Department of Research and Evaluation
Austin Independent School District

AISD REACH Program Update, 2012–2013:

Peer Observation

December 2013
Publication 12.89

Lisa N. T. Schmitt, Ph.D.

EXECUTIVE SUMMARY

In 2012–2013, 1,754 teachers were each observed twice in their classrooms during the school year by one of 15 peer observers. Teachers met with their peer observers prior to receiving the unannounced classroom visits, then received written feedback and a request for a follow-up conversation within 48 hours of each observation. Those earning a combined score representing 80% of possible points were eligible to receive a stipend of $500. Approximately 80% of Austin Independent School District (AISD) REACH teachers (*n* = 1,410) earned a stipend for peer observation in 2012–2013.

On average, teachers earned 75% to 81% of possible peer observation points, depending on their grade level. Elementary and high school teachers earned significantly more points than did middle school teachers, though all groups improved significantly from observation 1 to 2. The teachers in most need of improvement at observation 1 increased most during the year, though those in the top scoring group at observation 1 declined.

Scores generally did not differ for core area, non-core area, and special education teachers. However, experienced teachers received higher ratings, on average, than did their novice peers. Additionally, experienced teachers for whom peer observation was part of their appraisal scores received more points, on average, than did those for whom peer observation was used for stipends and feedback alone.

In general, teachers reported favorable attitudes toward peer observation. The majority of teachers at every level agreed that peer observation is a good idea. However, novice teachers were more likely than their more experienced peers to agree that peer observation is a good idea, as were teachers who had received higher peer observation scores. Additionally, high school teachers for whom peer observation was part of their appraisal score were more likely than other high school teachers to report confidence in the accuracy of their peer observer's ratings and satisfaction with the support they received from their peer observer.

Peer observation scores were moderately stable over time and were moderately related to other measures of instruction, including administrators' ratings, students' ratings, and students' growth. Relationships between peer observation scores and students' growth were strongest at the middle school level. There was little relationship, however, between peer observation scores and teachers' reported instructional practices. Additionally, no relationship was found between changes in peer observation rating from year to year and changes in teachers' reported self-efficacy.

Although data indicate peer observation is a sound addition to multiple measures of evaluating teacher effectiveness, little evidence supports its influence on changing instructional practices. The cost of the peer observation program was approximately $1,635,000, or $932 per teacher served at the 38 participating schools. Despite the generally favorable attitudes toward the program, more information is needed regarding its utility for improving the quality of teaching and learning.

TABLE OF CONTENTS

LIST OF FIGURES

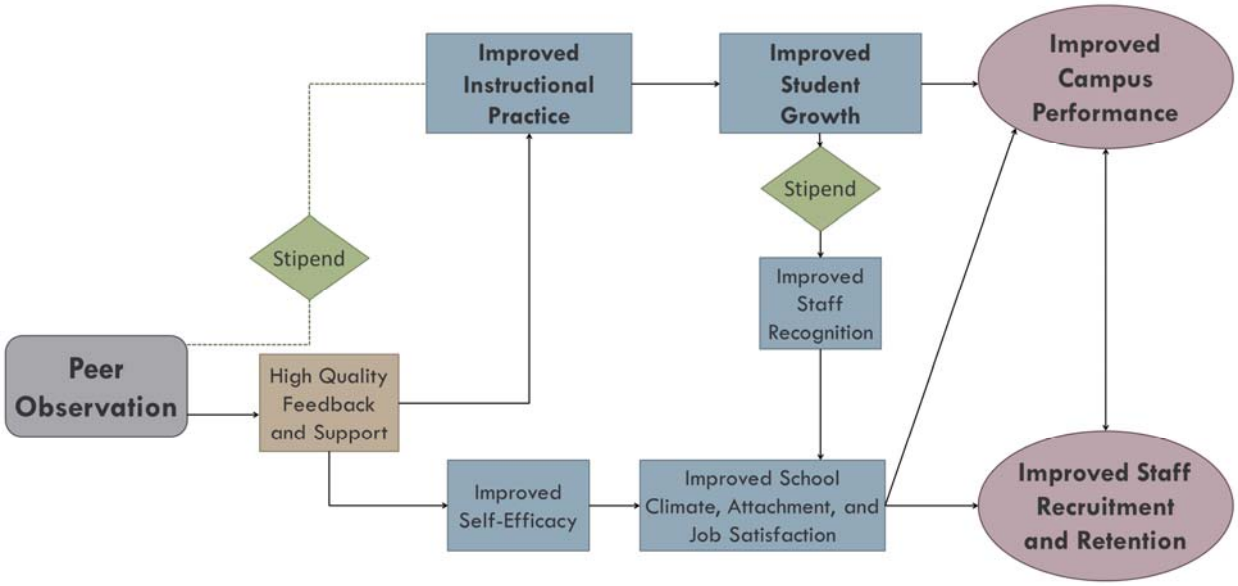
LIST OF TABLES

INTRODUCTION

The AISD REACH program, implemented in Austin Independent School District (AISD) at 38 high-needs schools in 2012–2013, provides educators with a framework for targeting instruction to meet students' needs. The program includes a variety of supports and opportunities for teachers to enhance their own instructional skills. Peer observation was designed to operate in conjunction with campus goals; teachers' goals for students (student learning objectives, or SLOs); and professional development opportunities (professional development units, or PDUs), to enhance the quality of teaching and learning at participating schools.

In 2012–2013, 15 peer observers met with their assigned teachers to discuss the observation rubric, teachers' goals for the school year, and the stipend requirements, then observed each teacher during unannounced classroom visits twice during the school year. Peer observers provided teachers with feedback, usually within 48 hours, and requested post-observation conferences to discuss their observations. The high-quality feedback and support from peer observers was expected to improve teachers' self-efficacy and instructional practices, thereby contributing positively to students' growth and teachers' attitudes regarding factors that influence retention on campus (Figure 1).

