again.*

Source. http://archive.austinisd.org/inside/initiatives/compensation/index.phtml
Note. Items on the kindergarten through Grade 2 survey are marked with an asterisk.
Response options were *always (4), sometimes (3), not a lot (2),* and *never (1).*
Items were selected to represent the highest rubric score for each observational domain and do not necessarily reflect the title of the domain.

Grades 3 through 5. The instrument for students in grades 3 through 5 included 21 items and a more traditional scale accompanied the faces that appeared on the primary-grade version. Students were asked to respond about their homeroom teacher and a randomly selected special area teacher. Surveys for 3rd through 5th graders were administered by teams of two researchers, as opposed to three for the primary grade students. After explaining the instructions, students were given an opportunity to ask questions. Although the number of questions varied among classrooms, students did ask several common questions. Among the most common was, "Why are we taking this survey?" Researchers explained that in college classes, students are asked to rate their professors at the end of the year, and we (i.e., the researchers) want to ask students in elementary school to do this, too. Another common question was "Will my teacher see my answers?" Although students had just heard in the instructions that their responses were private and would not be seen by their teachers, many wanted extra reassurance their responses would not be shared. Finally, students commonly asked, "Will my teacher get in trouble/get fired because of the answers to the survey?" These students seemed concerned that the survey would be punitive in some way. This question and answer time was valuable and should be incorporated into future administrations.

Most older elementary students completed both surveys in about 20 minutes. Students worked attentively on the surveys, and most were able to read the items without help. In a few cases, researchers read the questions directly to students who were struggling. If classroom teachers are responsible for survey administration in the future, Instructions should be provided for addressing this situation. Additionally, although most students completed the surveys within 20 minutes, some had not finished (or even gotten close to finishing) the survey after 30 minutes. Researchers agreed a time limit should be imposed so the survey does not consume too much class time.

When the surveys were completed and returned, students were given the opportunity to provide feedback and ask additional questions. Several students reported that some of the questions did not apply to their special area teachers, and they were unsure how to answer them. Although the survey allowed for students to select "don't know," a "does not apply" option should be considered. A few students reported some of the questions were difficult to understand; researchers also noted this during survey administration. These items were examined during reliability analyses and some were identified for removal or alteration.

Grades 6 through 12. Middle and high school students completed surveys during their advisory classes, with facilitation by advisory teachers. Packets with instructions and scan forms were distributed to advisory teachers the day before survey administration. Students received three or four scan forms each (at middle and high schools, respectively), with removable labels identifying the student (for ease of distribution) and the teacher and class to be evaluated. The administration process went smoothly. The participating middle school's principal lengthened advisory by 15 minutes to accommodate administration (for a total advisory time of between 45 and 50 minutes). The high school advisory period was not lengthened, but researchers received no indication that the survey took longer than the allotted time (35 minutes). Only about 75% to 80% of high school advisory teachers administered the survey, so data for high school teachers were not as comprehensive as intended.

PROPERTIES OF THE INSTRUMENTS

Sample of participants. Overall, 1,531 students participated in the pilot study, rating a total of 174 teachers (Table 1). The sample of students was considerably smaller for each of the elementary instruments than for the secondary instrument, resulting in far fewer teachers evaluated at the elementary level (27 elementary versus 147 secondary teachers). Also, because each elementary student in Grades 3 through 5 provided feedback about one of three special area teachers at the school, special area teachers were evaluated by nearly three times as many students as were homeroom teachers at those grades (56 versus 17, on average).[3]

Table 2. Summary of participants and surveys received, by grade span

	Kindergarten–2nd grade	3rd–5th grade	6th–12th grade	Overall
Number of students participating	220	172	1,139	1,531
Number of student surveys completed	220	347	4,303	4,870
Number of teachers assessed	14	13	147	174
Average number of surveys per teacher assessed*	16	26**	29	28
Range in number of surveys per teacher assessed*	1 to 23	1 to 66	2 to 75	1 to 75

