Department of Research and Evaluation
AUSTIN INDEPENDENT SCHOOL DISTRICT

September 2013

DRE Publication Number: 12.82

Austin Independent School District Pilot Teacher Appraisal System Update: Results of the 2012-2013 Student Response Survey

Karen M. Cornetto, Ph.D.
Raymond Gross, M.Ed.

EXECUTIVE SUMMARY

In 2011–2012, the Austin Independent School District's (AISD) Department of Research and Evaluation (DRE) pilot tested three versions of the Student Response Survey for students in kindergarten through 12th grade. Each survey was designed to measure students' perceptions of their teachers' instructional practices and their classroom environments along 10 instructional domains aligned with AISD competencies of effective teaching. Overall results from the pilot study revealed favorable ratings by students of their teachers and classroom environments, and further demonstrated that reliable student feedback can be obtained from students at grade levels 3 through 12. The results also suggested three broad areas for improvement for the survey instrument and administration process: item wording for students in early grades, student participation and teacher coverage rates, and administration time and demands on campus resources. Notably, a policy change made to the AISD pilot teacher appraisal system incorporated the Student Response Survey as a component of the teacher appraisal system in 2012–2013; therefore, refinements to the instrument and the process of administration became critical.

During the fall of 2012, DRE staff met with pilot appraisal school principals, interviewed classroom teachers, and conducted a focus group with district peer observers to help guide survey revisions and to identify procedural changes that addressed each of the identified improvement areas. The feedback obtained from AISD educators was incorporated into a revised pre-K through 2nd-grade survey that was pilot-tested with 89 students. Based on the findings, a new online pre-K and kindergarten survey was developed and further tested with 48 pre-K and kindergarten students. The pilot-study administrations were conducted with tablet computers in one-on-one and small group (five students) settings.

Final revisions to the administration procedures and instruments included refining survey items, improving the response scales, developing online surveys for all elementary students, reducing the number of teachers rated by students in pre-K through 2nd grade to one teacher, and reducing the number of teachers rated by secondary students from four to three. Special education students' participation was limited to students enrolled in a teachers' homeroom class at the elementary level, and secondary campuses determined the level of special education student participation appropriate for their campus, including the survey version and administration setting most appropriate to meet their students' needs. Trained facilitators were contracted to conduct the pre-K and kindergarten survey administrations to relieve campus staff from the additional demands required by the one-on-one and small group administrations.

Nine thousand pre-K through 12th-grade students from the 12 AISD pilot appraisal schools participated in the survey. Several survey formats were employed; students in pre-K were surveyed one at a time using tablet computers, kindergarten students were surveyed in small groups using tablets and/or computers, students in grades one through five were surveyed in a whole class setting using computers, and secondary students were surveyed via paper forms distributed in their advisory classes. In total, feedback was obtained on 683 teachers and 1,620 classes during more than 1,000 survey administrations. Campus contacts and survey facilitators indicated that the administration process went smoothly at each of the grade levels, with a few isolated exceptions. The feedback received from campus contacts and survey facilitators indicated that 3rd- through 12th-grade students were able to meaningfully participate; however, some of the survey items proved difficult for lower elementary students to understand. The trained facilitators questioned the reliability of pre-K students' responses. Several life-skills and resource teachers who had administered elementary versions of the survey to their students also reported that their students were not able to respond meaningfully.

Students' ratings of their teachers were generally very favorable. Reliability and factor analyses indicated that the versions of the survey for grades 3 through 5 and grades 6 through 12 were internally consistent, and that each survey measured two distinct student observational factors: teacher behavior and student behavior. These factors were consistent with the findings in 2011–2012. The survey versions used with younger students, however, were less internally consistent and were not clearly related to the rubric domains they were designed to measure. Correlations between teachers' mean survey scores and the other observation-based ratings and measures included in the pilot appraisal system revealed weak to moderate relationships. Given, however, that students' responses on the survey were based on many hours of observation rather than a single classroom visit, and that the survey measured a more narrow set of constructs than did the comprehensive rubric used by peer and administrator observers, these moderate correlations are meaningful.

The official personnel and administration costs of the survey in 2012–2013 are estimated to be $51,331.19, equating to $5.70 per student or $75.16 per rated teacher. However, the cost associated with surveying pre-K students was much higher than for older students. Additionally, campus contacts spent an estimated 9 hours each working on the preparation and administration of the survey, and total personnel time spent administering the 1st- through 5th-grade, and 6th- through 12th-grade surveys were approximately 73 hours and 211 hours, respectively.

Key recommendations:

1. Exclude student response survey data from the teacher appraisal for teachers in pre-K through grade 2 until such time as the survey can be conducted more efficiently and the data produced are more reliable, consistent, and varied.
2. Include student response survey data in the teacher appraisal for special education teachers only on a limited basis, at the discretion of the campus administrator and special education teachers.
3. Include a formal verification process that requires principals to identify the specific courses in the master schedule that are eligible for the survey.
4. Provide training for all facilitators to familiarize them with procedures and improve the consistency of the survey administration process.
5. Eliminate paper surveys and administer the survey online to all students to improve efficiency and reduce cost.
6. Examine the extent to which the student course evaluation adds value to the appraisal results and improves differentiated support among teachers.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

INTRODUCTION

Recent research has demonstrated the importance of using multiple measures to provide teachers with comprehensive feedback about their instruction. The Measures of Effective Teaching (MET) project (2013), funded by the Bill and Melinda Gates Foundation, demonstrated that using a combination of multiple classroom observations, teachers' value-added data, and student perception surveys predicted students' achievement gains better than any single measure in isolation. The study also found that evaluation systems incorporating multiple measures limited the variability of teacher ratings from year to year.

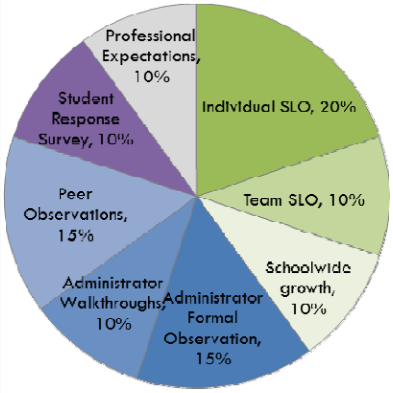
In 2011–2012, staff from the Austin Independent School District (AISD) developed a new teacher appraisal system that included observational components and measures of student growth (i.e., school-wide growth data using the state assessment and Student Learning Objectives [SLOs] results). In the same year, staff from AISD's Department of Research and Evaluation (DRE) and the Office of Educator Quality developed three versions of the Student Response Survey for students in kindergarten through 2nd grade, 3rd through 5th grade, and 6th through 12th grade. The surveys incorporated a combination of items from the Tripod instrument,[1] along with items developed by AISD. Each survey was designed to measure students' perceptions of their teachers' instructional practices and their classroom environments along 10 instructional domains aligned with AISD's competencies of effective teaching.[2] To have the ability to analyze students' responses along with multiple other measures of teacher effectiveness, DRE staff piloted the Student Response Survey with 1,521 students at the three schools participating in the pilot appraisal: Sunset Valley Elementary, Webb Middle School, and Lanier High School. Twenty-seven elementary teachers and 147 secondary teachers were rated by their students in 2011–2012. Researchers documented challenges with the survey administration, analyzed results of the survey, and examined the properties of each instrument.

Results from the pilot study revealed generally favorable ratings by students of their teachers, while confirming that reliable student feedback can be obtained at grade levels 3 through 12

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

2. The competencies of effective teaching were selected by the AISD appraisal system working group, which was a committee comprising teachers, principals, an associate superintendent, district chief officers, and other community stakeholders.

AISD Pilot Teacher Appraisal Program



AISD redesigned its teacher appraisal system with the goal of creating a more robust feedback loop for teachers that draws on the results of statewide and teacher-developed assessments and observations of classroom practice. The newly designed system differs from the state system of PDAS in that it provides a significant weight to student growth and focuses on three domains of practice: instructional practice, classroom climate, and professional expectations.

The new teacher appraisal system was initially piloted at 3 REACH campuses in 2011–2012, then revised and piloted at 12 REACH campuses during the 2012–2013 school year. Ultimately, AISD is committed to implementing a system that works toward the continuous improvement of teaching, learning and leadership, that promotes student success.

For more information, please visit http://www.austinisd.org/edquality/teacher-appraisal-system

(Schmitt, 2012). The results further suggested that feedback from students can distinguish teachers with high and low value-added scores across grade levels and subject areas. Although the initial pilot results were encouraging, findings from the study identified three broad areas for improvement:

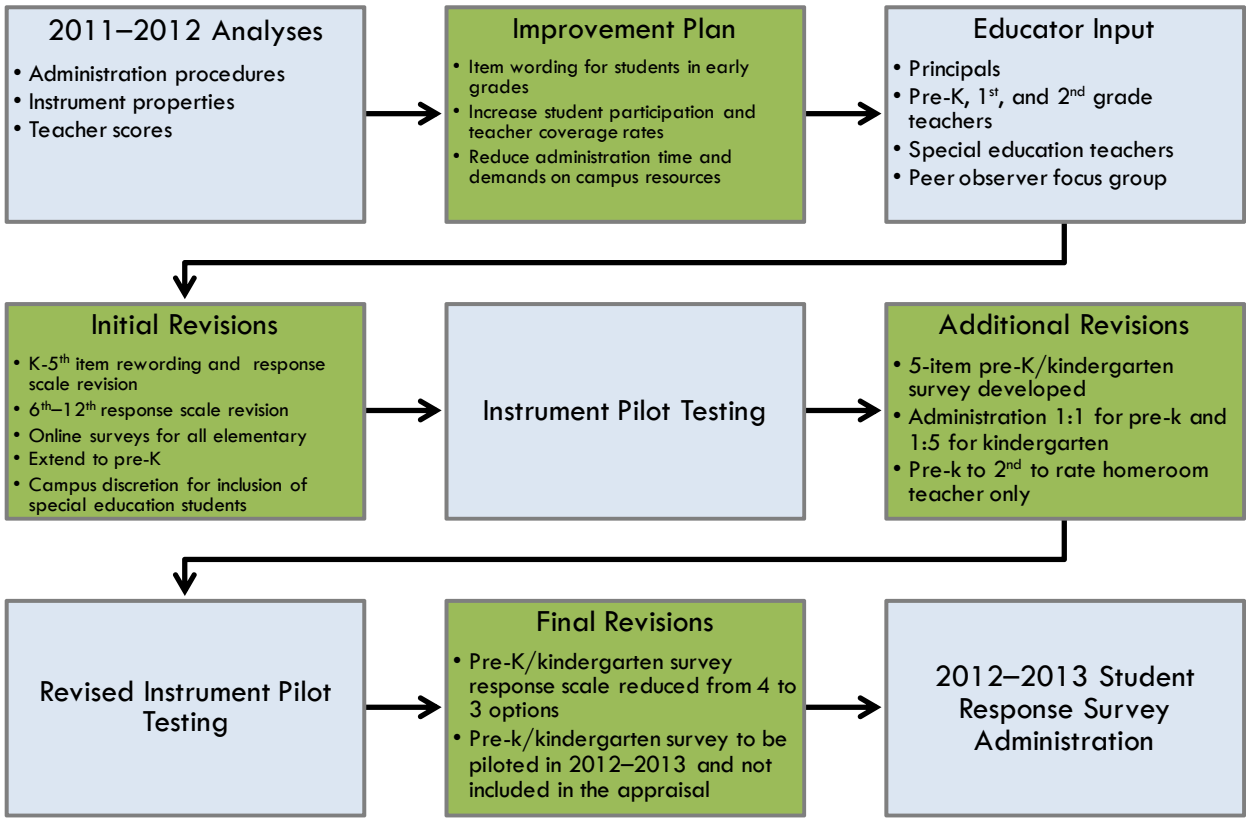
- Item wording for students in early grades
- Student participation and teacher coverage rates
- Administration time and demands on campus resources

In 2012–2013, program managers revised the teacher appraisal system to include the Student Response Survey along with classroom observations, professional expectations, SLOs, and school-wide growth measures (EVAAS). The pilot appraisal system also was expanded from three schools to 12 schools. This report provides an account of the revision activities undertaken, documents changes made to the Student Response Survey instrumentation and procedures, and summarizes findings from the 2012–2013 Student Response Survey administration at the 12 AISD pilot teacher appraisal schools.

Instrument and Process Revisions

During the fall of 2012, DRE staff worked with AISD educators to revise the survey instruments and to identify procedural changes that addressed each of the identified improvement areas. Based on the AISD educators' feedback and the results of extensive pilot testing, several changes were implemented during the 2012–2013 Student Response Survey administration (Figure 1).

Figure 1. 2012-2013 Student Response Survey Revision Process



Each of the final revisions to the 2012–2013 Student Response Survey administration incorporated findings from the 2011–2012 Student Response Survey pilot study; feedback gained through the principal, teacher, and peer observer meetings; and the subsequent pilot testing of a revised pre-K through 2nd-grade survey instrument. A summary of the teacher interviews and pilot testing can be found in Appendix A. Selected survey items at each level were revised or eliminated, response scales were modified, and a fourth survey for pre-K and kindergarten students was developed. Procedural changes included switching from paper to online survey administration at the elementary level and reducing the number of surveys to be completed by pre-K through 2nd-grade students and by high school students. Each of the revisions enacted was designed to enhance students' comprehension of survey items, improve the reliability of the survey instruments, increase teacher coverage through broader based student participation, and limit lost instructional time and demands on campus resources. This revision process is discussed in detail in the following sections.