Figure 1. Hypothesized Relationships Among Peer Observation and Outcomes for Students and Teachers



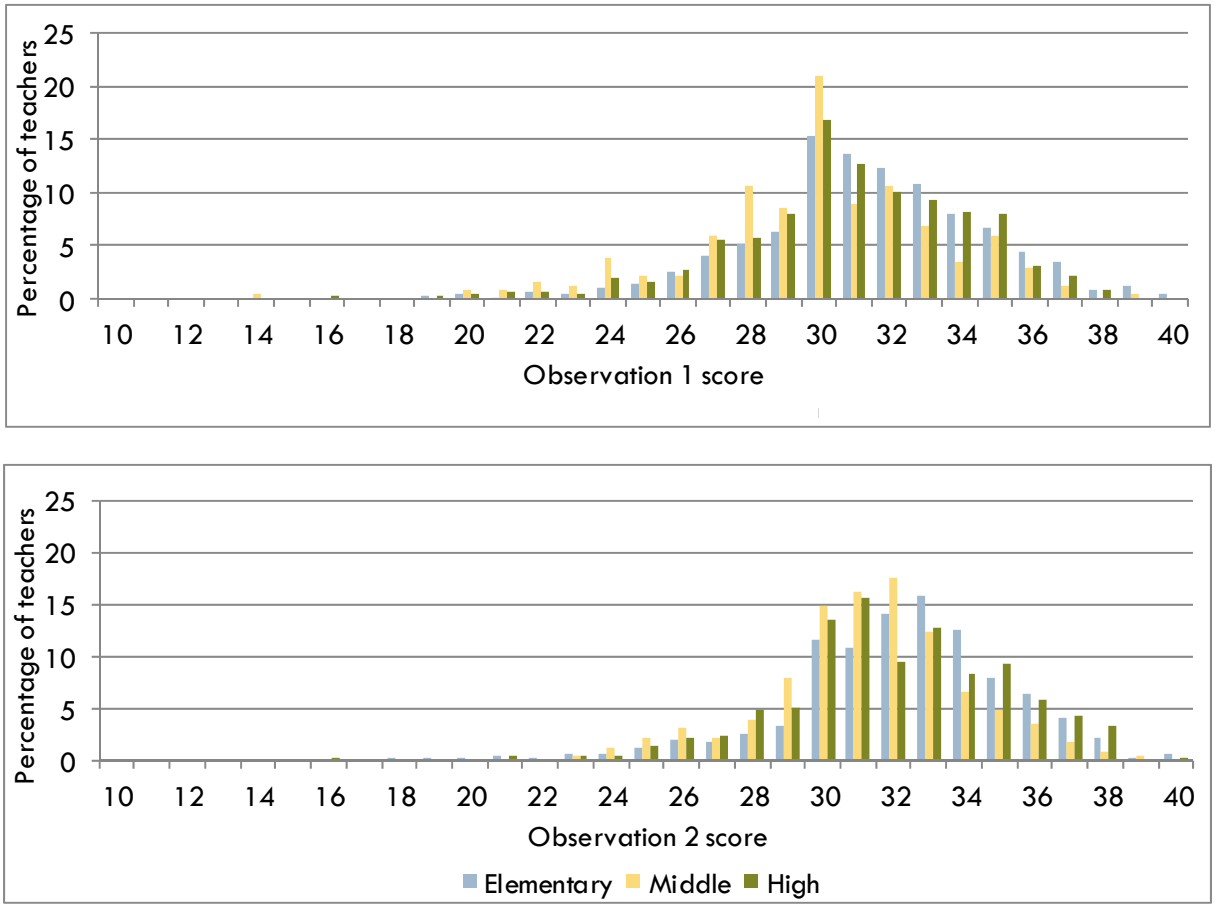
This report summarizes the results of the 2012–2013 program, including the distribution of peer observation scores, the stability of peer observation ratings over time, teachers' perceptions of peer observation, the relationship between peer observation ratings and instructional practices, and the relationship between peer observation ratings and other measures of teaching. The report also describes results with regard to the appraisal context of peer observation (i.e., whether peer observation was included in teachers' appraisal scores or not), and details the cost of the program. Finally, the report includes a summary of conclusions drawn from the results.

PEER OBSERVATION SCORES IN 2012–2013

Peer observation ratings were examined for each level, teacher type, and teacher appraisal context to identify any systematic differences in peer observation scores for teachers in specific circumstances. Analyses examined ratings for observations 1 and 2 during the 2012–2013 school year, the change in ratings from observation 1 to 2 during the 2012–2013 school year, and the change in total combined observation scores from 2011–2012 to 2012–2013.

Scores, by Level. Peer observation scores varied, ranging from 14 to 40 for observation 1 and from 16 to 40 for observation 2 (Figure 2). The most frequently occurring score for observation 1 was 30 for teachers at each level, but for observation 2 the most frequent score was 33 for elementary teachers, 32 for middle school teachers, and 31 for high school teachers.

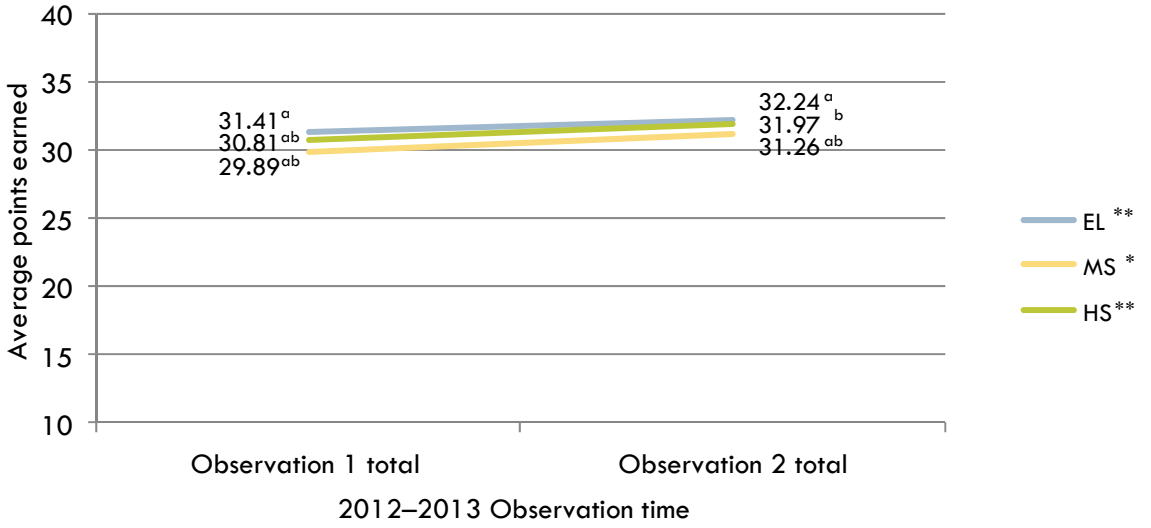
Figure 2. Distribution of Peer Observation Scores, by Level, 2012–2013



Source. 2012–2013 Peer Observation Database

Elementary and high school teachers' average scores were significantly higher than were those for middle school teachers, but scores for teachers at each level improved significantly from the first to the second time point during the year (Figure 3).