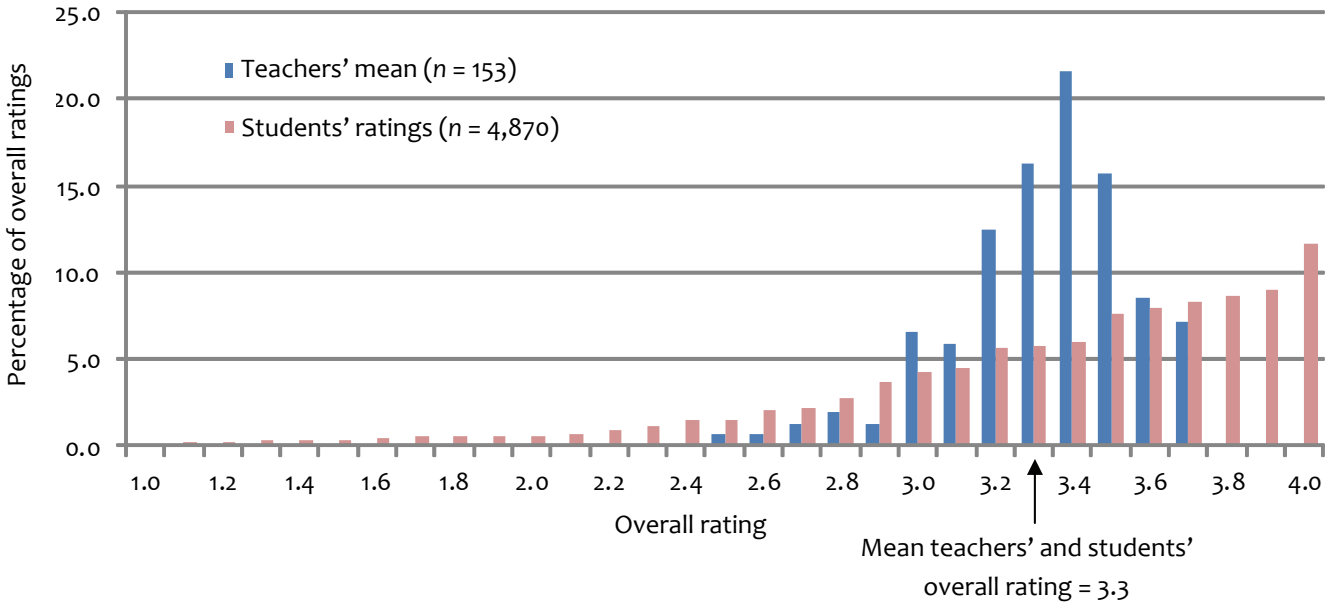
* For subsequent analyses involving teacher-level data, only teachers with 10 or more surveys were included.

** Special area teachers (i.e., art, music, and physical education) were evaluated by samples of students from each grade; the average number of surveys for homeroom teachers at grades 3 through 5 was 17, and the average number of surveys for special area teachers was 56.

Distribution of ratings. Students' ratings for each item were converted to a scale ranging from 1 (*never*) to 4 (*always*). Responses of "don't know" were excluded. Scores for survey items were averaged to create one overall mean score per survey completed. Although students' ratings for their classroom experiences ranged from 1.0 to 4.0, students' ratings of experiences varied within the same classroom; therefore, no teacher received an average rating below 2.5 or above 3.7 (Figure 1). Teachers received an average rating of 3.3, and students rated their classroom experiences an average of 3.3 on the scale of 1 to 4.

[3] Students in grades kindergarten through Grade 2 also provided feedback about special area teachers; however, those results were excluded due to concerns about their validity.