Principal Meetings

In November of 2012, DRE staff met with eight of the 12 pilot appraisal school principals to update them on the 2012–2013 survey administration and determine their school's readiness to administer the survey online. The elementary principals indicated that their respective campuses were sufficiently equipped with computer labs, laptop/tablet computer carts, or both to efficiently administer the survey. The elementary principals also indicated that large scale online assessments (e.g., the Texas English Language Proficiency Assessment System [TELPAS]) required scheduling and administration procedures that could be replicated for the Student Response Survey administration. Secondary principals reported limited technology resources, logistical and scheduling constraints, and concerns about extensive time requirements of an online administration; their preference was for a paper survey administered during an instructional or advisory period of their choosing.

Teacher Interviews

Elementary grade teachers. Interviews were conducted with seven pre-K through 1st-grade teachers, including special education, bilingual, and general education teachers, throughout November of 2012. During the interviews, teachers were asked to review proposed revisions to the survey for students in kindergarten through grade 2 to identify any questions that they perceived might be difficult for their students to understand. Teachers also were asked to evaluate pre-K students' abilities to meaningfully respond to the survey items reviewed, as well as to offer alternatives to the graphics used to visually represent each response option (e.g., faces ranging from happy to sad).

Most of the concerns teachers expressed centered around problematic wording rather than the concepts measured by the survey items. For example, most teachers commented that the words "challenging" in "this teacher gives me challenging work" (item 9) and "examples" in "this teacher shows us examples of work that is very good" (item 17) may be difficult for primary grade students to understand. Additionally, teachers felt that wording for some items should be more concrete. Phrases such as "this teacher", "share your ideas" (item 1), "checks to make sure" (item 2), "uses different ways" (item 3), and "shows respect" (item 11) were singled out as potentially confusing for lower grade level elementary students. Teachers' recommendations included changing "this teacher" to "my teacher" and "shows respect" to "treats with respect." They also suggested describing how the instructional practice or class environment depicted in the questions would "look, feel, and sound" in ways that were familiar to students. The pre-K teachers felt equally confident that their students would be able to meaningfully

respond if provided specific examples and reading support. However, many teachers noted that students' behavior during the administration may be a problem, especially if the survey were conducted in a large group with unfamiliar facilitators. Each of the teachers also expressed the concern that having other teachers switch classes to facilitate the survey, having smaller group administrations, or both would impose significant demands on staff time and would require substantial support from the campus administration.

Most teachers indicated some degree of concern or dislike with the happy/sad faces provided for each response option. They thought students might have a difficult time discerning between the different faces, and that greater differentiation was needed. Alternative suggestions included reducing the response choices from four options to three; color coding response options (e.g., green to red); aligning the response scale with other school-wide systems commonly used (e.g., behavioral charts, other rubric based monitoring systems); and using a range of stars from 1 to 4. Teachers' concerns with the response scale mirrored DRE researchers' observations made during the May 2012 pilot-administration, when it was noted that elementary school students may have associated the range of happy faces from smiling to sad with their personal feelings (i.e., like or dislike of the teacher) rather than taking them as an indication of the frequency of occurrence for each of the instructional activities or classroom environments depicted in the survey items.

Special education teachers. Two elementary and one high school special education resource teachers were interviewed to help DRE staff identify the range of abilities of the students they serve, and to better understand the level of accommodation that would be required for resource and life-skills students to participate. All three teachers reported that their resource students, and many inclusion students, would need additional support, including oral administration in small groups (six or fewer) outside their inclusion setting. The two elementary teachers did not identify any major concerns about providing these supports, and one noted that "special education resource teachers could switch classes with one another and administer the survey to each others' small groups as they receive them throughout the day." Each teacher indicated that such an arrangement would not require additional personnel beyond bilingual support, and could feasibly be completed in one to two days. This sentiment was echoed by the secondary special education teacher, who added that the surveys could be administered in English classes to avoid pulling special education students out of their inclusion setting to receive the appropriate accommodations. This teacher was concerned that administering the survey during advisory period, or another class in which resource students are included, would require substantial coordination on the part of campus staff to provide students appropriate accommodations, and might signal differences among students in front of their peers. The two elementary teachers also indicted that any attempt to administer the survey to their resource students in a whole group setting (i.e., inclusion classroom) would require significant administrator support for scheduling and relief of staff from other duties to assist. All three teachers additionally noted that it would be difficult for the majority of life-skills students to meaningfully participate. Although each of the teachers stated that it would be valuable to capture students' perceptions of their inclusion teachers, none were able to identify a feasible method that would work in all settings.

An additional issue was raised by the special education teachers that highlighted concerns about the applicability of the survey to both the inclusion and resource settings. Each teacher reported that a student's ability to observe and associate the instructional activities depicted by each survey item with their inclusion teacher would depend on the level of co-teaching taking place in the classroom. They felt that the classroom teachers often controlled the degree to which inclusion teachers were able to actively

partake in classroom instruction, and that variation occurred from class to class. Additionally, they felt that the survey items as currently worded did not necessarily apply, given the specialized instruction their students required. Specific differences between what takes place more regularly in the resource classroom, compared with what is expected in the general education classroom, included, in the words of these teachers, "greater use of direct teaching methods" and "continuous repetition."

Specific survey change recommendations (e.g., offering concrete examples, and anchoring the response scale) offered by the resource teachers paralleled those made by the general education elementary teachers. The teachers expressed the opinion that these changes would be required to avoid inconsistent responses caused by, in the words of one teacher, "different perceptions of what the questions and answers choices may look like in practice." Each of the special education teachers additionally indicated that the 11-item kindergarten through grade-2 version of the survey would be most appropriate for their students, given its simplified language and reduced number of items.

Peer Observer Focus Groups

In December of 2012, DRE staff met with 16 AISD peer observers to obtain additional input on changes to the 2012–2013 survey. The peer observers' feedback was particularly valuable given their in-depth knowledge of the classroom observation rubric domains and scoring system, as well as their experience observing a variety of classes in different contexts. The peer observers initially focused on developing concrete examples of what each teacher activity would "look, feel, and sound like" in practice. The group, however, determined that arbitrarily defining specific examples could exclude worthy instructional practices from the example list. They also acknowledged that the survey items must be written with some level of abstraction to more fully capture the essence of each instructional domain, rather than a single indicator. The peer observers additionally provided guidance for anchoring the response options to the peer and administrator observation scoring rubric (e.g., "a lot of the time" indicates "every day every subject") and recommended modifying the graphics that accompanied the survey response scale to show different size clocks, with successively smaller clocks illustrating lesser degrees of time, corresponding with the response scale options (Appendix B).

Pilot Testing

In January of 2013, students (n = 89) from five classes at two elementary schools participated in a pilot test of a revised survey intended for students in pre-K through grade 2. The pilot was used to determine students' comprehension of the survey questions and to assess their abilities to navigate the online survey. Feedback from teachers about their impressions of the survey instrument and the administration process was obtained through debriefing meetings held with the classroom teachers following the pilot administrations. Students from three classes (pre-K, kindergarten, and 1st grade) participated in the online survey pilot, and two classes (pre-K and 1st grade) completed paper surveys. Administration for each class was conducted in whole group settings (by class). First-grade students were asked to rate their homeroom teacher and to respond to a second survey about a special areas teacher. Pre-K and kindergarten students responded to one survey about their homeroom teacher only. The teachers swapped classes to administer the paper survey, and the three online administrations were facilitated by the homeroom teachers, with support from DRE staff.

Feedback from the facilitating teachers and observations made by DRE researchers revealed that many students required significant support, including individual reading assistance, technology instruction, and behavioral redirection. The paper administration was particularly challenging; DRE observers noted that

students' engagement declined as they progressed through the survey. Many students displayed a general lack of comprehension, an overall inability to follow along, and disruptive behavior during both the pre-K and 1st-grade paper administrations. The pre-K pilot survey was not completed, and the 1st-grade administration lasted nearly 45 minutes.

DRE researchers noted higher levels of student engagement with the online pilot survey. However, the students' abilities to navigate the online system varied widely, posing a unique set of challenges for survey facilitators. Despite the additional technology support required, online survey administration times were lower than times for paper administration. Overall administration times, including logging students into the online survey and reviewing survey instructions with students, averaged toward the upper target range of 30 to 35 minutes. The reduced administration times appeared to be directly related to better management of the survey administration process and to greater student interest during the survey administration, both of which were attributed to the survey's online format.

Some of the instructional activities described in the pilot survey were difficult for students to identify in practice. Pre-K and kindergarten students in particular had trouble with pilot survey items related to specific instructional strategies designed to promote meta-cognition and deeper learning, such as having students talk with their peers about what they "know and think" (item 1) and being asked to explain "why [they] think what [they] think" (item 4). Elementary students also appeared to have difficulties discerning between "different ways" of learning (item 3) and fully identifying the array of methods a teacher may use to check for students' understanding (item 2). The students displayed fewer difficulties responding to survey items depicting more concrete concepts that are often more explicitly discussed in class, such as "learn new things" (item 5), "stay busy working" (item 7), "behave the way the teacher wants" (item 8), and "treat with respect" (items 10 and 11).

Based on the results of the interviews, focus groups, and pilot testing, a new version of the survey was developed for pre-K and kindergarten students that included five items that tested well during the pilot (Table 1). It also became clear that pre-K and kindergarten students would need one-on-one and small group (five students) administrations, respectively, in order to receive the level of reading support needed to comprehend the survey items, and the computer support needed to successfully navigate the online survey. It was determined that students in pre-K through grade 2 would complete one survey about their homeroom teacher only.

The new pre-K and kindergarten survey was tested with three small groups of kindergarten students (n = 15) and 33 pre-K students from four separate classes. DRE staff noted that the small group and one-on-one survey administrations allowed survey facilitators to provide greater individualized support to students and significantly helped maintain the integrity of the survey administrations. The kindergarten revised pilot administrations were conducted in the computer lab, using desktop computers. The administration times averaged roughly 10 minutes, and the kindergarten students appeared to understand the questions and respond appropriately with the small group support. The pre-K online revised pilot survey administrations were conducted one-on-one with tablet computers. Average administration time was 6 minutes for students to complete the survey, and roughly another 6 minutes to retrieve students and to return them to their classrooms. The pre-K students appeared to comprehend the questions and meaningfully respond in most cases. However, many students demonstrated difficulty distinguishing between the *sometimes* and *a little of the time* response options. Final revisions to the 2012–2013 pre-K and kindergarten survey reduced the response options from four choices to three by eliminating *a little of the time* as a response choice. Due to the extensive revisions made to the finalized

pre-K and kindergarten survey and outstanding questions about comprehension and administration, program managers decided that 2012–2013 would serve as a pilot year and that the results would not be included as part of the pre-K and kindergarten teachers' appraisal.

Table 1. 2012-2013 Prekindergarten (Pre-K) and Kindergarten Student Response Survey Revision Summary

Domain	Pre-K through 2nd grade piloted version  A lot of the time  Sometimes  A little of the time ○ Never ? Don't know	Final pre-K and kindergarten version  A lot of the time  Sometimes Never ? Don't know
Student Engagement	1. How often does your teacher ask you to talk with each other about what you know and think?	
Checks for understanding	2. How often does your teacher check to make sure you understand what you are learning?	
Differentiated instruction	3. How often do you get to learn in different ways in your class?	
Problem-solving & critical thinking	4. How often does your teacher ask you why you think what you think?	
Rigorous academic expectations	5. How often does your teacher help you learn new things?	1. How often does _____help you learn new things?
Relevant and useful feedback	6. How often does your teacher help you understand why your answers are right or wrong?	
Classroom routines & procedures	7. How often do students in your class stay busy working and not wasting time?	2. How often do students in _____'s class stay busy working?
Classroom management	8. How often do students in your class behave the way the teacher wants them to?	3. How often do students in your class behave the way _____wants them to?
Classroom safety, security, organization	9. How often does your teacher show your class work that is very good?	
Classroom fairness, respect, diversity	10. How often do students in your class treat the teacher with respect?	4. How often do students in _____'s class treat the teacher with respect?
Classroom fairness, respect, diversity	11. How often does your teacher treat students in your class with respect?	5. How often does _____treat students in your class with respect?

Source. 2012–2013 revised pre-K through grade 2 Student Response Survey

Note. In the online survey, teacher name was inserted in areas marked "____."

No further refinements were made to the resulting 1st and 2nd grade Student Response Survey.