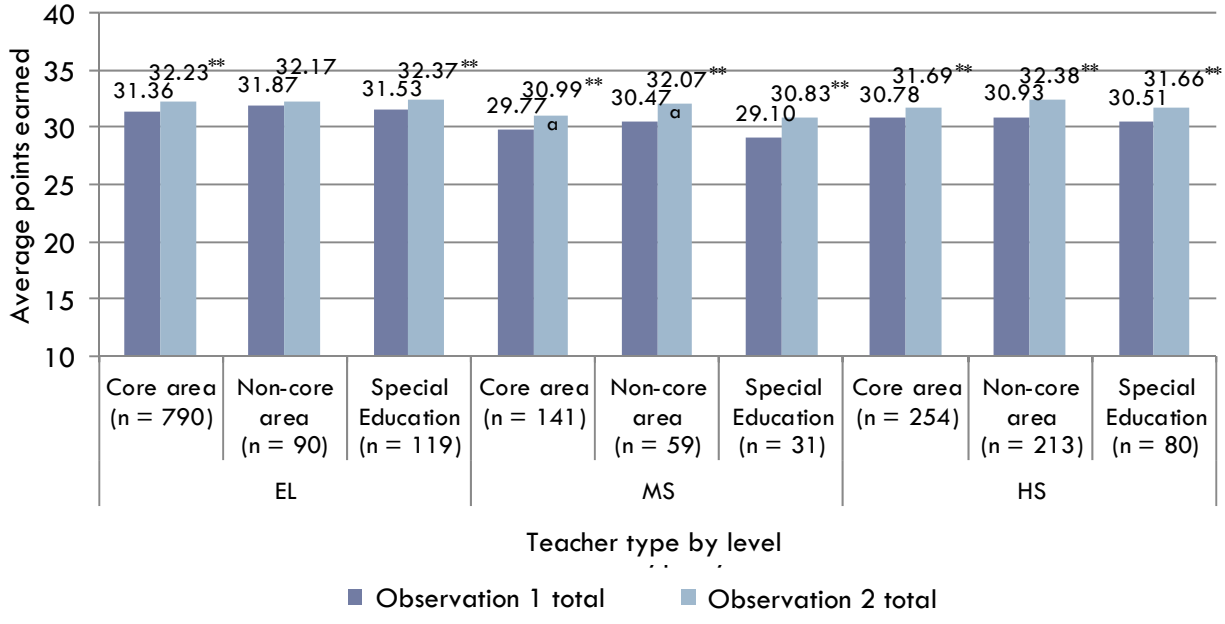
Figure 3. Average Points Teachers Earned for Observations in 2012–2013, by Level



Source. 2012–2013 Peer Observation Database
Note. Average scores for an observation time sharing the same superscript were significantly different.
**change was significant at $p < .01$; *change was significant at $p < .05$

Scores, by Teacher Type. An examination of results for each teacher type showed ratings improved significantly from observation 1 to 2 for all but elementary non-core area teachers, and the change was comparable for teacher types within each level, with the exception of core and non-core area middle school teachers (Figure 4).

Figure 4. Average Points Teachers Earned for Observations in 2012–2013, by Teacher Type and Level



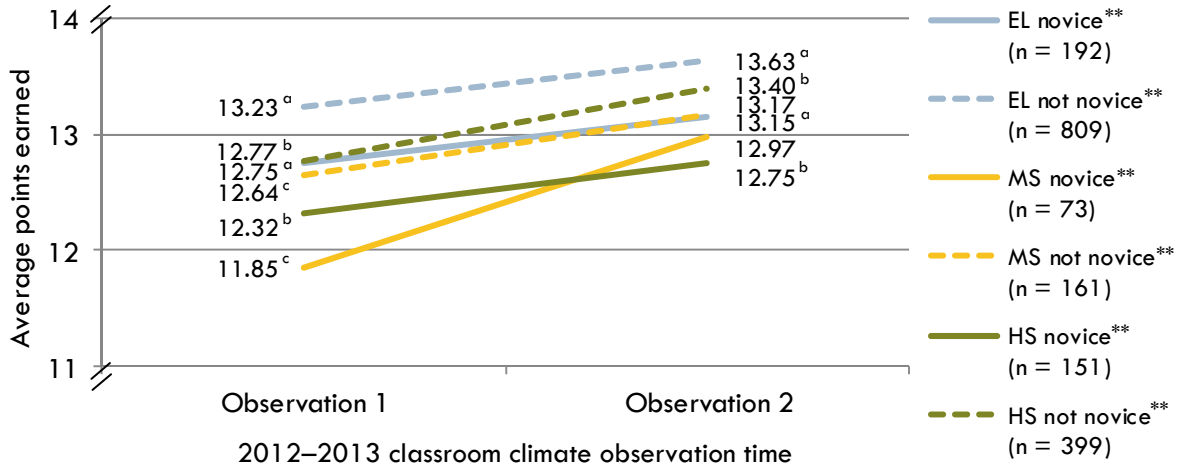
Source. 2012–2013 Peer Observation Database
Note. Changes in ratings for teacher types sharing the same superscript were significantly different within level.
**$p < .01$

Scores, by Teacher Experience. Evidence suggests novice teachers may struggle more so than do more experienced teachers with classroom management, and improved professional development opportunities regarding classroom management are critical to retention of novice teachers in the work force (Oliver & Reschly, 2007). Because peer observation scores represented the combination of scores in two domains (i.e., classroom climate and instructional practice), we examined whether differences occurred between the scores for novice and non-novice teachers for one or both domains. Figures 5 and 6 show the classroom climate and instructional practice ratings for novice and non-novice teachers at each level for observations 1 and 2. Teachers were rated in four areas for the classroom climate domain and six areas for instructional practice.

Elementary novice teachers were rated significantly lower than were their non-novice peers for all four categories (i.e., classroom climate and instructional practice, observations 1 and 2), and high school novice teachers were rated significantly lower than were their non-novice peers for three categories (i.e., classroom climate, observations 1 and 2, and instructional practice, observation 2). Middle school novice teachers, however, were rated lower than were their non-novice peers only for one category (i.e., classroom climate, observation 1).

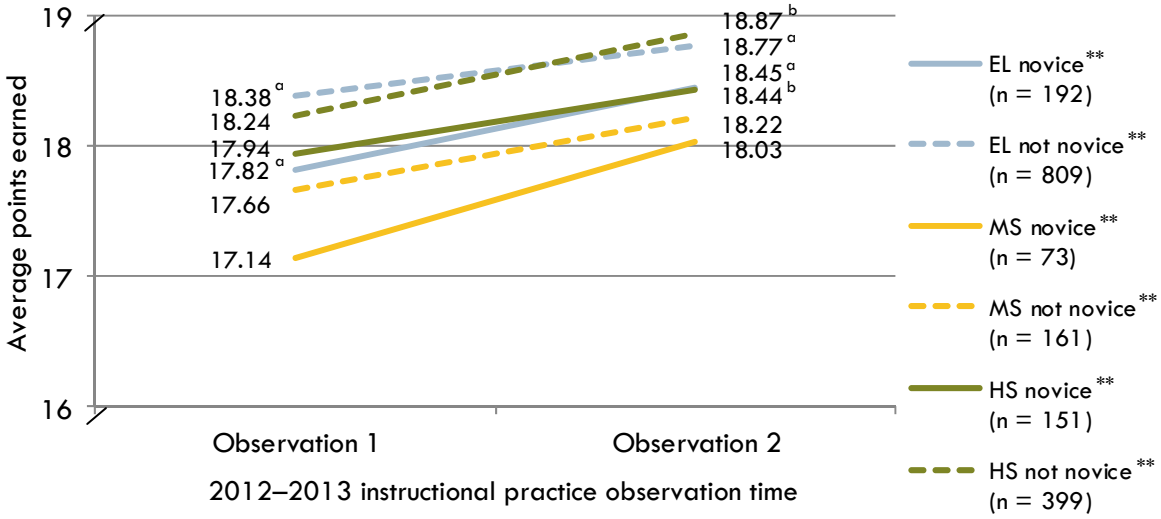
Each group improved significantly from observation 1 to 2 for both classroom climate and instructional practice (Figures 5 and 6, respectively). In five of six comparisons, the rate of improvement was not different for novice teachers than it was for their non-novice peers. However, middle school novice teachers improved significantly more from observation 1 to 2 than did their non-novice peers on Classroom Climate ($t[226] = 2.24$, $p < .05$; Figure 5).