Figure 1. Distribution of Students' and Teachers' Mean Ratings Across All Items



Factor analysis and reliability. Items initially were examined separately for each grade span, with principal components analyses using Varimax rotation. Results identified four problematic items that split loaded on the two factors identified. These four items, which included the statement previously identified as problematic for younger students, were removed from subsequent analyses. After preliminary results showed similar response patterns for each grade span, responses for the 20 items asked of Grades 3 through 12 were combined for a subsequent analysis, which revealed two factors corresponding to concepts of *instructional strategies* (Factor 1) and *student participation* (Factor 2). Table 3 identifies the observational domain each item was intended to assess, along with item descriptive statistics and results from exploratory factor analysis. It should be noted that items were written to assess students' perceptions of their classroom experiences of the highest level of each observational domain; however, the domain titles for the observational rubric did not necessarily reflect the indicators that defined the highest quality teaching in that domain (see Appendix A for an example).

Table 3. Factor Loadings and Descriptive Statistics for Student Course Feedback Survey Items (Grades 3–12)

	Item	Factor 1 loadings	Factor 2 loadings	Mean	SD
19.	This teacher shows respect to all students.	.79	.15	3.63	.70
18.	I feel like this teacher really cares about me.	.75	.34	3.45	.82
4.	This teacher answers my questions when I don't understand something.	.74	.26	3.54	.75
3.	This teacher has several good ways to explain each topic so I understand.	.73	.35	3.50	.77
5.	This teacher uses different ways to help us learn.	.72	.35	3.49	.76
12.	This teacher helps me understand why my answers are right or wrong.	.69	.36	3.42	.84
9.	This teacher pushes (tells) everybody to work hard.	.64	.30	3.50	.80
17.	This teacher shows us examples of high quality work.	.63	.37	3.42	.80
10.	This teacher tells me I can do challenging work.	.62	.37	3.35	.86
14.	Everybody knows what they should be doing and learning in this class.	.62	.35	3.44	.74

(continued on next page)

Table 3. Factor loadings and descriptive statistics for student course feedback items (grades 3— 12) (continued)

	Item	Factor 1 Loadings	Factor 2 Loadings	Mean	SD
2.	This teacher asks us to lead lessons.	.16	.78	2.87	1.01
7.	This teacher asks us to explain our ideas to each other.	.38	.70	3.18	.89
20.	Students in this class celebrate when someone does a good job.	.24	.69	2.94	1.06
8.	This teacher asks me to explain (say) why I think what I think.	.39	.68	3.18	.90
1.	This teacher asks us to share our thoughts	.38	.67	3.24	.90
16.	Students in this class make sure everyone follows the rules.	.40	.51	3.20	.89
13.	This class stays busy and does not waste time	n.a.	n.a.	3.26	.82
6.	This teacher gives me a new assignment when I already understand something.	n.a.	n.a.	3.22	.89
15.	This teacher does not allow students to break the rules.	n.a.	n.a.	3.45	.89
11.	This teacher tells us how our work will be graded.	n.a.	n.a.	3.33	.86

Note. Items with factor loadings of n.a. were removed after preliminary analyses concluded they were problematic.

Items were combined into scales according to the factor loadings. Internal reliability coefficients were computed and suggested strong reliabilities for each subscale and for all items combined[4] for Grades 3 through 12. Reliability was moderate for the eight items asked of kindergarten through Grade 2 (Table 4).

Table 4. Internal Reliability Coefficients for Instructional Strategies Scale, Student Participation Scale, and Overall

	Cronbach Coefficient α (standardized)	Mean	SD
Instructional strategies (3–12)	.91	3.46	.60
Student participation (3–12)	.84	3.09	.72
Overall (3–12)	.88	3.31	.60
Overall (kindergarten–2)	.64	3.39	.44

Students also were asked one additional item, "I would choose to have this teacher again," to determine the extent to which their ratings of classroom experiences reflected their opinions about the teacher. Teachers with high mean ratings for instructional strategies and student engagement were more likely than were teachers with low mean ratings for these scales to have students who would choose them again. Mean ratings teachers received for instructional strategies were somewhat more related (r = .77) than were ratings of student engagement in the classroom (r = .60) to students' reports they would choose that teacher again (Table 5).