Summary of Revisions to Grades 3 Through 5 and 6 Through 12 Versions

Revisions to the surveys for grades 3 through 5 and grades 6 through 12 were made in response to the item analyses conducted in 2011–2012 and to feedback from educators during the focus groups and interviews. For 2012–2013, the response scales were changed from a 4-point general frequency scale (ranging from *always* to *never*) to a more specific 4-point scale designed to capture the amount of time teachers engaged in each activity (ranging from *a lot of the time* to *never*). The clocks identified for the

pre-K and kindergarten version (Appendix B) also were used for all the elementary versions of the survey to visually represent the response options. In addition, the item analyses of the 2011–2012 student responses revealed several problematic items at each grade level, and with the input of AISD educators, those items were reworded or removed (Table 2).

Based on feedback from the facilitators who conducted the survey in 2011–2012, several procedural changes were made, the most important of which was a reduction in the number of surveys that students answered. In 2011–2012, students in kindergarten through grade 2 responded to two surveys: one about their homeroom teacher and one about a special area teacher. Observations made by DRE researchers during the elementary administrations revealed that many students had a difficult time responding to two surveys because it was challenging to hold their attention, and some children became confused about whom they were rating. In 2012–2013, students in kindergarten through grade 2 responded only about their homeroom teacher. Similarly, feedback from the advisory teachers at high schools suggested that responding about four teachers was challenging during one class period; therefore, the number of teachers about whom high school students responded was reduced to three in 2012–2013. Finally, all elementary surveys were moved from paper to a web-based platform. This offered the opportunity to improve the readability and ease with which students could understand the question by inserting their teachers' names into each of the survey items (Table 2).

Table 2. Summary of Revisions to Student Response Survey, by Survey Version

Domain	2011–2012 survey items	Grades	2012–2013 revised survey items
Student engagement	1. This teacher asks us to share our thoughts.	1–2	How often does _____ask you to talk with each other about what you know and think?
		3–5	How often does _____ ask you to share your ideas with each other?
		6–12	This teacher asks us to talk about our ideas with each other.
Student engagement	2. This teacher asks us to lead lessons.	3–5	How often do you get to do a lot in this class, not just listen to _____ talk?
		6–12	We get to do a lot in this class, not just listen to the teacher talk.
Checks for understanding	3. This teacher has several good ways to explain each topic so I understand.	3–5	How often does _____ use several good ways to explain each topic so that you understand?
		6–12	This teacher has several good ways to explain each topic so I understand.
Checks for understanding	4. This teacher answers my questions when I don't understand something.	1–2 & 3–5	How often does _____check to make sure you understand what you are learning?
		6–12	This teacher checks to make sure we understand what we are learning.
Differentiated instruction	5. This teacher uses different ways to help us learn.	1–2	How often does _____ give you special attention to help you learn something?
		3–5	How often does _____ use different ways to help you learn?
		6–12	This teacher uses different ways to help us learn.
Differentiated instruction	6. This teacher gives me a new assignment when I already understand something.	3–5	How often does _____ explain difficult things clearly?
		6–12	This teacher explains difficult things clearly.

Table 2. Summary of Revisions to Student Response Survey, by Survey Version (Continued)

Domain	2011–2012 survey items	Grades	2012–2013 revised survey items
Problem-solving and critical thinking	7. This teacher asks us to explain our ideas to each other.	3–5	How often does ____ help you think about things in new ways?
		6–12	This teacher helps me think about things in new ways.
Problem-solving and critical thinking	8. This teacher asks me to explain (to say)* why I think what I think.	1–2	How often does ____ ask you to explain why you answered the way you did?
		3–5	How often does ____ ask you to explain why you think what you think?
		6–12	This teacher asks me to explain why I think what I think.
Rigorous academic expectations	9. This teacher pushes (tells)* everybody to work hard.	1–2 & 3–5	How often does ____ give you challenging work?
		6–12	This teacher gives me challenging work.
Rigorous academic expectations	10. This teacher tells me I can do challenging work	3–5	How often does ____ show you why this class is important?
		6–12	This teacher shows me why this class is important.
Relevant and useful feedback	11. This teacher tells us how our work will be graded.	3–5	How often does everybody know what they should be learning in ____'s class?
		6–12	Everybody knows what they should be learning in this class.
Relevant and useful feedback	12. This teacher helps me understand why my answers are right or wrong.	1–2 & 3–5	How often does ____ help you understand why your answers are right or wrong?
		6–12	This teacher helps me understand why my answers are right or wrong.
Classroom routines and procedures	13. This class stays busy and does not waste time.	1–2	How often do students in ____'s class stay busy working?
		3–5	How often do students in ____'s class stay busy working and not waste time?
		6–12	This class stays busy and does not waste time.
Classroom routines and procedures	14. Everybody knows what they should be doing and learning in this class.	3–5	How often does everybody know what they should be doing in ____'s class?
		6–12	Everybody knows what they should be doing in this class.
Classroom management	15. This teacher does not allow students to break the rules.	3–5	How often does ____ allow students to break the rules?
		6–12	This teacher allows students to break the rules.
Classroom management	16. Students in this class make sure everyone follows the rules.	1–2 & 3–5	How often do students in your class behave the way ____ wants them to?
		6–12	My classmates behave the way the teacher wants them to.
Classroom safety security organization	17. This teacher shows us examples of high quality work (work that is very good).*	1–2	How often does ____ show examples of work that is very good?
		3–5	How often does ____ show you examples of high quality work?
		6–12	This teacher shows us examples of high quality work.

Table 2. Summary of Revisions to Student Response Survey, by Survey Version (Continued)

Domain	2011–2012 Survey Items	Grades	2012–2013 Revised Survey Items
Classroom safety, Security, organization	18. I feel like this teacher really cares about me.	3–5	How often do you feel comfortable talking about your ideas in ____'s class?
		6–12	I feel comfortable talking about my ideas in this class.
Classroom fairness security diversity	19. This teacher shows respect to all students.	1–2	How often does ____ show respect to all students?
		3–5	How often does ____ show respect to all students?
		6–12	This teacher shows respect to all students.
Classroom fairness security diversity	20. Students in this class celebrate when someone does a good job.	3–5	How often do students in ____'s class celebrate when someone does a good job?
		6–12	Students in this class celebrate when someone does a good job.
n/a	21. I would choose to have this teacher again.	All levels	Item removed from survey
Classroom fairness security diversity	New Item developed for 2012–2013 SRS for all grade level survey versions	1–2 & 3–5	How often do students in ____'s class treat the teacher with respect?
		6–12	Students in this class treat the teacher with respect.

Source. 2011–2012 Student Response Survey items, 2012–2013 Student Response Survey items

Note. In the online survey, the teacher's name was inserted in areas marked "____."

Kindergarten students answered the same set of questions during the 2011–2012 Student Response Surveys as did 1st and 2nd grade.

*2011–2012 Student Response Survey alternate item wording for kindergarten through grade 2 students are indicated in parentheses.

2012–2013 Student Response Survey Administration

Sampling

A summary of all teachers and students included in the survey can be found in Tables 3 and 4.

Elementary. All elementary homeroom teachers for pre-K through grade 5, and all art, music, and physical education teachers with 10 or more students were included (n = 199) and all pre-K through 5th-grade students assigned to a homeroom class were eligible to participate (n = 3,158)[3]. Students in pre-K through 2nd grade completed one survey about their homeroom teacher, and 3rd- through 5th-grade students completed two surveys: one about their homeroom teacher and a second about either their art, music, or physical education teacher. The special area course rated by a student for their second survey was determined by the students' birth month to ensure equitable samples.

Secondary. Teachers with a minimum of 10 students in three or more classes, yielding a minimum of 30 student surveys, were included (n = 499). Secondary students (n = 7,342) rated their teachers from three randomly selected courses. Special education resource and life-skills teachers who met the 30 record criteria also were included. Each campus developed an administration process that would work best with their students' schedules. Each campus also determined the level of special education student participation and the Student Response Survey version that best matched the needs of their resource and life-skills students.

During the meeting in November, principals were asked to provide feedback about their master schedules to verify teacher assignments and to identify courses that should be excluded when sampling courses for students (e.g., double-blocked courses and non-instructionally oriented courses, such as office aide and special assignment). Secondary principals noted multiple errors in their master schedules, including inaccurate teacher and course assignments. Additionally, slight variations in course naming conventions occurred across schools. The combination of master schedule errors, variations in course naming conventions, and principals' decisions about which courses should be ineligible for the survey added an additional level of complexity to the sampling process. In the end, the sampling process for each secondary campus was unique. In addition, the elementary class rosters did not always accurately reflect the teachers providing the students' primary instruction. For example, the 5th-grade teachers at one elementary school employing "flexible grouping" did not directly instruct the majority of the students enrolled in their homeroom class.

DRE staff worked closely with campus contacts to verify teachers' assignments and class rosters throughout the sampling processes. Master schedule and class roster verification required unanticipated additional time from and effort by campus contacts and staff, and proved to be difficult for some campuses to easily complete. Overall, the inclusion criteria captured 100% of homeroom teachers at each elementary campus (Table 3), but despite efforts to verify teacher and student rosters, errors were later uncovered during and after the 2012–2013 administration (e.g., one 4th-grade teacher was not listed as a homeroom teacher in AISD data records, and was therefore not included in the survey). Teacher coverage rates at the secondary level were slightly lower than for elementary teachers due to teachers not meeting the 30-student record minimum criterion (Table 4).

[3] Special education life-skills students, and resource students not enrolled in a general education homeroom teachers' class, were not included in the elementary student sample.

Table 3. Elementary Student Participation Rates and Teacher Coverage

School	Brown	Harris	Norman	Rodriguez	Sims	Sunset Valley	All Elementary
Sampled students	456	654	311	926	287	524	3,158
Student participants	417	612	276	854	243	499	2,898
Participation rate	91%	94%	89%	92%	85%	95%	92%
Homeroom teachers assigned to campus	29	43	21	52	20	34	199
Teacher coverage rate	100%	100%	100%	100%	100%	100%	100%
Average number of student ratings per homeroom teacher	17	17	16	20	14	18	18
Average number of student ratings per special area teacher	56	36	36	61	32	55	50

Source. 2012–2013 Student Response Survey; AISD human resources records

Table 4. Secondary Student Participation Rates and Teacher Coverage

School	Akins	Lanier	Martin	Reagan	Travis	Webb	All secondary
Sampled students	2,472	1,421	561	973	1,222	693	7,342
Student participants	2,165	1,083	464	793	1,002	595	6,102
Participation rate	88%	76%	83%	82%	82%	86%	83%
Adjusted rate minus leavers	85%	88%	90%	79%	86%	85%	86%
Teachers assigned to campus	169	113	50	81	111	58	582
Teacher coverage rate	83%	83%	74%	88%	83%	84%	83%
Average number of student ratings per teacher	45	34	36	33	32	35	37

Source. 2012–2013 Student Response Survey; AISD human resources records

Notes. The sample of students included special education students who were later determined to be unable to respond to the survey in its present form.

The adjusted student rate excludes students who left their school in May and were therefore not enrolled during the survey administration period.

The teacher rates reflect the absence of both the number to teachers who did not meet the 10 student/30 record criteria and any teachers included in the sample who did not receive at least 10 student ratings.

ADMINISTRATION

During May of 2013, 9,000 pre-K through 12th-grade students from the 12 AISD pilot teacher appraisal schools participated in the survey. Four versions of the survey were used, with one for each of the following groups: pre-K and kindergarten, grades 1 and 2, grades 3 through 5, and grades 6 through 12 (Table 5). Elementary grade level surveys were offered in both English and Spanish, and secondary surveys were offered in English only.

Feedback was obtained for 683 teachers and 1,620 classes during more than 1,000 survey administrations. In total 21,955 individual surveys were collected. DRE staff developed administration protocols, instructions, and scripts that were distributed by campus contacts to all staff members responsible for administering the survey. Additionally, campus contacts were trained on survey procedures and provided direct support as needed. Details about each administration are provided in the following sections.

Table 5. 2012–2103 Student Response Survey: Number of Items, Response Scale and Delivery Method, by Survey Version

Grades	Number of items	Response scale					Delivery method
Pre-K and kindergarten	5	 A lot of the time	 Sometimes	 Never	 Don't know		1:1 via tablet (pre-K) 5:1 via computer (kinder)
1 and 2	11	 A lot of the time	 Sometimes	 A little of the time	 Never	 Don't know	Whole class, via computer
3 to 5	21	Same as 1 and 2					Whole class, via computer
6 to 12	21	Same as 1 and 2, without graphics					Whole class, on paper

All survey facilitators were asked to complete an administration feedback form at the end of each administration. The feedback form provided facilitators the opportunity to report any irregularities and provide feedback about what happened during the sessions (Appendix C). The data gathered through these forms were reported for each survey version, when available. Four of the six elementary campuses completed some or all of their 1st-grade through 5th-grade administration feedback forms; however, feedback forms were returned by the facilitators of all pre-K and kindergarten administrations. One of the two campus contacts did not return the forms because that contact assumed the forms were to "be completed only when an irregularity occurred." The other campus contact provided no details regarding the survey administrations at that campus. Due to the limited administration feedback form return rates it is difficult to gauge the precise rate at which survey irregularities occurred across all campuses. The data were further limited by apparent inconsistencies in the reporting of time requirements and irregularities by administrators at the various campuses. It is possible that some survey administrators were more likely to complete the administration form and report an irregularity for a survey administration when an irregularity occurred, as noted by the one campus contact. Therefore, the administration feedback form data summarized for each survey version are meant to provide a general overview of the reported irregularities to help guide future improvement efforts. Roughly 26% (n = 40) of all completed elementary survey administrations forms and 9% of secondary forms (n = 20) indicated that some type of irregularity occurred. Detailed information about the nature of these irregularities can be found in Appendices D and E.