Figure 5. Average Classroom Climate Rating, by Novice Status for Observations 1 and 2, 2012–2013



Source. 2012–2013 Peer Observation Database
Note. Averages for teacher types sharing the same superscript were significantly different within level and time. The total number of possible points for Classroom Climate ranged from 4 to 16.
**change was significant at $p < .01$

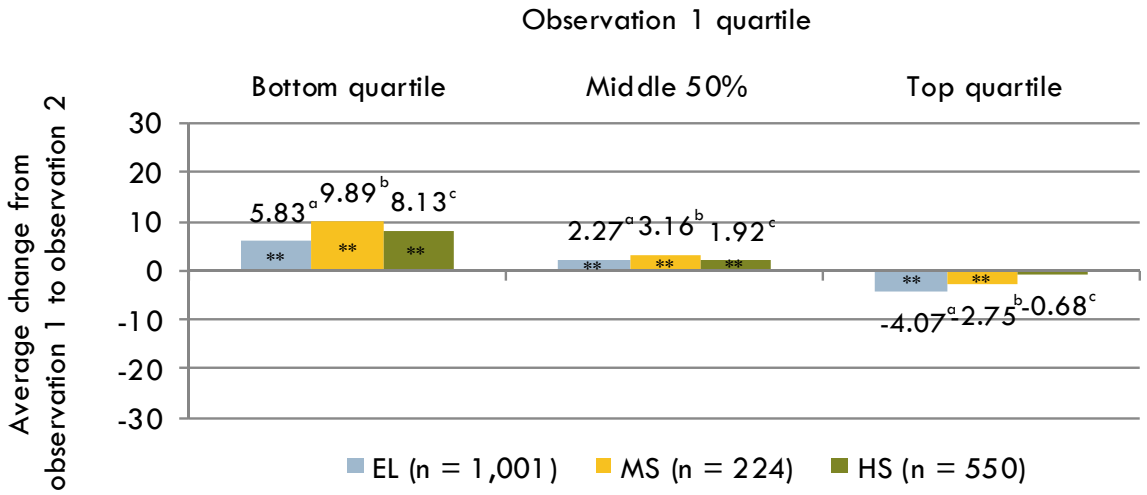
Figure 6. Average Instructional Practice Rating, by Novice Status for Observations 1 and 2, 2012–2013



Source. Peer Observation Database, 2012–2013
Note. Averages for teacher types sharing the same superscript were significantly different within level and time.
The total number of possible points for Instructional Practices ranged from 6 to 24.
EL = elementary school; MS = middle school; HS = high school
**change was significant at $p < .01$

Scores, by Previous Observation Rating Quartile. We wanted to determine whether those in greatest need of improvement did improve. Thus, we divided teachers into quartiles by level according to their scores for observation 1. Results showed teachers who had scored in the bottom quartile at observation 1 improved significantly more than did those who had scored in

Figure 7. Change from Observation 1 to 2, by Observation 1 Quartile and Level, 2012–2013

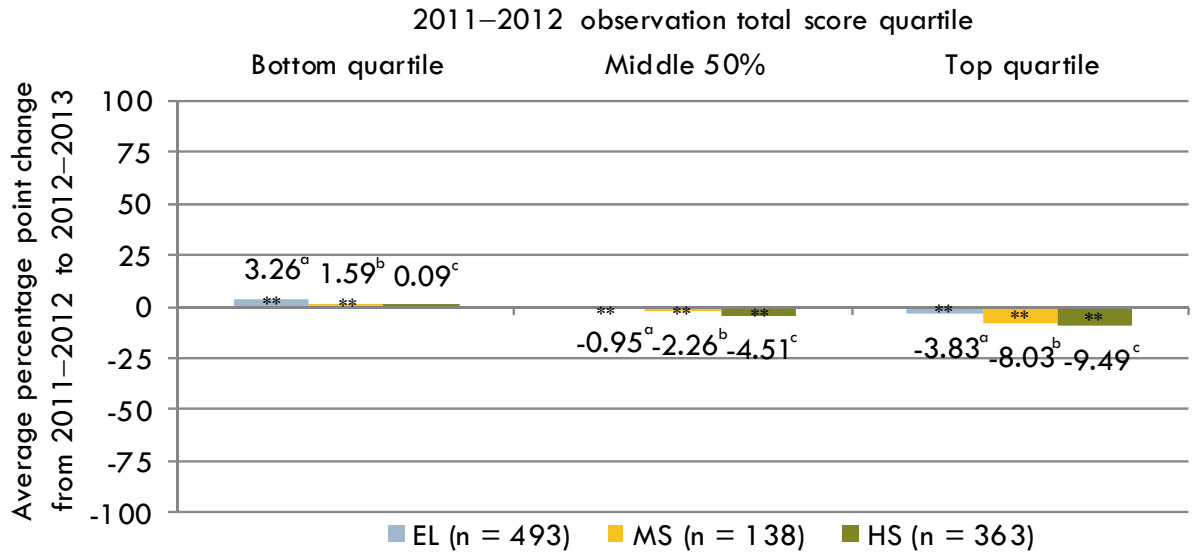


Source. 2012–2013 Peer Observation Database
Note. Average change for quartile groups sharing the same superscript was significantly different within level.
The scale for 2012–2013 ranged from 10 to 40; thus, change scores may range from –30 to 30.
EL = elementary school; MS = middle school; HS = high school
**Change is significant at $p < .01$

the middle 50% or bottom quartile, and teachers who had scored in the middle 50% improved significantly more than did those who had scored in the bottom quartile for the first observation (Figure 7). With the exception of high school teachers who scored in the top quartile for observation 1, the change from observation 1 to 2 was statistically significant for all groups of teachers. On average, the scores of elementary and middle school teachers who were in the top quartile at observation 1 declined, while the scores of all other groups of teachers improved.