Table 5. Pearson's *r* Coefficients for Correlations of Teachers' Mean Scores With Each Other and With Their Mean Ratings Received for the Item, "I would choose to have this teacher again."

Scale	Instructional strategies	Student participation	Overall	"I would choose to have this teacher again."
Instructional strategies	1.0			
Student participation	.75	1.0		
Overall (n = 153)	.96	.90	1.0	
"I would choose to have this teacher again."	.77	.60	.72	1.0

[4] Items from each subscale were combined with one item that did not load on either factor but was believed to be important (number 13) to create an overall score.

Validity. Students' ratings were correlated with teachers' value-added scores. Results were analyzed separately for each subject area and school level due to differences in value-added methodology and data distributions. Sample sizes for disaggregated groups were too small to allow for specific conclusions, but the pattern of correlation coefficients suggests positive relationships between students' feedback and teachers' value-added scores (Table 6).

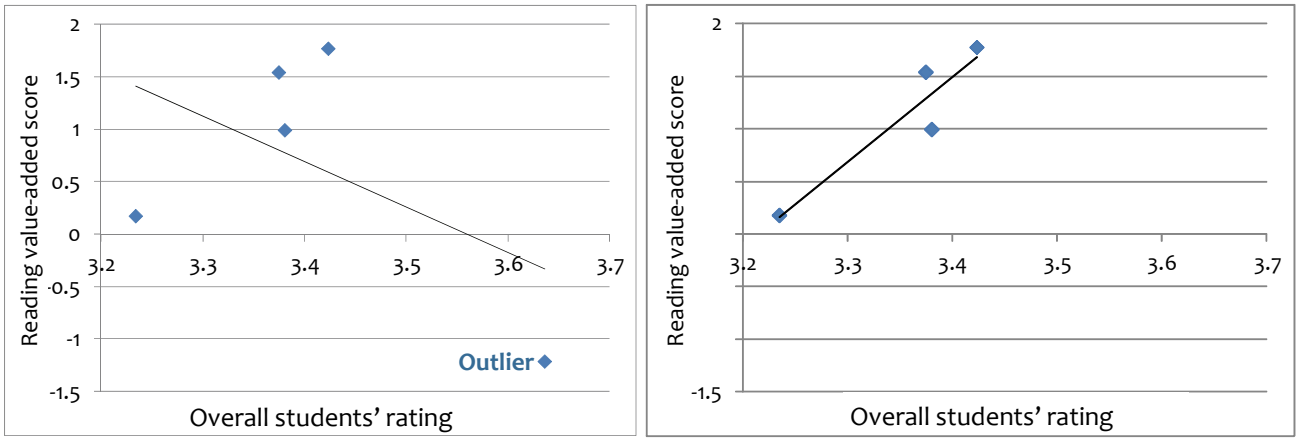
Table 6. Correlations Between Subject Area Value-Added Scores and Students' Ratings of Teachers, by Level

	Grades 9–10				Grades 6–8				Grades 4–5	
	M (n = 7)	R (n = 6)	S (n = 5)	SS (n = 3)	M (n = 9)	R (n = 10)	S (n = 3)	SS (n = 3)	M (n = 4)	R (n = 4)
Overall rating	0.20	0.81	0.26	0.70	0.52	0.06	0.96	0.77	0.74	0.94
Instructional practice	0.15	0.81	-0.18	0.56	0.53	0.03	0.97	0.69	0.80	0.12
Student participation	0.22	0.81	0.18	0.92	0.41	0.13	0.91	0.98	0.63	0.92

Note. Due to small cell sizes, p values were not considered meaningful and results must be interpreted with caution.
One extreme outlier was removed from the elementary analyses above. See Figure 2.
M = mathematics; R = reading/English language arts; S = science; SS = social studies

Because the sample of teachers with value-added data for each subject and grade range was small, data must be interpreted with caution. Extreme values influence these analyses and must be considered. For example, one elementary outlier made the following difference in correlations of student feedback with reading value-added scores (Figure 2). However, although sample sizes were very small, the pattern of correlations suggests feedback from students can indeed distinguish teachers with high and low value-added scores across grade levels and subject areas.