Elementary Administration

In total, 2,898 pre-K through 5th-grade students responded to the online survey, providing feedback about 199 homeroom and special area teachers. Special education students who were not assigned to a homeroom teacher did not participate. Overall, student participation rates ranged from 85% to 95% across campuses, with 92% of all eligible students responding. No evidence of any systematic denial of student participation was reported.

Administration feedback forms and post-administration interviews with campus contacts and survey facilitators indicated a smooth administration process in general. In a few instances, campus contacts reported problems associated with students' and administrators' use of the online survey system, including user errors such as students inadvertently closing the survey window before completing the survey, and survey administrators unable to log into or navigate the online survey. In each case, campus staff were able to resolve the issues with only minor disruptions to the administration process. Campus contacts and survey facilitators also reported that the online survey displayed differently on various devices (e.g., iPads, Nooks, Kindles, and laptops), requiring some students to scroll down the screen to either read the entire survey item, select an answer choice, or advance to the next question. The inability to standardize the online survey to be compatible with the variety of devices available across campuses required facilitators to provide extra support to students, and reportedly extended administration times at schools using Kindles and Nooks. Seventy-eight percent of the reported irregularities (n = 31) "did not compromise student responses," while 22% percent (n = 9) of the reported irregularities "possibly compromised student responses" (Table 6). Reported irregularities varied across grade levels and campuses. Although the data obtained from the administration forms were limited, evidence indicated that school-specific factors, including students' familiarity with the technology and survey settings (i.e., facilitators, survey group size, survey location), may have a differential impact on the reliability of students' scores across campuses.

Table 6. 2012–2013 Administration Forms: Reported Irregularities by Elementary Campus

	Brown (n=23)	Harris (n=60)	Norman (n=24)	Rodriguez (n=19)*	Sims (n=17)	Sunset Valley (n=13)*	All elementary (n=157)
Administration feedback form completion rate	100%	92%	94%	30%	94%	32%	62%
No Irregularities Encountered	70%	77%	79%	79%	65%	69%	75%
Irregularities encountered, did not compromise student responses	22%	13%	21%	21%	29%	31%	20%
Irregularities encountered, possibly compromised student responses	9%	10%	0%	0%	6%	0%	6%

Source. 2012–2013 Student Response Survey
Note. Percentages may not equal 100% due to rounding.
*Includes prekindergarten and kindergarten administrations only.

Pre-K/kindergarten. Although the results of the survey were not included in the appraisal for pre-K and kindergarten teachers, their students were surveyed to provide some feedback to teachers and to assess the feasibility of including these students in the future. The instrument for pre-K and kindergarten included five items selected from the survey for 1st and 2nd grade. The response scale included three choices: *a lot of the time*, *sometimes*, and *never*. Students also had the option to select *don't know*.

To reduce the burden on teachers, DRE staff contracted with a team of trained facilitators to schedule and conduct the pre-K and kindergarten survey. The online survey was administered one-on-one with pre-K students using IPADs. Kindergarten students participated in small groups of 5 students or less using tablet, laptop and/or desktop computers. Nine hundred twenty students completed the survey over approximately 550 individual survey administrations (table 7).

Table 7. Prekindergarten (Pre-K) Through Kindergarten Student Response Survey Participation, 2012 –2013

Grade level	Individual survey administrations	Participating students	Student participation rate	Rated teachers	Teacher coverage rate
Prekindergarten	450	450	93%	27	100%
Kindergarten	94*	470	96%	28	100%

Source. 2012–2013 Student Response Survey

Note. Teacher coverage rate represents the percentage of grade level homeroom teachers receiving student feedback.

*Survey administrations for kindergarten are estimated based on a 5:1 administration ratio. In some instances, fewer than five students were included in the small group administrations.

Instructions asked students to select the clock that best represented the amount of time each statement was true for their class. To maintain consistency across survey administrations, survey instructions and survey items were read aloud to all students. Survey items were displayed one question on the screen at a time to help pace students and ensure that students responded to the correct survey item. Teachers and researchers helped students navigate the online survey and answered questions regarding word meaning. Students who did not understand the question, students who were not sure, or students who did not have an opinion were instructed to select the *don't know* option.

Pre-K administration times averaged 5.8 minutes, and kindergartens administration times averaged 8.6 minutes. The total time required to administer the surveys, when including the time to retrieve and return students to their classes between administrations, increased to approximately 12 minutes per one-on-one pre-K administration, and 16 minutes per small group kindergarten administration. In total, 163.75 hours of total survey administration time (including setting up and transporting students to and from the survey location) were logged by the external researchers, averaging 14 minutes per surveyed student across all pre-K and kindergarten classes. An additional 24.5 hours were logged for scheduling, logistics, and communication with campus contacts required to prepare for the survey administrations.

Consensus among the facilitators was that kindergarten students were able to complete the survey successfully with support. They reported that students appeared to understand the questions and knew the questions were pertaining to their homeroom teacher. They did express concerns with students being interested in each others' answers and sharing their answers out loud during the administration, which may have swayed their peers. Each of the facilitators noted that smaller groups of about 3 students expedited the kindergarten administration process, and limited behavioral disruptions. The facilitators were less confident that pre-K responses were valid and/or reliable. They reported that "while some [pre-K] students seemed to understand the questions, others clearly did not", and noted that pre-K students were "easily distracted." In many instances, questions had to be "repeated several times" and students had difficulty with some of the item wording.

Total reported irregularities on the feedback forms for the pre-K and kindergarten students were fairly consistent with other elementary grade levels. However, the percentage of administrations with reported irregularities that "possibly compromised student responses" were lower at the pre-K and kindergarten level compared to other elementary grades (Appendix D), suggesting the small group administrations successfully helped survey administrators maintain control over the administration environment, and allowed them to provide more individualized attention to student needs, than is possible with whole class

administrations.

Between 4% and 5% of the pre-K and kindergarten administrations included some irregularity that "possibly compromised student responses." The most commonly cited irregularities included, "one or more students unable to follow along" and "students did not understand questions." It was further reported that an additional 20% of pre-K administrations and 17% of kindergarten survey administration encountered some type of irregularity that did not disrupt the evaluation or compromise student responses.

Grades 1 and 2. A summary of participants can be found in Table 8. The survey instrument for 1st and 2nd grade students included a subset of 11 items from the survey administered to students in grades 3 through 12. They responded using a 4-point scale, ranging from *a lot of the time* to *never* and also were given a *don't know* option. To provide some support for students in choosing a response option, different size clocks were used to depict the amounts of time. The surveys were conducted online in whole groups by class, and sessions were scheduled and facilitated by campus staff. Teachers did not administer surveys to their own students.

The facilitators were responsible for reading the instructions and items aloud, guiding students through the online survey, and answering any questions about the meaning of words. Students were instructed to select *don't know* in cases in which they did not understand the question or had no opinion. In total, 880 1st- and 2nd-grade students provided feedback about 53 teachers.

Table 8. First and Second Grade Student Response Survey Participation, 2012–2013

Grade level	Survey administrations	Student surveys completed	Participating students	Participation rate	Rated teachers	Homeroom teacher coverage rate
First grade	27	463	463	94%	27	100%
Second grade	26	417	417	95%	26	100%

Source. 2012–2013 Student Response Survey

Administration times ranged from 20 minutes to 45 minutes, with an average time of 36 minutes. Average administration times for 1st-grade students (38 minutes) were slightly longer than for 2nd-grade students (34.5 minutes), most likely due to greater need for reading and technical support at the 1st-grade level. Campus contacts reported relatively few problems with the 1st-grade and 2nd-grade survey administrations. However, 1st- through 2nd-grade survey administration forms (n = 28) collected from four of the six participating elementary campuses revealed wide variations in survey administration irregularities reported by facilitators. Nearly half of the reported irregularities indicated that "one or more students were unable to follow along" (n = 5) or "students did not understand questions" (n = 7). Specific concerns noted by survey administrators included students not being familiar with the vocabulary (e.g., frequently, *desafiantes*); students' lack of computer skills (e.g., manipulating the mouse, closing the survey inadvertently); and students' behavior (e.g., excessive talking, sharing answers out-loud, and not following along with the pacing of the class). One administrator specifically noted that there were "too many students for one administrator."

Grades 3 through 5. A summary of participants can be found in Table 9. The instrument for students in grades 3 through 5 included 21 questions, and the response scale included four response options, ranging

from *a lot of the time* to *never*, as well as a *don't know* option. Each of the 3rd- through 5th-grade students completed two surveys: one about their homeroom teacher and one about either their art, music, or physical education teacher. The survey was administered online in a whole group setting for each class, and was facilitated by designated campus staff other than the classroom teacher. Students were allowed to complete the survey at their own pace after receiving a brief orientation and instructions provided by the survey facilitator. Survey facilitators guided students through the online survey and provided individualized reading support, as needed. A total of 1,098 3rd- through 5th-grade students participated, providing feedback about 65 3rd- through 5th-grade homeroom and special areas teachers.

Table 9. Grades 3 Through 5 Student Response Survey Participation, 2012–2013

Grade level	Survey administrations	Student surveys completed	Participating students	Student participation rate	Homeroom teacher coverage rate
Third grade	23	780	390	94%	100%
Fourth grade	23	760	380	89%	100%
Fifth grade	19	656	328	92%	100%

Source. 2012–2013 Student Response Survey
Note. Life skills and resource teacher with fewer than 10 students were excluded from the teacher sample. Each

Survey administration times ranged from 27 minutes to 45 minutes, averaging roughly 37 minutes across all 3rd- through 5th-grade administrations. Analysis of administration times revealed little discernible difference across grade levels; however, large differences were observed between schools. The level of facilitators' familiarity with the online survey; the degree to which survey facilitators followed the administration procedures and used the script; and inaccurate or inconsistent reporting of administration times may have contributed to the observed variation across schools.

Overall, elementary campus contacts and survey administrators reported that the 3rd- through 5th-grade students were able to complete the survey with minimal support, and indicated a high level of confidence in the students' comprehension of survey items and understanding of the response scale. Some irregularities were reported, including some behavioral issues (e.g., excessive talking, students not taking seriously, and sharing answers), which may have compromised students' responses during some survey administrations. Reported irregularities possibly compromising students' responses ranged from 0% to 40% across the four schools completing the irregularity section of the administration feedback form.

Secondary Administrations

A summary of participation can be found in Table 10. Responses from 6,102 middle and high school students, including 17,959 individual surveys, provided feedback about 484 secondary teachers and 1,421 classes. The survey was administered on paper during a class period of the school's choosing. All six secondary schools initially elected to administer the survey during their respective advisory periods, but one middle school later requested to switch the administration to a regular 8th-period class, due to advisory period scheduling changes enacted toward the end of the year. Administration packets containing survey protocols and instructions, student rosters, survey scan forms, a survey collection envelope, and an official DRE security seal were distributed to each of the survey administrators. Students each received a survey booklet containing three scan forms each, with a detachable, perforated strip identifying the student at top of the survey. Survey administrators instructed students to remove their

names before submitting their surveys to ensure their confidentiality, and a student was selected to collect completed forms from his or her peers and to seal the survey collection envelope. Eighty-nine percent of the returned survey collection envelopes (n = 402) were properly sealed.

Table 10. Secondary Student Response Survey Participation, 2012-2013

Grade level	Student participation				Teacher coverage		
	Sampled students	Student participants	Participation rate	Adjusted rate minus leavers	Teachers assigned to campus	Rated teachers	Teacher coverage rate
All Secondary	7,342	6,102	83%	86%	582	484	83%

Source. 2012–2013 Student Response Survey; AISD human resources records

Note. The sample of students included special education students who were later determined to be unable to respond to the survey in its present form.

The adjusted student rate excludes students who left their school in May and were therefore not enrolled during the survey administration period.

The teacher rates reflect the absence of both the number to teachers who did not meet the 10 student/30 record criteria and any teachers included in the sample who did not receive at least 10 student ratings.

Overall, 96% of advisory or other class period teachers administered the survey to their students (Table 11), a significant increase from the 75% to 80% administration rate in 2011–2012. This improvement was due in part to the increased communication efforts with campus contacts throughout the survey preparation process, but was primarily due to the increased stakes associated with the survey results because the survey was included in the teacher appraisal. Eighteen teachers did not complete the survey administrations; 15 were life-skills or resource class teachers and three did not return surveys for unknown reasons. No evidence suggests any intentional effort to circumvent the survey administration process on the part of any of the facilitators, or to systematically deny any student participation.