We also examined the change in total scores for teachers according to how they had scored in the previous year (2011–2012). As with change from observation 1 to 2 during the 2012–2013 school year, teachers who had scored in the bottom quartile in 2011–2012 improved more in 2012–2013 than did those from the middle 50% or top quartile, and teachers who had scored in the middle 50% improved more than did those who had scored in the top quartile in 2011–2012 (Figure 8). Additionally, change for all but one group (high school teachers from the bottom quartile) was significant. Only elementary and middle school teachers from bottom quartile in 2011–2012 improved significantly in 2012–2013. All others declined.

Figure 8. Change from 2011–2012 to 2012–2013, by 2011–2012 Quartile and Level, 2012–2013



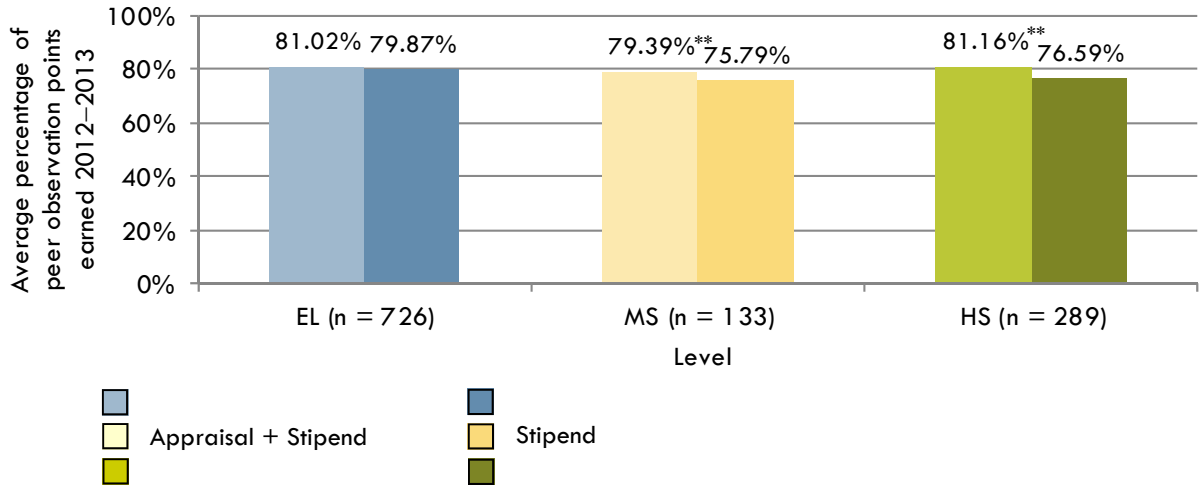
Source. 2011–2012 and 2012–2013 Peer Observation Database

Note. Due to different possible point values for observation scores in 2011–2012 and 2012–2013, scores have been converted to represent the percentage of possible points earned each year.
Average change for quartile groups sharing the same superscript was significantly different within level.
EL = elementary school; MS = middle school; HS = high school

**change from 2011–2012 to 2012–2013 is significant at $p < .01$.

Scores, by Appraisal Context. Peer observation scores were used as one measure included in experienced (i.e., non-novice) teachers' appraisal scores at 12 of the 38 schools. We were interested in determining if scores differed systematically according to whether peer observation was used in teachers' appraisals. Indeed, scores were significantly higher for middle and high school teachers whose peer observation ratings were incorporated into the appraisal than for those whose peer observation ratings were used for stipend criteria alone (Figure 9).

Figure 9. Percentage of 2012–2013 Peer Observation Points Earned, by Appraisal Context



Source. 2012–2013 Peer Observation Database; 2012–2013 Appraisal Database
Note. EL = elementary school; MS = middle school; HS = high school
**$p < .01$

Middle and high school teachers whose peer observations were included in the appraisal score also were significantly more likely to earn peer observation stipends than were their equally experienced peers whose peer observations were not used for appraisal purposes ($\chi^2(1, n = 133) = 6.48, p < .01$ and $\chi^2(1, n = 289) = 14.72, p < .01$ for middle and high school teachers, respectively).[1]

Change in Peer Observation Versus Other Measures. To determine whether change in peer observation scores reflected anything other than simply a tendency for scores to be higher at observation 2 than at observation 1, we assessed whether teachers whose peer observation scores increased over time also demonstrated an increase in other measures of teaching. Because no other measures of teaching were consistently implemented during the same year for all teachers, we examined whether the change in peer observation scores from 2011–2012 to 2012–2013 corresponded with the change in administrators' appraisal ratings from 2011–2012 to 2012–2013. Data suggest a weak-to-moderate correlation between changes in peer observation scores and changes in administrators' appraisal ratings for teachers who had scored in the bottom quartile at peer observation 1 in 2011–2012 (Appendix A). Elementary and middle school teachers who were in the bottom quartile were likely to have improved on both peer observations and administrator appraisal ratings in 2012–2013. However, high school teachers who were in the bottom quartile were likely to improve on one measure but decline on the other in 2012–2013. These results are limited by the exclusion of teachers who were appraised with different systems in 2011–2012 and 2012–2013; change in appraisal scores could not be assessed for over half the teachers with peer observation ratings both years.

[1]Peer observation was included in the appraisal score for experienced teachers not in contractual difficulty. Those in contractual difficulty were included in the group whose appraisal did not include peer observation.

STABILITY OF PEER OBSERVATION SCORES OVER TIME

Within the 2012–2013 school year, peer observation scores for observations 1 and 2 were moderately and significantly correlated (Table 1), and this result was similar to that found in 2011–2012 (Lamb & Schmitt, 2011). Because the peer observation program was implemented in 2011–2012, previous analyses could not examine the stability of peer observation from year to year. With two years of data, we measured the relationship between peer observation scores in 2011–2012 and 2012–2013.

Table 1. Correlations Between Peer Observation Scores Within 2012–2013 and From 2011–2012 to 2012–2013

				Correlation with 2012–2013 peer observation score	
Year	Level	Score	n	Observation 1	Observation 2
2012–2013	Elementary	Observation 1	990	n/a	.52**
	Middle	Observation 1	228	n/a	.48**
	High	Observation 1	538	n/a	.59**
2011–2012	Elementary	Observation 1	491	.35**	.33**
		Observation 2	487	.29**	.37**
	Middle	Observation 1	140	.11	.04
		Observation 2	137	.17*	.13
	High	Observation 1	357	.36**	.42**
		Observation 2	360	.36**	.39**

Source. 2012–2013 and 2011–2012 Peer Observation Database
*$p < .05$
**$p < .01$

Scores were slightly more related within school year than they were from year to year. However, the peer observation scoring rubric changed in 2012–2013; therefore, we would expect a greater correlation within year than across years for that reason alone. Additionally, 10 of the 15 peer observers were new to the role in 2012–2013. Peer observation ratings for elementary and high school teachers were more consistent across years than were those for middle school teachers. For context, we compared the stability of peer observation scores with that of teachers' appraisal scores over time.