Figure 2. Scatterplots Showing Relationship Between Overall Students' Ratings and Reading Value-Added Scores for Elementary Teachers, With and Without One Outlier



Results from the student course feedback survey also were compared with ratings teachers received from announced and unannounced peer and administrator observations using the rubric from which the student course feedback survey was developed. Data suggest students' course feedback is moderately correlated with observational ratings from other teachers and administrators (Table 7).

Table 7. Correlations Between Student Ratings of Teachers and Observational Ratings by Peer Observers and Administrators

	Students' ratings of teachers' instructional strategies	Students' ratings of students' participation	Overall students' ratings
Announced peer observation (*n* = 130)	.32**	.29**	.33**
Unannounced peer observation (*n* = 134)	.20*	.20*	.20*
Announced administrator observation (*n* = 77)	.42**	.41**	.37**
Unannounced administrator walkthrough average (*n* = 87)	.42**	.39**	.44**

* $p < .05$
** $p < .05$

Results from correlations of students' feedback with teachers' value-added and other observational scores suggest the student feedback instrument is a valid measure of teacher quality, defined as students' growth in reading, mathematics, and science and the dimensions on the observational rubric. It must be noted that student survey items represented specific indicators within each rubric domain, and did not capture the totality of indicators representing the highest score for each respective domain (see Appendix A). Thus, the moderate correlations between students' and other sources of feedback may reflect the limited behaviors assessed on the student survey. Additionally, students' feedback was based on their experiences as a whole, while peer and administrator ratings were intended to represent a point in time. It also is possible that administrators' ratings were influenced by other knowledge, resulting in a correlation between students' and administrators' ratings that was greater than that between ratings of students and peer observers.

CONCLUSION

Evidence from this pilot study suggests feedback about teachers can be reliably obtained from students at all grade levels, and that such feedback can provide a valid supplemental measure of teacher quality to accompany additional indicators. Results suggest the pilot instrument reflected two distinct factors of classroom teaching: instructional strategies and student participation, and items generally performed well together. Although sample sizes were limited, results suggest students' ratings were related positively to teachers' value-added scores. Students' ratings also were related positively to ratings teachers received from peer observers and administrators. With some minor modifications, the student course feedback process can be a valuable resource for educators in AISD.

APPENDIX

Appendix A. Example Observational Domain and Respective Student Feedback Items

Domain	Level 1 description	Level 2 description	Level 3 description	Level 4 description	Student survey
Provides relevant and useful feedback to students	Students do not engage in self-assessment Students do not understand assessment criteria and how they will be evaluated Provides incorrect feedback or no feedback during a lesson Does not provide feedback to students that explains why they have or have not met the standard	Some students engage in self-assessment Students understand some assessment criteria and how they will be evaluated some of the time Provides limited feedback during a lesson that affirms correctly understood content, clarifies misunderstood content, and extends student thinking. Provides feedback to students that explains why they have or have not met the standard some of the time.	Students engage in self-assessment Students understand assessment criteria and how they will be evaluated Provides feedback during a lesson that affirms correctly understood content, clarifies misunderstood content, and extends student thinking Provides high-quality, timely feedback to students that explains why they have or have not met the standard	Students self-assess on their understanding of lesson objectives and provide feedback to the teacher Students use assessment criteria to guide their learning Provides feedback during a lesson that affirms correctly understood content, clarifies misunderstood content, and allows students to synthesize concepts Provides high-quality, timely feedback to students that allows them to explain why they have or have not met the standard	This teacher helps me understand why my answers are right or wrong. This teacher tells us how our work will be graded.