Table 11. Survey Administration Rates

	Martin MS	Webb MS	Akins HS	Lanier HS	Travis HS	Reagan HS	All secondary
Scheduled survey administrations	39	51	160	98	64	58	470
Completed survey administrations	38	50	156	95	58	55	452
Survey administration rate	97%	98%	98%	97%	91%	95%	96%

Source. 2012–2013 Student Response Survey

In cases in which the life-skills and resource special education teachers did administer the survey, the facilitating teachers reported students were unable to meaningfully participate even when provided assistance. Two life-skills teachers attempted to administer adapted versions of the pre-K and kindergarten survey using Boardmaker Share (i.e., a program that creates picture symbols), with which many life-skills students are familiar (Appendix F). However, both teachers reported their students were unable to respond.

Participation rates of secondary students were affected by various factors, including work-study programs in which students were officially enrolled in an advisory period on campus but were off site working during the survey administration; general absenteeism, which tends to be higher toward the end

of the year (approximately 8% to 10% during last 6 weeks of school at the high school level); and decisions to not survey some students (i.e., life-skills and resource students, English language learners). In a few instances, some students reportedly elected not to participate, although this did not appear to be widespread. Of the original 7,342 secondary students identified to participate in the Student Response Survey, 6,102 completed one or more surveys, representing an 83% student participation rate overall, and an 86% participation rate when accounting for students who were no longer enrolled at the school at the time of the survey administration (leavers).

In general, secondary campus contacts reported that the administration process went well, with a few exceptions. For example, at one high school, a substitute teacher instructed students to take the surveys home to complete, and another substitute teacher at a different campus lost the survey collection envelope. Campus contacts noted these difficulties and reported that although they had attempted to ensure that substitute teachers were adequately informed about the survey administration procedures and purpose, they doubted the substitutes had sufficient time to read through the provided instructions and procedural guidelines prior to administering the surveys. In other instances, teachers reportedly had to survey their students over two days because the students did not have time to complete all three surveys during the regular period. Overall, administration times ranged from 15 minutes to 60 minutes, averaging slightly under 30 minutes. Only 14% of the administrations exceeded the desirable 30-minute range, and in most instances, students were able to complete the surveys during a regularly scheduled class period at both the middle and high school levels, with few disruptions.

Administration feedback form completion rates varied substantially across campuses, ranging from a low of 51% to a high of 80%, and averaging 62% across all secondary campuses. Although limited, the feedback provided suggests irregularities were relatively uncommon across campuses. Overall, 90% of the administration forms received indicated that "no irregularities [were] encountered," while 3% indicated that an irregularity occurred that "possibly compromised student responses." The most common reported irregularity included "the students' teacher was present during the administration" (n = 23).[4] In addition, "students did not understand the questions" (n = 9), and "one or more students unable to follow along" (n = 7) applied exclusively to survey administrations with life-skills and resource students. "Other" reported irregularities reported (n = 21) are summarized in Table 12.

Table 12. Secondary Student Response Survey Administration Irregularities

	Martin MS (n=16)	Akins HS (n=69)	Lanier HS (n=60)	Reagan HS (n=32)	Travis HS (n=42)	All secondary (n=219)
Administration feedback form completion rate	54%	51%	65%	71%	80%	62%
No irregularities encountered	88%	91%	93%	88%	88%	90%
Irregularities encountered, did not compromise student responses	6%	7%	3%	6%	10%	6%
Irregularities encountered, possibly compromised student responses	6%	1%	3%	6%	2%	3%

Source. 2012–2013 Student Response Survey
Note. Survey administration forms were not provided to Webb because they administered the survey early.

4. Due to the sampling method applied, it was expected that some students would respond about a classroom teacher who was also their advisory teacher. Additional procedures including students removing their names from their survey forms and students collecting completed forms were enacted to ensure confidentiality at the secondary level.

Teacher coverage. The teacher coverage and class coverage measures provide a way to assess the Student Response Survey's overall reach, inclusiveness, and representativeness. Teacher coverage specifically measures the percentage of eligible instructional staff that were included in the Student Response Survey sample. The measure is particularly valuable for evaluating the extent to which the current sampling criteria captured teachers rated under the new appraisal system. Teacher coverage rates were affected by the courses selected for rating by campus administrators and by the teacher sampling criteria established by DRE researchers. In total, 63 teachers were removed from the sample because they were assigned fewer than 10 students, or had less than 30 total student records; 23 of these were resource, inclusion, or life-skills teachers. An additional seven teachers and 13 staff members (i.e. counselors, campus administrators) were removed because campus administrators omitted their courses from being rated.

Class coverage. Class coverage measures the percentage of a teacher's courses that were captured in the student response survey sample. The measure provides a way to assess the extent to which a teacher's scores incorporate feedback across all of the their teaching assignments. Teachers' strengths, teaching styles, and student needs vary from course to course, making it important to capture a full representation of their work in any appraisal system. Secondary students rated 1,421 courses; one hundred and thirty-five courses were removed from the list given their non-instructional nature. Across all secondary schools, teachers received feedback from an average of 34% of their total students across all of their courses taught.

Analyses of Survey Properties

As detailed previously, the extensive revision and pilot process yielded substantial changes to the Student Response Survey instruments used in 2012–2013. Four final versions of the survey were used (Table 5), and although the concepts measured were consistent across versions, item wording varied somewhat across versions; the online versions that were administered to pre-K through grade 5 were programmed so the teacher's name was inserted into each question stem (see Table 2 for the wording in each of the survey versions). The goals of the present analysis were to assess the reliability and factor structures of these four instruments and to examine the extent to which student ratings correlated with other observational ratings of the same teacher. The tables that follow contain descriptive results as well as the results of both reliability and factor analyses.

Item Means and Distribution of Scores

Mean responses for each item by survey version and the percentage of students who selected each response option can be found in Tables 13 and 14. Across versions, ratings were generally very favorable. Item means on the 4-point scale ranged from 2.77 to 3.63 at the secondary level, 2.91 to 3.70 for grades 3 through 5, and 3.03 to 3.56 for grades 1 and 2. For pre-K and kindergarten, means on the 3-point scale ranged from 2.44 to 2.69. For all items on all versions, no more than 10% of students, and in most cases far fewer, selected *don't know*. This provides some preliminary evidence that the questions were clear and relevant. Younger students did not choose *don't know* any more frequently than did older students. Also, similarities occurred in the way students responded to the survey items across versions. Students gave their highest ratings to items related to teachers' respect for students and checking for understanding; they gave their lowest ratings to items related to students celebrating when someone does a good job and students behaving the way the teacher wants them to.

Table 13. Means and Response Percentages for Each Item, Grades Prekindergarten (Pre-K) and Kindergarten

Survey Items	Percentage of students				Mean
	A lot of the time (3)	Sometimes (2)	Never (1)	Don't know	
1. How often does ____ help you learn new things?	69%	21%	4%	6%	2.69
2. How often do ___'s students stay busy working?	61%	28%	5%	6%	2.60
3. How often do students in your class behave the way __ wants them to?	52%	28%	8%	8%	2.44
4. How often do students in your class treat __ with respect?	55%	28%	11%	9%	2.51
5. How often does __ show respect to all students?	67%	21%	5%	6%	2.66

Source. 2012–2013 Student Response Survey

Note. Means for pre-K and kindergarten version are out of 3 points.

Responses of "don't know" were not included in the mean computation.

Table 14. Means and Response Percentages for Each Item, by Survey Version

Survey items	Response percentages						
	Survey version	A lot of the time	Some-times (3)	A little of the time (2)	Never (1)	Don't know	Mean
This teacher asks us to share our ideas with each other.	1 & 2	30%	41%	16%	3%	10%	3.10
	3-5	36%	37%	12%	10%	5%	3.05
	6-12	48%	32%	9%	7%	4%	3.27
We get to do a lot in this class, not just listen to the teacher talk.	3-5	41%	33%	12%	5%	9%	3.23
	6-12	58%	27%	9%	4%	3%	3.42
This teacher has several good ways to explain each topic so I understand.	3-5	68%	21%	5%	3%	3%	3.60
	6-12	59%	26%	9%	4%	3%	3.43
This teacher checks to make sure we understand what we are learning.	1 & 2	53%	26%	10%	1%	10%	3.44
	3-5	72%	19%	4%	3%	3%	3.64
	6-12	64%	23%	8%	4%	3%	3.50
This teacher uses different ways to help us learn.	1 & 2	45%	26%	14%	6%	9%	3.22
	3-5	67%	22%	5%	3%	4%	3.58
	6-12	55%	28%	9%	4%	4%	3.39
This teacher explains difficult things clearly.	3-5	55%	30%	6%	3%	6%	3.45
	6-12	52%	30%	10%	4%	4%	3.35
This teacher helps me think about things in new ways.	3-5	58%	29%	6%	2%	4%	3.49
	6-12	47%	31%	11%	6%	4%	3.25
This teacher asks me to explain why I think what I think.	1 & 2	34%	34%	16%	5%	11%	3.10
	3-5	46%	32%	8%	5%	8%	3.30
	6-12	42%	31%	14%	9%	5%	3.11

Distribution of scores. The distribution of scores was examined in two ways. First, the distribution of

Table 14. Means and Response Percentages for Each Item, by Survey Version (continued)

Survey items	Survey version	Response percentages: A lot of the time (4)	Some-times (3)	A little of the time (2)	Never (1)	Don't know	Mean
This teacher gives me challenging work.	1 & 2	36%	36%	15%	6%	6%	3.10
	3-5	31%	43%	13%	7%	6%	3.05
	6-12	43%	37%	10%	5%	4%	3.23
This teacher shows me why this class is important.	3-5	65%	17%	6%	4%	8%	3.56
	6-12	55%	26%	9%	6%	5%	3.36
Everybody knows what they should be learning in this class.	3-5	54%	32%	7%	2%	6%	3.46
	6-12	54%	28%	9%	3%	6%	3.41
This teacher helps me understand why my answers are right or wrong.	1 & 2	40%	30%	17%	5%	8%	3.14
	3-5	61%	24%	6%	4%	5%	3.49
	6-12	56%	26%	9%	5%	4%	3.38
This class stays busy and does not waste time.	1 & 2	52%	26%	13%	4%	6%	3.34
	3-5	30%	44%	16%	5%	5%	3.03
	6-12	48%	33%	11%	5%	4%	3.29
Everybody knows what they should be doing in this class.	3-5	46%	37%	8%	3%	6%	3.34
	6-12	53%	30%	9%	3%	5%	3.40
This teacher allows students to break the rules.*	3-5	9%	6%	6%	72%	7%	3.52
	6-12	9%	10%	11%	61%	9%	3.36
My classmates behave the way the teacher wants them to.	1 & 2	33%	35%	17%	7%	8%	3.03
	3-5	32%	44%	14%	4%	6%	3.10
	6-12	38%	39%	13%	5%	5%	3.16
This teacher shows us examples of high quality work.	1 & 2	54%	27%	11%	2%	5%	3.41
	3-5	59%	24%	6%	3%	8%	3.50
	6-12	51%	30%	9%	5%	6%	3.35
I feel comfortable talking about my ideas in this class.	3-5	48%	28%	12%	6%	6%	3.26
	6-12	46%	28%	12%	8%	6%	3.19
This teacher shows respect to all students.	1 & 2	64%	21%	6%	3%	7%	3.56
	3-5	77%	12%	5%	2%	4%	3.70
	6-12	72%	16%	5%	3%	4%	3.63
Students in this class celebrate when someone does a good job.	3-5	31%	34%	14%	13%	10%	2.91
	6-12	30%	26%	16%	17%	10%	2.77
Students in this class treat the teacher with respect.	1 & 2	47%	30%	13%	4%	6%	3.27
	3-5	46%	34%	11%	3%	6%	3.30
	6-12	55%	29%	8%	4%	5%	3.41
Overall mean	1 & 2						3.23
	3-5						3.36
	6-12						3.31

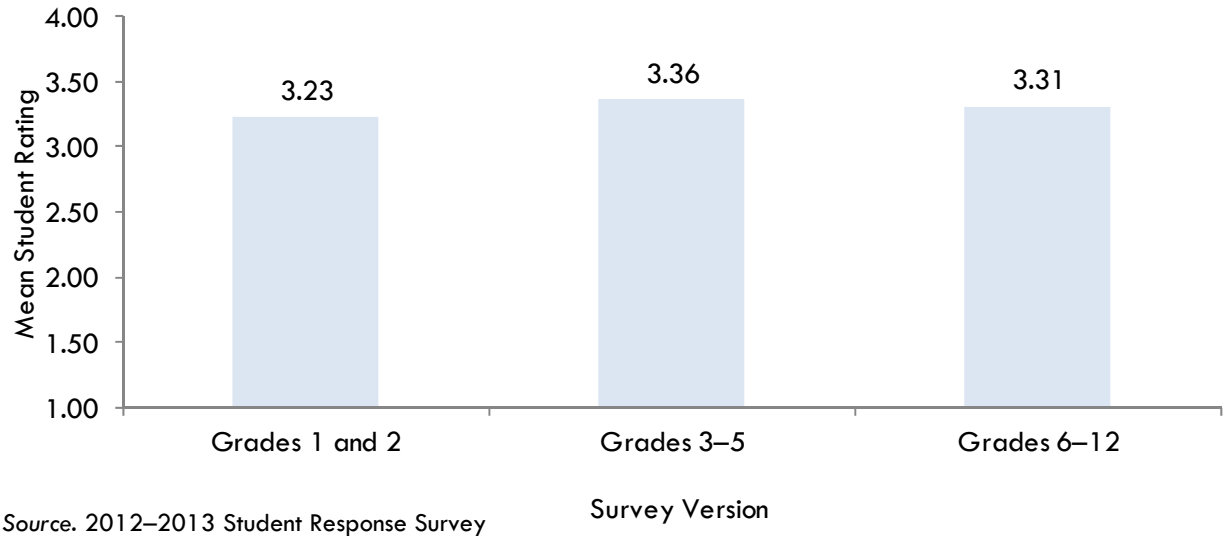
Source. 2012–2013 Student Response Survey

Note. For the alternate wording for grades 1 and 2 and grades 3 to 5 versions, see Table 2.