Peer observation scores for elementary and middle school teachers were less consistent across years than were appraisal scores [$r(607) = .60$, $p < .01$ and $r(88) = .58$, $p < .01$ for elementary and middle school appraisal scores, respectively]. High school peer observation and administrative appraisal scores were equally stable over time [$r(114) = .39$, $p < .01$ for high school appraisal scores].

PEER OBSERVATION AND OTHER MEASURES OF INSTRUCTION

To assess the validity of peer observation ratings, we correlated peer observation scores with other measures of instruction including teacher appraisal ratings, students' instructional ratings, teachers' self-ratings of their instructional practices, and students' growth scores. In general, peer observation scores were moderately related to several other measures of instruction.

Teachers with high peer observation total scores were significantly more likely than those with low peer observation scores to also have received high appraisal observation ratings from their administrators on either the district's Professional Development and Appraisal System (PDAS) or the pilot teacher appraisal system (Table 2). In addition, high school teachers with high peer observation total scores were significantly more likely than their peers with low peer observation scores to have received high ratings of their instructional practices from students and high ratings for professional expectations from their administrators.

Table 2. Correlations Between Peer Observation Total and Other Observational Measures, 2012–2013

		Correlation with peer observation total score		
		Elementary	Middle	High
Original appraisal	Percentage of possible PDAS points earned	.41**	.29**	.36**
Pilot appraisal	Administrator walkthrough rating total	.22**	.45**	.40**
	Administrator formal observation rating	.16	.48**	.42**
	Professional expectations rating	.15	.08	.38**
	Student ratings	.17	.27	.26**

Source. 2012–2013 Peer Observation Database; 2013 Pilot Appraisal Database
*$p < .05$
** $p < .01$

Peer observation ratings were minimally related to teachers' self-reported instructional practices (Table 3), though some correlations were statistically significant. The weak connection between peer observation ratings and these instructional practices is not surprising, however, given the

Table 3. Correlations Between Peer Observation Total and Teachers' Instructional Practices, 2012–2013

	Correlation with peer observation total score		
	Elementary	Middle	High
Reflective teaching	.07	.00	.12
PLC behaviors (ECS)	.12*	.00	-.02
PLC behaviors (TELL)	.00	.09	-.02
Data use	.07*	.06	-.01

Source. 2012–2013 Peer Observation Database; 2013 TELL AISD Survey; 2013 Employee Coordinated Survey
Note. PLC = professional learning community
*$p < .05$

differences in the behaviors assessed with each measure. See Appendix B for sample sizes and Appendix C for survey items. We also examined whether changes (improvements or declines) in peer observation ratings from one year to the next were correlated with changes in teachers' reported self-efficacy. No relationship was found between changes in peer observation rating from year to year and changes in teachers' self-efficacy.

However, results did show some significant relationships between peer observation scores and students' growth data for those teachers with student Educational Value Added Assessment System (EVAAS®) results (Table 4). Teachers with higher observation 1 scores had significantly higher EVAAS scores than did those with lower observation 1 scores in 27% of possible instances (all at the middle school level), and teachers with higher observation 2 scores had significantly higher EVAAS scores than did those with lower observation 2 scores in 45% of possible instances. Overall, in 45% of possible instances, teachers with higher observation totals (i.e., sum of observations 1 and 2) were significantly more likely to have high EVAAS scores than were teachers with lower observation totals.

Table 4. Correlations Between Peer Observation Scores and EVAAS Scores, by Subject, 2012–2013

Level	Subject EVAAS score	*n*	Observation 1	Observation 2	Observation total
Elementary	Reading/ELA	161	-.03	.02	-.01
	Math	146	.09	.36**	.25**
	Science	59	.12	.36**	.25*
Middle	Reading/ELA	40	.54**	.28	.49**
	Math	38	.41**	.34*	.46**
	Science	13	.55*	.09	.51
	Social studies	13	.06	-.35	-.13
High	Reading/ELA	54	.24	.36**	.33**
	Math	54	.11	.24	.19
	Science	39	.09	.31*	.24
	Social studies	40	.27	.28	.30

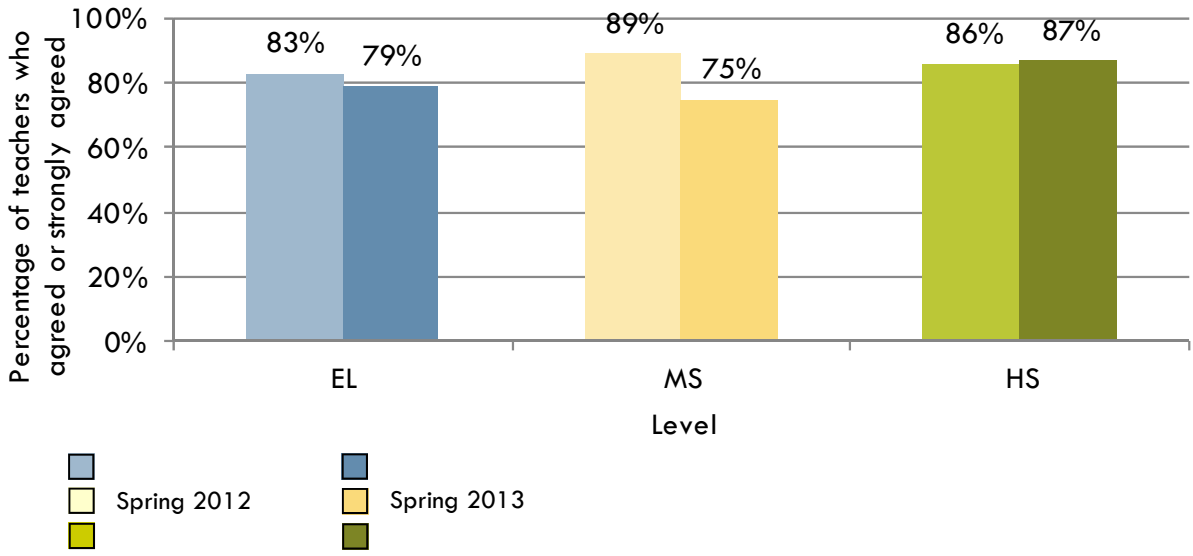
Source. 2012–2013 Peer Observation Database, 2013 EVAAS
*$p < .05$
**$p < .01$

Results suggest peer observation, administrator observation, student feedback, and EVAAS measure some similar aspects of teaching.

TEACHERS' PERCEPTIONS OF PEER OBSERVATION

In general, teachers reported favorable attitudes toward peer observation. At least three-quarters of teachers at each level agreed or strongly agreed that peer observation is a good idea (Figure 10). Ratings of middle school teachers in 2012–2013 were less favorable than were those of middle school teachers in 2011–2012, and the decline in attitudes of middle school teachers who responded to the survey item in both Spring 2012 and Spring 2013 approached statistical significance ($t[16] = -2.06$, $p = .06$). However, 75% of middle school teachers still agreed that peer observation is a good idea.

Figure 10. Percentage of Teachers Who Agreed or Strongly Agreed That Peer Observation is a Good Idea, Spring 2012 and Spring 2013



Source. 2012 and 2013 Employee Coordinated Survey
Note. EL = elementary school; MS = middle school; HS = high school

Teachers with high peer observation scores were significantly more likely than those with low scores to agree that peer observations are a good idea (Table 5). Additionally, those who agreed peer observation is a good idea were more likely than those who did not agree to believe strategic compensation is a good idea.

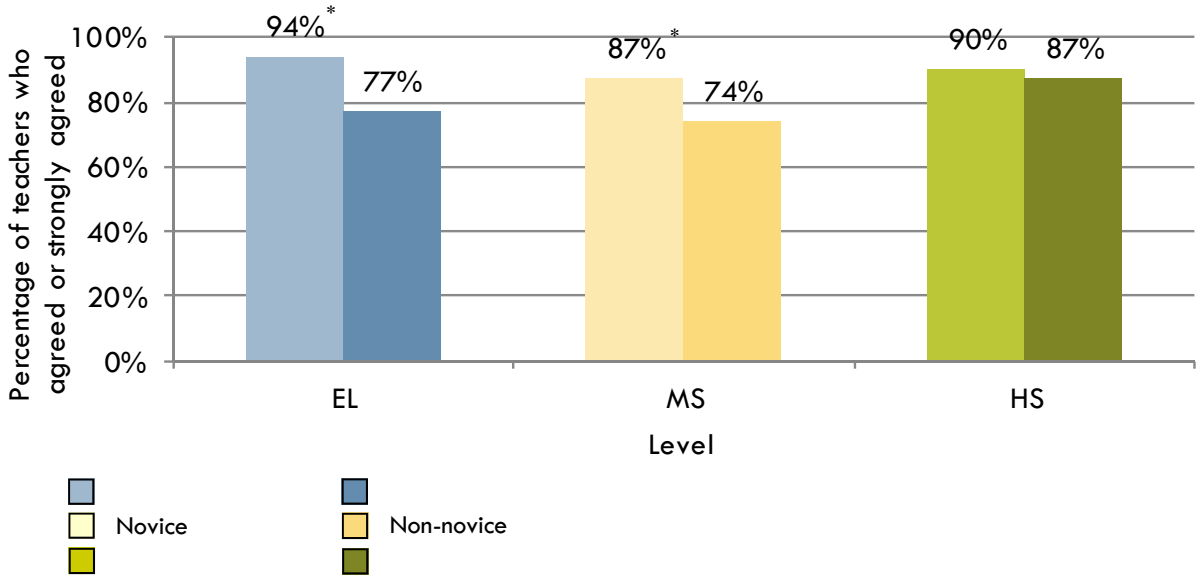
Table 5. Correlations Between Peer Observation Scores and Attitudes Toward Peer Observation and Strategic Compensation

	Correlation with rating for "Peer observation is a good idea."		
	Elementary	Middle	High
Observation 1	.27**	.19	.33**
Observation 2	.28**	.34*	.20**
"Strategic compensation is a good idea."	.32**	.32*	.21**

Source. 2012–2013 Peer Observation Database; 2013 Employee Coordinated Survey
*$p < .05$
**$p < .01$

Novice teachers at the elementary and middle school levels also were more likely than their more experienced peers to believe peer observation is a good idea (Figure 11).

Figure 11. Percentage of Novice and Non-novice Teachers Who Agreed or Strongly Agreed That Peer Observation is a Good Idea, Spring 2013

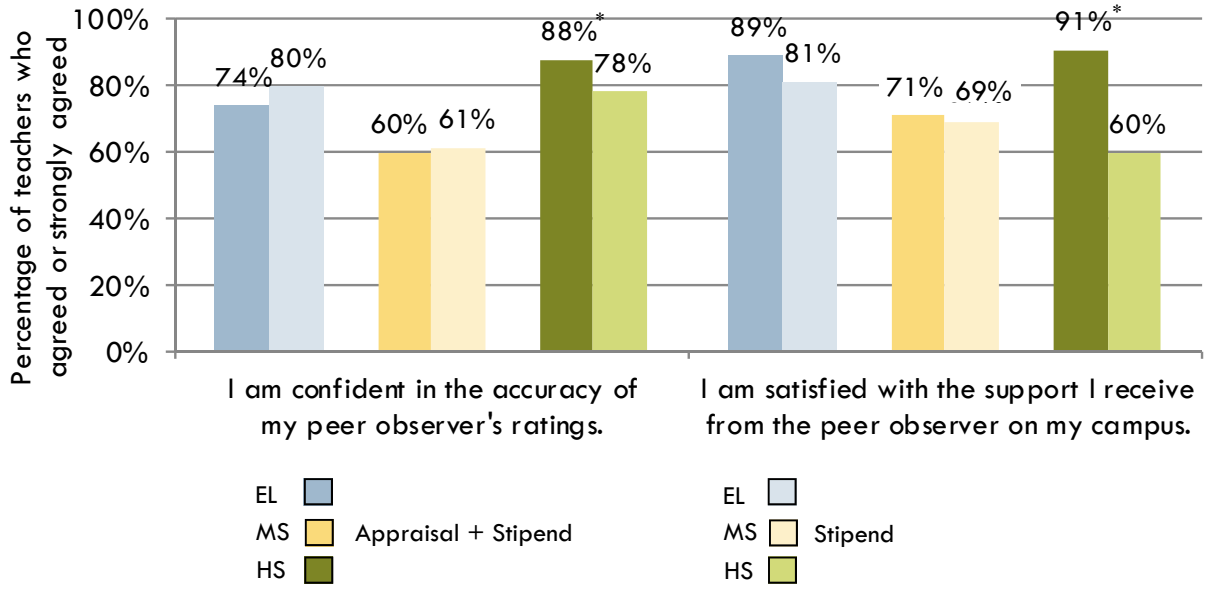


Source. 2013 Employee Coordinated Survey
Note. EL = elementary school; MS = middle school; HS = high school
*$p < .05$

We also were interested in whether appraisal context may have influenced teachers' opinions about peer observation. Because novice teachers were generally more favorable toward peer observation than were non-novice teachers, and novice teachers could not be part of the appraisal system that incorporated peer observation, we limited our analysis to a comparison of non-novice teachers whose appraisal incorporated peer observation versus those for whom peer observation was used for stipends alone.