*The average for this item was reverse-coded (e.g.,1= *a lot of the time* and 4= *never*); therefore, higher numbers are desirable

overall ratings *given by students* are presented in Figures 2 and 3. Significant differences were found among all the average overall student ratings across the survey versions. Students in grades 3 through 5 gave the highest ratings, followed by students in grades 6 through 12; students in grades 1 and 2 gave ratings lower than both students in grades 3 through 5 and student in grades 6 through 12. The rating distributions also were slightly different across elementary and secondary versions. Although the means were nearly identical (i.e., all elementary, 3.32; all secondary, 3.31), the distribution of scores varied by survey version.

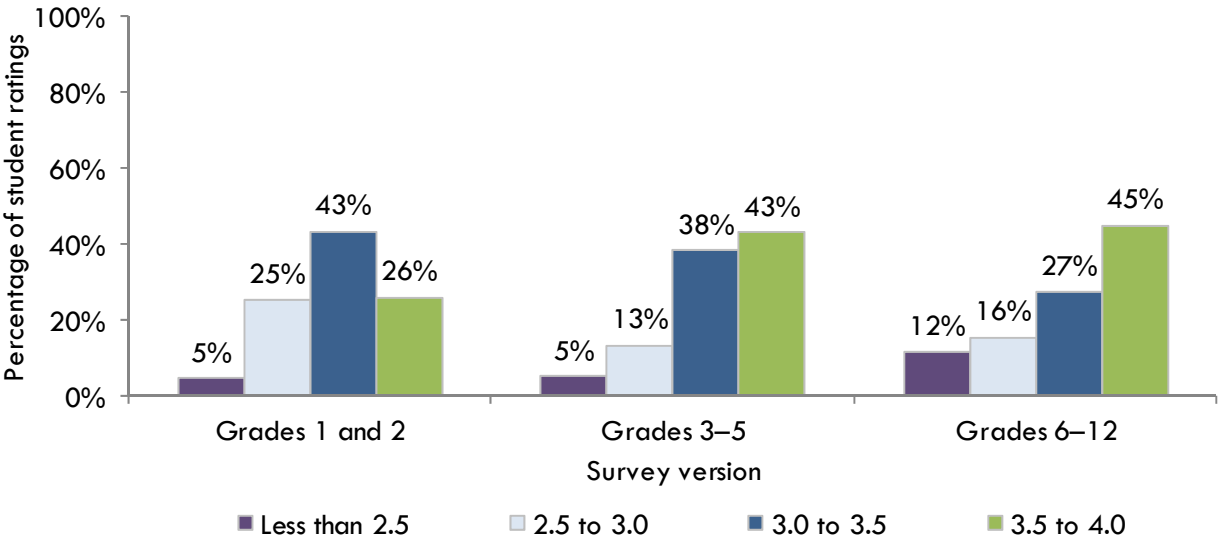
Figure 2. Mean Ratings Given by Students, by Survey Version



Source. 2012–2013 Student Response Survey

Note. $F = 13.99$; $p < .01$; All means were different from each other at the $p<.05$ level.

Figure 3. Distribution of Mean Ratings Given by Students, by Survey Version



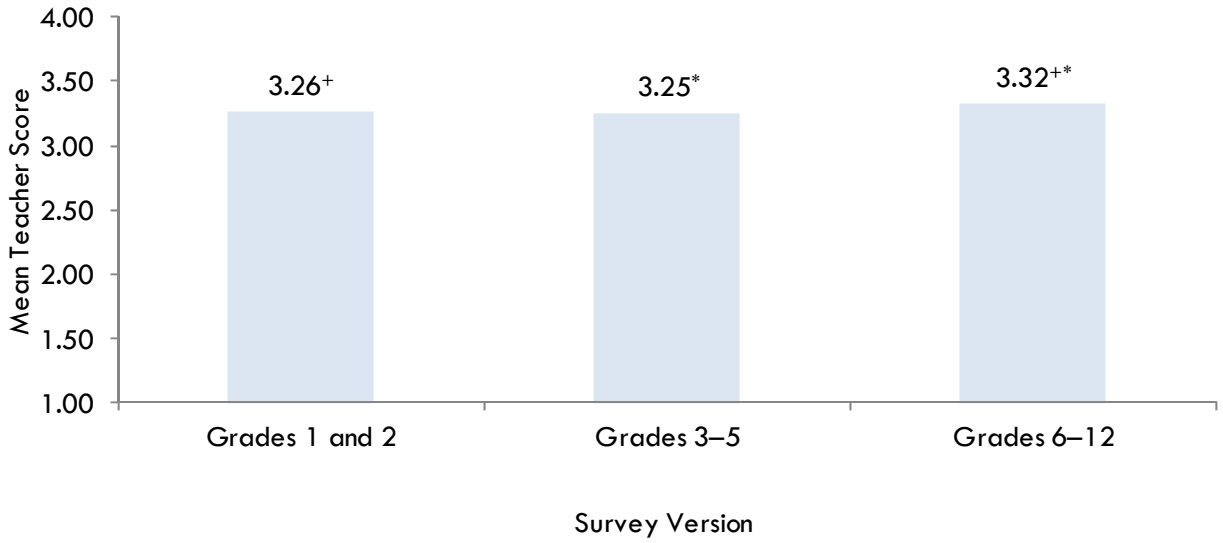
Source. 2012–2013 Student Response Survey

Note. Total scores were not computed for pre-K and kindergarten because this survey was not included in the appraisal for those teachers in 2012–2013.

Second, the distribution of scores *received by teachers* was examined (Figures 4 and 5). The overall means that teachers received were different for each survey version; secondary teachers received the highest scores (3.32), followed by teachers in grades 1 and 2 (3.26) and teachers in grades 3 through 5 (3.25). It is interesting to note that no teacher in grades 3 to 5 received a score lower than 3.0.

Survey Reliability and Factor Structure

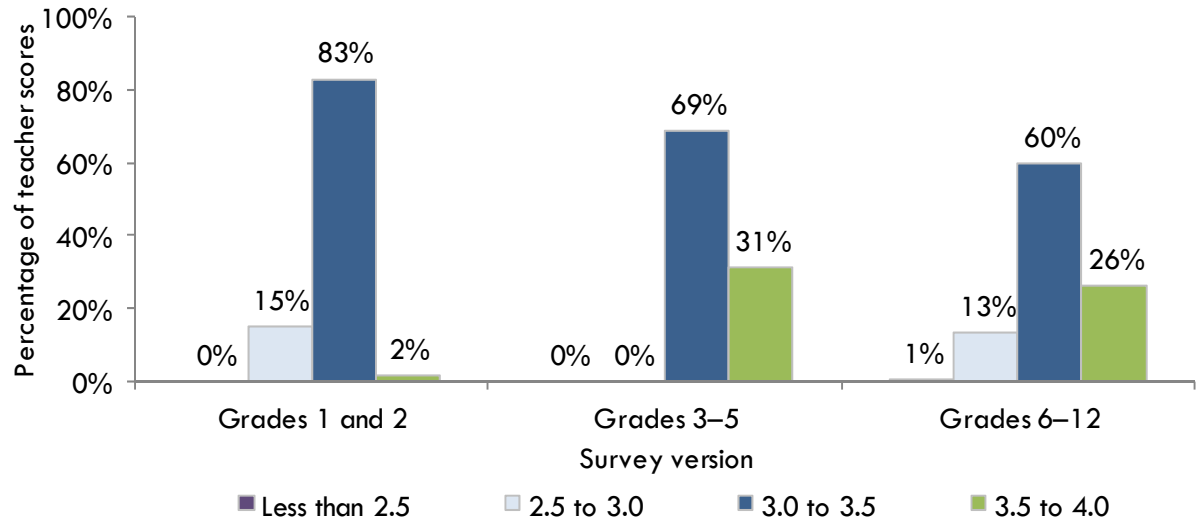
Figure 4. Mean Scores Received by Teachers, by Survey Version



Source. 2012–2013 Student Response Survey

Note. $F = 8.13$; $p < .01$; Means with the same superscripts are significantly different at $p \leq .05$

Figure 5. Distribution of Mean Scores Received by Teachers, by Survey Version



Source. 2012–2013 Student Response Survey

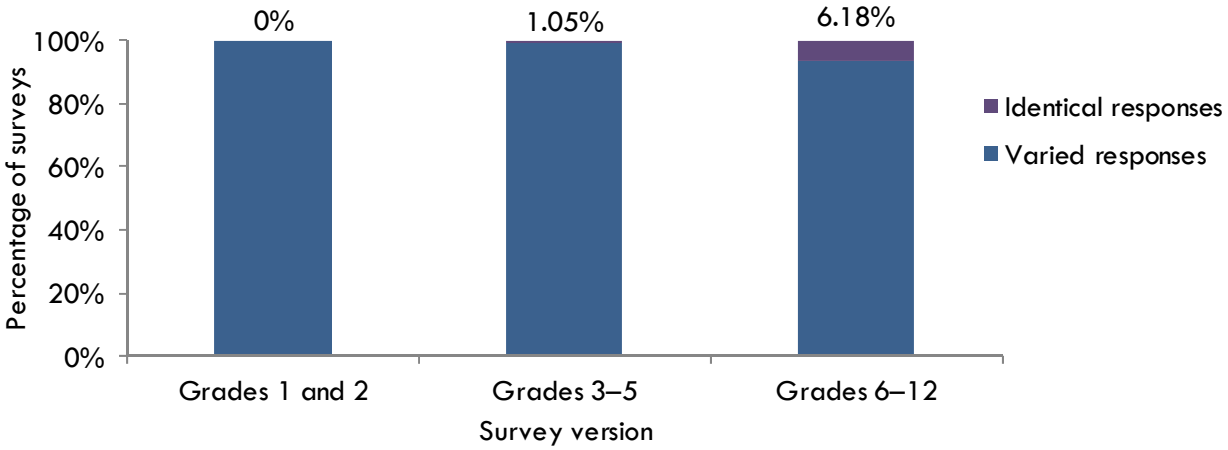
Note. Total scores were not computed for pre-K and kindergarten teachers because the survey was not included in their appraisal in 2012-2013.

To address the quality of the survey instruments, response patterns, reliabilities, and factor structures were examined for all survey versions.

Response patterns. To assess the extent to which students responded thoughtfully to the survey items, the percentage of students who gave the same response for every item was examined. The survey included one negatively worded item; therefore, answering all questions with the same response may be an indication of a less mindful response. Analyses were conducted for all versions except the pre-K and kindergarten instrument due to limitations in the data (i.e., the survey had only five items and no negatively worded items). Figure 6 displays the percentages of surveys with the same response for every item for each version of the survey. The percentage of problematic responses was very low for all versions, and this provides some evidence to suggest that students read and responded to each item.

Reliability. Results indicate that the reliability of the survey versions for both grades 3 through 5 (α = .89) and grades 6 through 12 (α = .96) were high, but that the versions for grades 1 and 2 (α = .64) and pre-K and kindergarten (α = .59) were much less internally consistent (Table 15). When examined for each grade, the responses of kindergarten students (α = .54) were less consistent than the responses given by pre-K students (α = .60). This may be due in part to the different administration methods; pre-K students answered the items one on one with a facilitator using tablets, but kindergarten students took the survey in small groups (approximately five students). Analyses of the Student Response Survey instruments used in 2011–2012 yielded similar, although slightly better, results.

Figure 6. Percentage of Surveys With Identical Responses for All Items, by Survey Version



Source. 2012–2013 Student Response Survey

Factor analyses. Table 16 displays the results of the present factor analyses. Consistent with the previous analyses, the results for the grade 3 through 5 version and the grade 6 through 12 version indicate that

Table 15. Reliability Coefficients, by Survey Version and Grade Level

	Survey version							
	Prekindergarten /kindergarten			Grades 1 and 2			Grades 3–5	Grades 6–12
	Pre-K	Kindergarten	All	1	2	All		
Cronbach's α	.60	.54	.58	.64	.59	.64	.89	.96

Source. 2012–2013 Student Response Survey

the survey measures two clear factors: one focused on teacher behavior (factor 1) and the other focused on student behavior (factor 2). Reliabilities for each of these factors were moderate to high for both the 3 through 5 version and the 6 through 12 version of the survey (Table 16).

Several items, however, did not load on either factor for the either the grades 3 through 5 or grades 6 through 12 version. The items "This teacher gives me challenging work" and "Students in this class

Table 16. Rotated Factor Solution for All Items, by Survey Version

	Survey version grade level								
	Grades 1 and 2			Grades 3–5			Grades 6–12		
Survey item Factor	1	2	3	1	2		1	2	
1. This teacher asks us to share our ideas with each other.			0.72	0.61			0.70		
2. We get to do a lot in this class, not just listen to the teacher talk.				-	-	-	0.59		
3. This teacher has several good ways to explain each topic so I understand.				0.71			0.79		
4. This teacher checks to make sure we understand what we are learning.			0.50	0.74			0.78		
5. This teacher uses different ways to help us learn.	-	-	-	0.74			0.79		
6. This teacher explains difficult things clearly.				0.61			0.77		
7. This teacher helps me think about things in new ways.				0.69			0.78		
8. This teacher asks me to explain why I think what I think.	0.59			0.67			0.74		
9. This teacher gives me challenging work.	-	-	-	-	-	-	-	-	-
10. This teacher shows me why this class is important.				0.59			0.69		
11. Everybody knows what they should be learning in this class.					0.62			0.61	
12. This teacher helps me understand why my answers are right or wrong.	0.73			0.70			0.75		
13. This class stays busy and does not waste time.			0.51		0.66			0.73	
14. Everybody knows what they should be doing in this class.					0.69			0.69	
15. This teacher allows students to break the rules.*						0.74			0.85
16. My classmates behave the way the teacher wants them to.		0.71			0.77			0.80	
17. This teacher shows us examples of high quality work.	0.62			0.62			0.64		
18. I feel comfortable talking about my ideas in this class.				0.50			0.62		
19. This teacher shows respect to all students.		0.58		0.53			0.57		
20. Students in this class celebrate when someone does a good job.				-	-	-	-	-	-
21. Students in this class treat the teacher with respect.		0.71			0.77			0.79	
Cronbach's α	.44	.49	.32	.88	.79	-	.94	.87	-

Source. 2012–2013 Student Response Survey

Note. The factor solution was produced using principal components analysis with varimax rotation. For simplicity item wording from the 6 through 12 version was used in this table. Please refer back to Table 2 for all wording variations.

celebrate when someone does a good job" were both problematic, and the only negatively worded item, "This teacher allows students to break the rules," failed to load on either factor for the second year in a row. The results for the grades 1 and 2 version and the pre-K and kindergarten version, however, were not consistent with the other versions.

The results for the grades 1 and 2 version suggest a lack of cohesiveness. There were three factors, and the items did not cluster together in a conceptually clear way. In addition, the analyses of the pre-K and kindergarten version produced two different factor solutions (i.e., one for pre-K and one for Kindergarten) for the same five items. At pre-K, the instrument appeared to be unidimensional, while at kindergarten, the items clustered into the teacher behavior and student behavior factors in the analyses described here (Table 17).

Taken together, the results of the reliability and factor analyses suggest that the grades 3 through 5 and grades 6 through 12 versions of the survey are internally consistent, and that they both tap into the same

Table 17. Rotated Factor Solution for All Items, Prekindergarten (Pre-K) and Kindergarten Version

		Grade level		
		Pre-K	Kindergarten	
Survey item	Factor	1	1	2
1. How often does _____ help you learn new things?		0.55		0.87
2. How often do ___'s students stay busy working?		0.59	0.58	
3. How often do students in your class behave the way __ wants them to?		0.63	0.75	
4. How often do students in your class treat __ with respect?		0.72	0.79	
5. How often does __ show respect to all students?		0.69		0.68
Cronbach's α		.59	.51	.35

Source. 2012–2013 Student Response Survey

Note. Factor solution was produced using principal components analysis with varimax rotation.

two distinct student observational factors: teacher behavior and student behavior. These factors are very consistent with the observational rubric domains that the survey was designed to measure. When compared with the results of the previous year's analyses, the results are consistent and suggest some degree of reliability. The versions used with younger students, however, were less internally consistent. Additionally, the analyses of the grades 1 and 2 version produced factors that were not clearly related to the constructs that the survey was designed to measure.

Relationships With Appraisal Components

The final analyses addressed the extent to which the students' ratings correlated with other observation-based ratings of the same teacher (Table 18). Given the concerns outlined previously about the surveys used for students in grades pre-K through 2, these analyses were limited to the grades 3 through 5 and grades 6 through 12 versions. To assess convergent validity, correlations were examined between a teacher's Student Response Survey means (overall, and for each of the two factors) and the observation scores that the teacher received from his or her administrator and peer observer. Because the survey was based on the same observational protocol that the administrators and peer observers used to produce their observation scores, the ratings were expected to be similar in some ways. The results indicate weak to moderate correlations between student ratings and the other observation-based ratings. Given that the student survey was based on many hours of observation rather than a single classroom visit, and that

it measured a more narrow set of constructs than the comprehensive rubric, these moderate correlations seem reasonable.

In addition, correlations between the Student Response survey results and the teacher's student learning objective (SLO) results were examined. SLOs are learning goals set by the teacher at the beginning of the school year. Teachers set one SLO for their own students (i.e., individual SLO) and either work with colleagues to set a goal for their entire group of students (i.e., team SLO) or set a second individual SLO. The SLO metric included here is the percentage of students who met each goal. If SLOs are a proxy for teacher effectiveness, and if the constructs measured on the observational rubric (and captured in the survey) are associated with effective teaching, it is reasonable to expect a correlation between SLO

Table 18. Correlations Among Appraisal Observer Ratings and Student Learning Objective Results, by Survey Version

			Student Response Survey			Teacher self-rating	Administrator observations			Peer observations		SLOs	
		Survey version	Overall mean	Instruct factor	Student factor		Walk 1	Walk 2	Formal	1	2	Ind % met	Team % met
Student Response Survey	Instruct Factor	3 to 5	.93**										
		6 to 12	.99**										
	Student Factor	3 to 5	.88**	.65**									
		6 to 12	.93**	.88**									
Teacher self-rating		6 to 12	.16	.13	.26**								
Administrator observations	Walk-through 1	3 to 5	-.06	-.04	-.08	—							
		6 to 12	.30**	.27**	.34**	.23*							
	Walk-through 2	3 to 5	.20	.22	.12	—	.48**						
		6 to 12	.25**	.21**	.33**	.27*	.66**						
	Formal	3 to 5	.41**	.41**	.33*	—	.31*	.50**					
		6 to 12	.20**	.17**	.27**	.19	.66**	.66**					
Peer observations	Obs 1	3 to 5	.29	.28	.23	—	.20	.28	.29				
		6 to 12	.21**	.18**	.27**	.22*	.29**	.38**	.41**				
	Obs 2	3 to 5	.31	.25	.32*	—	.21	.29	.32*	.41**			
		6 to 12	.27**	.24**	.34**	.16	.43**	.35**	.47**	.53**			
SLOs (% met)	Individual	3 to 5	.08	-.02	.19	—	.16	.02	-.05	.03	.05		
		6 to 12	.05	.02	.13*	.30**	.11	.17*	.01	.00	.06		
	Team/ Individual	3 to 5	.20	.16	.22	—	-.01	.07	.30	.25	.34*	.14	
		6 to 12	.06	.04	.11*	.38**	.08	.10	.09	.09	.14*	.44*	
EVAAS	Reading/ ELA	3 to 5	-.20	-.28	-.09	—	-.04	-.04	-.30	-.22	-.42	.01	-.55*
		6 to 12	.34*	.28	.46**	-.10	.23	.16	-.06	.12	.16	.42**	.19
	Math	3 to 5	.34	.33	.29	—	-.20	.24	.18	.36	.28	-.32	.38
		6 to 12	.36	.36	.33	-.23	.08	-.20	-.04	.31	.18	.25	.27

Source. 2012–2013 Student Response Survey; District student learning objectives, peer observation, and administrator observation databases; 2013 Employee Coordinated Survey; District EVAAS records

Note. The overall mean was computed without problematic items 9, 15, and 20

Grades 3 through 5 n = 37, grades 6 through 12 n = 278; teacher self-rating not available for grade 3-5 teacher due to small cell size; EVAAS results for Grades 3 through 5 Reading/ELA (n = 17), Math (n = 18); EVAAS results for Grades 6 through 12 Reading/ELA (n = 43), Math (n = 20).

*p<.05 **p<.01

results and student ratings. Results indicated few significant relationships between the results of student SLOs and student ratings.

Although statistically significant only for secondary students, the magnitude of the correlations was similar for grades 3 through 5 and grades 6 through12, and the relationships were weak. The strongest correlations between SLO results and student response survey results were for the overall mean and the student behavior factor. It also is notable that the correlations among all of the appraisal factors were weak to moderate, and that the Student Response Survey results were correlated with the administrator observations and peer observations at the same magnitude with which they were correlated with each other. In addition, the teachers' self-rating of the instructional strategies measured on the student response survey and included on the observational rubric also was examined. This metric was not included as part of the teacher appraisal, but offers a fourth perspective on the teachers' instructional practice. The self-ratings of teachers in grades 6 through 12 were weak to moderately correlated with several of the appraisal components, including the instructional factor of the Student Response Survey ($r = .26$).

Finally, teacher reading/ELA and mathematics (math) value-added data (i.e., Education Value-Added Assessment System [EVAAS]) were examined to assess the extent to which the results of the Student Response Survey for a teacher were related to the academic growth of his/her students. In general, EVAAS results were not strongly related to any of the appraisal components, including the Student Response Survey. However, secondary reading/ELA EVAAS results were significantly correlated with the Student Response Survey overall score ($r = .32$; $p < .05$) and to the student behavior subscale score ($r = .46$; $p < .01$), suggesting that there is a relationship between students' perceptions of their classroom, particularly perceptions of student behavior, and their academic growth. Secondary reading/ELA EVAAS results also were correlated with the percentage of students who met that teacher's SLO ($r = .42$; $p < .01$). This provides some validation that there is an association between growth as demonstrated by the SLO and growth as measured by EVAAS. The magnitude of the elementary and secondary correlations between math EVAAS and the Student Response Survey results was similar to that for secondary reading/ELA, though not statistically significant (likely due to small cell sizes). In addition, correlations between elementary reading/ELA EVAAS and the other appraisal components were small but negative. Future research should continue to address this unexpected relationship.

Fiscal Considerations

The official personnel and material costs were estimated to be $51,331.19 (Table 19), which was $5.70 per student or $75.16 per teacher (Table 20). However, several issues should be considered when assessing the value of the survey results in relation to the cost. First, the cost per unit for the pre-K and kindergarten surveys were much higher than were the surveys for students in other grades, due to the

Table 19. Estimated Student Response Survey Expenditures

Source	Description	Amount
Staff salaries	.50 FTE (salary+benefits) for 10 months; +100 hours	$36,094.10
Paper surveys	Custom printed survey forms for grades 6 to 12	$5,742.09
Contractors	Fee for administration of pre-K/kindergarten survey	$7,995.00
Online survey tool	Cost of contract for online survey tool ($187.50/survey)	$1,500.00
Total		$51,331.19

Source. District financial records

individual and small-group administration. Second, the cost of DRE staff time to develop and pilot the survey instruments is not an anticipated future cost. Table 21 displays only the material and facilitator costs for each survey version. The materials and facilitation costs per student for the pre-K and kindergarten survey was nearly $8 higher than for the students in grades 1 through 12 and more than 10 times higher per teacher for the pre-K and kindergarten teachers than for students in other grades.

Table 20. Estimated Cost per Participant, by Survey Version

Participants	Pre-K/ Kindergarten	1 through 5	6 through12	Total	Cost per participant
Students	920	1,098	6,102	9,000	$5.70
Teachers	55	144	484	683	$75.16

Source. District financial records

Table 21. Materials and Facilitator Contract Costs Only, by Survey Version

	Pre-K/Kindergarten		1 through 5		6 through 12	
	Participants	Cost per	Participants	Cost per	Participants	Cost per
Students	920	$8.69	1,978	$0.75	6,102	$0.94
Teachers	55	$145.36	144	$10.41	484	$11.86
		$7,995.00		$1,500.00		$5,742.09

Source. District financial records

Table 22. Summary of Time Estimates for Campus Contacts and Survey

	Survey Scheduling	Survey Administration		
Activity / grade level	Estimated time for campus contacts	Average survey administration times	Total number of administrations	Approximate total staff time
Training	1 hour			
Roster verification	3 hours			
Scheduling and coordination	5 hours			
Pre-K		12 minutes/student	450	90 hours
Kindergarten		16 minutes/5 students	94	25 hours
First		38 minutes/class	27	17 hours
Second		35 minutes/class	26	16 hours
Third		38 minutes/class	23	15 hours
Fourth		36 minutes/class	23	14 hours
Fifth		39 minutes/class	19	12 hours
Secondary		28 minutes/class	452	211 hours
Total	9 hours/contact=108		1,114	400 hours

Source. Administration feedback forms and campus contact debriefing interviews.

In addition to financial costs, the cost of teacher time also should be considered. The campus contacts spent an estimated 9 hours each working on the preparation and administration of the survey, although some secondary schools with particularly challenging master schedules required more time and support than these estimates capture. In addition, the average administration times are displayed in Table 22. In some cases, these averages may be underestimated, particularly in cases where technology was shared across classes and logins had to be reset between sessions.

Conclusions, Recommendations, and Implications for Policy

Several key findings emerged from the revision, pilot, administration, and analysis of the Student Response Survey in 2012–2013 that will inform this work as it moves forward. The following section addresses the most critical of these and proposes recommendations for policy changes or other actions.

Which students and teachers should be included in the survey?

A significant focus of the pre-pilot revision process was the adaptation of the Student Response Survey instrument for multiple student groups at different reading ability levels. This was important both to bring more student voices into the appraisal and to provide more teachers with students' feedback. For pre-K and kindergarten students, a five-item version of the survey was developed in consultation with district pre-K and kindergarten teachers and external consultants with expertise in early childhood education[4] and survey administration. Finally, the instrument for students in grades 1 and 2 was revised to simplify the language and to personalize the instrument by including the name of the teacher in the questions.

In spite of these efforts, the feedback received from the contractors who administered the survey to pre-K and kindergarten students, and the results of the item analyses, raised questions about the validity of these data. In particular, the concerns raised by the survey administrators about students' comprehension were significant. Also, the cost of the administration in terms of time and money were greatest for the teachers of younger students, further suggesting that it may not be feasible to continue to survey their students, nor reasonable to include the results of those surveys in their appraisal scores. The results of the analyses of the grades 1 and 2 data also are cause for concern. The instrument lacks internal consistency and the mean scores showed very little variation; 83% of teachers received a mean between 3.0 and 3.5. The consequences of these issues are twofold. First, the way in which students responded across items was not consistent, which was likely due to a combination of (a) the wide variation in younger students' ability to gauge subtle distinctions between the response scale options and (b) the 11-item subset of questions not tapping into a common construct. Second, because of the lack of variation in scores, the Student Response Survey results did not contribute to the aim of the pilot appraisal to develop a system that differentiates among teachers to better provide targeted support and professional development. Recently the Georgia Department of Education opted not to include children from Grades 2 and below in their student survey. "Based on [our pilot program] on how to implement and get reliable data, the K-2 survey really was not feasible," says Teresa MacCartney, Deputy Superintendent of Race to the Top Implementation. Vanderbuilt University researcher Ryan Balch, who developed and administered the survey in seven districts in Georgia recommends limiting student surveys to Grades 4 and higher, suggesting that that there are questions about surveying very young students who cannot read (Kloberdanz, 2012).

Recommendation 1: Exclude Student Response Survey data from the teacher appraisal

4. Experts were contracted during survey development in 2011-2012. See Schmitt (2012) for details.

for teachers in pre-K through grade 2 until such time as the survey can be conducted more efficiently and the data produced are more reliable, consistent, and varied.

Similarly, the experiences of special education teachers who administered or attempted to administer the survey to their students varied widely. Moreover, more than half of special education teachers were excluded from the Student Response Survey teacher sample because they worked with fewer than 10 students. Given the variety of needs represented by special education students, and the variation in the delivery of instruction they receive, it would be extremely difficult to develop a one-size-fits-all survey and survey administration process that provides a reliable or valid measure of a special education student's experience with his or her teacher. Each of the teachers who attempted to administer the survey to his or her life-skills and resource students reported an overall inability of the students to respond meaningfully. Future considerations to include special education students with more severe cognitive limitations should consider the potential gains in terms of teacher coverage and student participation against the cost of developing appropriate instruments and the additional demands placed on campus staff.

> **Recommendation 2**: Include Student Response Survey data in the teacher appraisal for special education teachers only on a limited basis, at the discretion of the campus administrator and special education teachers.

Given the status of district data systems and high variation among campuses in terms of their use of the master scheduling process, the process by which teachers were identified for inclusion in the Student Response Survey was extremely complex. In addition to eliminating staff from the sample who were not eligible for the pilot appraisal (e.g., counselors or assistant principals who were listed as teachers in the master schedule), determining the courses in which their students were enrolled that should be included in the sample also proved to be problematic. Each school had to be treated on a case-by-case basis, with multiple conversations with campus contacts and principals to decode their unique usage of the master schedule. It seems unlikely that all secondary principals will agree to a common set of course names or naming conventions for this purpose; therefore, the process of verifying the data should be formalized through an electronic verification system.

> **Recommendation 3:** Include a formal verification process that requires principals to identify the specific courses in the master schedule that are eligible for the survey.

What changes can be made to improve the process?

In addition to challenges determining which teachers and courses were eligible for the Student Response Survey, opportunities to train teachers who administered the survey to students in grades 1 through 12 were very limited. Teachers were provided detailed written instructions about how to log into the survey (or how to distribute the pre-coded paper surveys to students), what language to use when describing the contents of the survey, how to limit their interpretation of the survey language, how to complete the survey irregularity form, and other aspects of the process. However, not all teachers read the full instructions in advance, and several situations arose during facilitation that were unanticipated, such as the presence of a substitute teacher. Future surveys should include a brief mandatory training session for all teachers who will facilitate the survey process. The most efficient way to do this is via a video module wherein the specific procedures and policies are explained and relevant processes (e.g., the computer login procedure) are demonstrated.

> **Recommendation 4:** Provide training for all facilitators to familiarize them with

procedures and improve consistency of the survey administration process.

Feedback from the teachers who administered the survey online was largely positive. The technology worked well, the students were engaged in the activity, and the campus staff did not have to worry about distributing, collecting and returning physical forms. In addition, the district's new human capital platform includes a course evaluation module, so no additional administration costs would be incurred. Therefore, all future surveys should be administered online.

> **Recommendation 5:** To improve efficiency and reduce cost, paper surveys should be eliminated and all students should take the survey online.

Should the Student Response Survey be included in the appraisal?

The decision to include the results of the Student Response Survey in teacher appraisals has been recognized by leaders in this area as one that should not be taken lightly.

> *Rob Ramsdell, a director of the Tripod Project, which has been designing and administering student surveys since the late 1990s, advised caution. To Ramsdell, the point of student surveys is to give teachers more information about what is — and isn't — working in the classroom. "There probably is a place for them in teacher evaluation systems, but we think the use in that way needs to be handled very carefully," he said* (Butrymowicz, 2012).

The new pilot appraisal system was designed to provide feedback for teachers from multiple sources, and as a result, to allow for more meaningful opportunities for improving their practice. In addition, the inclusion of multiple sources of information was expected to better differentiate among teachers than would a single instrument, thus providing principals and district staff with opportunities to better target professional development resources. Future research should examine the extent to which the inclusion of the Student Response Survey in the appraisal contributes to these aims. Moreover, it is important to determine the extent to which it contributes in a way that justifies the time and expense of administering the survey to all students.

> **Recommendation 6:** Research should examine the extent to which the student course evaluation adds value to the appraisal results and improves differentiation among teachers.

References

Bill and Melinda Gates Foundation. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study*. Retrieved from http://www.metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

Butrymowicz, S. (2012, May 13). Student surveys for children as young as 5 years old may help rate teachers. *The Washington Post*. Retrieved September 30, 2013, from http://www.washingtonpost.com

Kloberdanz, K. (2012, May 25). Should Student Surveys Be Used to Rate Teachers? Georgia Says Yes. Retrieved September 30, 2013, from http://http://www.takepart.com/article/2012/05/25/should-student-surveys-evaluate-teachers-georgia-says-yes

Schmitt, L. (2012). *Student course feedback pilot, 2011–2012: Administration and instrumentation* (Publication No.11.76 RB). Austin, TX: Austin Independent School District.

Appendix A. Teacher Interviews and Prekindergarten through 2nd Grade Student Response Survey Testing Summary

School	Grade level/instructional area	Activity
Andrews EL	• Prekindergarten • 1st Grade	• Teacher interview/ 16 students tested (paper survey) • Teacher interview / 19 students tested (paper survey)
Jordan EL	• Bilingual prekindergarten • Bilingual 1st Grade • Bilingual kindergarten	• Teacher interview / 15 students tested (online survey) • Teacher interview / 21 students tested (online survey) • Teacher interview / 18 students tested (online survey)
Linder EL	• Special education resource kindergarten–5 • Bilingual special education resource kindergarten–5	• Teacher interview • Teacher interview
Dobie Pre-K Ctr.	• Bilingual Prekindergarten • Prekindergarten	• 11 students tested in one-on-one setting (online survey) • 12 students tested in one-on-one setting (online survey)
Blanton EL	• Kindergarten • Prekindergarten	• 15 students tested in small group setting (online survey) • 10 students tested in one-on-one setting (online survey)
Lanier HS	• Special education resource English 9–12	• Teacher interview

Appendix B. Revision to Elementary Response Scale Graphics

2011–2012 elementary response scale

Always Sometimes Not a lot Never Don't know

2012–2013 revised response scale

A lot of the time Sometimes A little of the time Never Don't know

Appendix C. Survey Administration Form

2013 Student Response Survey Administration Form

Please use the following form to document any irregularities encountered during administration of the student course evaluation. Include any specifics incidents/ occurrences that you feel may compromise the reliability or validity of student responses.

Survey Administrator (name):_____
Date:_____

Class Information:

Teacher's Name _____ Grade Level ____ Total students surveyed _____

Survey Administration:

Administration Setting: Individual Small Group Whole Class

Administration Time Requirements:_____ *(Please indicate the amount of time required to complete the survey. For pre-K and Kinder administrations include estimated average time required per each administration.)*

____No irregularities encountered

____Irregularities encountered, but did not disrupt evaluation and students' responses were not compromised

____Irregularities encountered, possibly compromised student responses

Please check all of the problems/concerns encountered during survey administration

	Student confidentiality compromised		Computer Session timed out
	Student's teacher present		Survey link not working
	One or more students unable to follow along		Student became ill during survey process -unable to complete
	Students did not understand questions		Other:_____

Please provide any additional information regarding administrative problems/concerns encountered.

_____

_____

_____

_____

Appendix D. Reported Irregularities for Prekindergarten through 5th Grade Student Response Survey Administrations

	Pre-K (*n*=44)	Kinder (*n*=53)	First (*n*=12)	Second (*n*=14)	Third (*n*=14)	Fourth (*n*=11)	Fifth (*n*=7)	All elementary (*n*=155)	All secondary (*n*=219)
No irregularities encountered	75%	79%	67%	64%	71%	82%	69%	74%	90%
Irregularities encountered, did not compromise student	17%	20%	17%	21%	14%	9%	16%	17%	6%
Irregularities encountered, possibly compromised student	5%	4%	17%	14%	14%	9%	16%	8%	3%

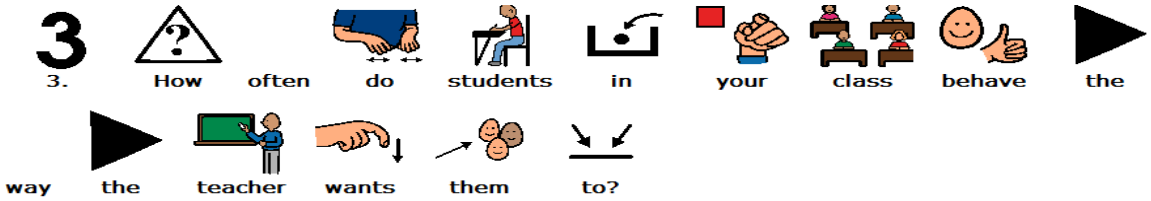
Source. 2012–2013 student response survey administration forms.
Note. Percentages may not equal 100% due to rounding.

Appendix E. 2012–2013 Administration Form: Reported Irregularities, by Category

	Elementary (n = 155)		Secondary (n = 219)	
Reported irregularities categories	*n*	%	*n*	%
Student confidentiality compromised	0	0%	2	1%
Students' teacher present	0	0%	23	11%
One or more students unable to follow along	14	9%	7	3%
Students did not understand questions	16	10%	9	4%
Computer session timed out	11	7%		
Survey link not working	0	0%		
Student answer choices influenced by other student(s) and/or school personnel	4	3%	3	1%
Other:	17	11%	21	10%
Technology disruptions / user error	9	6%		
Special education students unable to comprehend survey / not tested	4	3%	2	1%
Roster errors / absent students	2	1%	10	4%
Student behavior	2	1%	5	2%
Insufficient time to complete			1	<1%
Survey collection: confidentiality			2	1%
Survey damaged (ripped by student)			1	<1%
No irregularities		74%		90%

Source. 2012–2013 student response survey administration forms.

Appendix F. Modified Pre-Kindergarten and Kindergarten Survey Item for Secondary Life–Skills Students



AUSTIN INDEPENDENT SCHOOL DISTRICT

SUPERINTENDENT OF SCHOOLS

Meria J. Carstarphen, Ed.D.

CHIEF PERFORMANCE OFFICER

William Caritj, M.Ed.

DIRECTOR, DEPARTMENT OF RESEARCH AND EVALUATION

Holly Williams, Ph.D.

AUTHORS

Karen M. Cornetto, Ph.D.
Raymond Gross, M.Ed.

BOARD OF TRUSTEES

Vincent M. Torres, President
Gina Hinojosa, Vice-President
Jayme Mathias, Secretary
Cheryl Bradley
Ann Teich
Amber Elenz
Lori Moya
Robert Schneider
Tamala Barksdale