High school teachers for whom peer observation was part of the appraisal reported significantly more confidence in the accuracy of their peer observer's ratings and more satisfaction with the support they received from the peer observer on their campus than did their peers without peer observation as an appraisal component (Figure 12). They also were more likely to agree their peer observer collaborated with them to improve their teaching (Table 6). Attitudes did not differ, however, regarding other aspects of peer observation, nor was a difference found among elementary or middle school teachers according to appraisal context in their attitudes toward peer observation.

Figure 12. Percentage of Non-novice Teachers Who Agreed or Strongly Agreed They Are Confident in the Accuracy of Their Peer Observer's Ratings and Are Satisfied With the Support They Receive From Their Peer Observer, by Appraisal Status, Spring 2013



Source. 2013 Employee Coordinated Survey
Note. See Appendix D for cell sizes. EL = elementary school; MS = middle school; HS = high school
*$p < .05$

Table 6. Percentage of Non-novice Teachers Who Agreed or Strongly Agreed With Peer Observation Items, by Appraisal Status, Spring 2013

	Elementary		Middle		High	
	Appraisal + stipend	Stipend	Appraisal + stipend	Stipend	Appraisal + stipend	Stipend
My peer observer collaborates with me to improve my teaching.	87%	71%	56%	61%	82%*	71%
I often consider the feedback that I received during my post-observation conference when planning and conducting my daily work.	91%	78%	60%	57%	80%	74%
My students have benefitted from the feedback that I received during my post-observation conference.	86%	73%	44%	60%	83%	70%

Source. 2013 Employee Coordinated Survey
Note. See Appendix D for cell sizes.
*$p < .05$

Attitudes toward peer observation were favorable among experienced elementary and high school teachers. However, experienced middle school teachers were less likely than their elementary or high school peers to agree they or their students benefitted from peer observation.

COST OF PEER OBSERVATION

To implement the peer observation program in 2012-2013, 15 full-time peer observers were employed for intensive training beginning the month prior to the start of the school year for a cost of approximately $930,000. The cost equates to approximately $530 per teacher observed, before stipends. Including the $705,000 paid to those teachers who earned stipends for achieving satisfactory scores, the total cost for peer observation was approximately $1,635,000, or an average of $932 per teacher observed.

CONCLUSION

Evidence suggests peer observation is a fairly reliable and valid measure of teachers' classroom instruction. Teachers' scores from peer observations during the same school year were moderately related to each other, and also were moderately related to other measures, including their students' growth and their administrators' and students' ratings of their teaching. Peer observation ratings appeared to measure aspects of classroom instruction similar to those that are assessed in other ways. Changes to the observational rubric in 2012–2013 likely suppressed the magnitude of the stability in peer observation ratings over time we might otherwise have found; nevertheless, ratings were moderately stable from year to year. In fact, peer observation ratings were equally as stable across years as were administrators' appraisal scores at the high school level. Results support the inclusion of peer observation as part of the pilot teacher appraisal system at 12 schools.

Despite concerns voiced in Spring 2012 regarding the impending use of peer observation ratings in the appraisal system (Lamb & Schmitt, 2012), teachers did not appear to have been negatively affected by the new context, which included a dual purpose for peer observation. Rather, attitudes toward peer observation were equally if not more favorable among the experienced teachers whose scores were included in their appraisal than among their peers. This may have reflected the fact that middle and high school teachers with peer observation in their appraisal earned more points and consequently were more likely to earn stipends than were their peers. Data showed those who scored higher were more likely than those with lower scores to have agreed peer observation is a good idea.

It is noteworthy, however, that even though novice teachers scored lower than did non-novice teachers, they were in fact more likely than their experienced peers to have agreed peer observation is a good idea. Novice teachers at REACH schools were accustomed to regular classroom observation and feedback from their assigned mentor teachers, which may have predisposed them to more favorable attitudes about classroom observation in general.

The feedback teachers received after observation should have facilitated improvements in classroom climate and instructional practices following the conference. Indeed, teachers scored significantly higher, on average, at observation 2 than they had before. Scores increased for teachers of all types and all levels of experience. Scores increased most for teachers who had originally scored in the bottom quartile or middle 50%. Middle school novice teachers, who had scored lower on average than any other group at observation 1, increased to a greater degree than did their experienced peers. However, scores for novice elementary and high school teachers did not increase at a faster rate than those for their more experienced peers; thus, novice elementary and high school teachers still scored significantly lower than their peers at observation 2.

Elementary and middle school teachers who had scored in the top quartile at observation 1 actually received significantly lower scores at observation 2 than they had at observation 1. The decline of approximately 4 percentage points for elementary and 3 percentage points for

middle school teachers suggests some potential for regression to the mean, which occurs when successive measurements are taken using an instrument that incorporates both skill and some part chance. The magnitude of decline among the top quartile, however, was much smaller than the magnitude of increase among the bottom quartile (approximately 6 to 10 percentage points across levels). Therefore, we can assume that the increases found among teachers who had scored in the bottom quartile represented more than simply chance.

An increase in scores from observation 1 to observation 2 could reflect nothing more than a general tendency for observers to have rated teachers higher on the second observation. Thus, we should rely on an alternate measure to assess improvement in teachers' instruction. Indeed, changes in peer observation scores for elementary and middle school teachers who had scored in the bottom quartile corresponded somewhat with changes in administrators' appraisal ratings. However, changes in peer observation scores for high school teachers who had scored in the bottom quartile corresponded with inverse changes in administrators' appraisal ratings. In other words, when peer observation ratings improved, administrator ratings declined (and vice-versa). This surprising relationship should be examined more closely in the future if we are to understand whether changes in observation ratings do indeed reflect improvements and declines. It is likely these results were influenced by the exclusion of many teachers, particularly at the high school level, due to the lack of comparable administrator appraisal rating systems across both years. The lack of relationship between changes in the ratings from two different sources for teachers who had scored in the top quartile or middle 50% is not necessarily cause for concern, given the minimal changes that actually occurred in the scores for teachers in these two groups. The range in peer observation change scores for these groups may simply have been too narrow for correlations with change in administrators' ratings to have been found.

Unfortunately, the options for assessing changes in teachers' instructional practices over time were limited. Aside from administrators' ratings and peer observation scores, no other instructional measure was available for all teachers at multiple time points. Additionally, the survey measures that were used for assessing some instructional practices (i.e., data use, reflective teaching, and engagement in PLCs) did not necessarily address the behaviors that may be improved through peer observation and feedback. Because peer observers typically identified and discussed two specific areas for improvement during their post-observation conferences with teachers, the available observational and survey data may not have sufficiently addressed the areas on which teachers focused their efforts for instructional improvements. To truly evaluate the influence of peer observation on teachers' practice, another measure would be necessary.

However, we could examine teachers' opinions regarding whether peer observation was useful to them and their students. The majority of teachers did, in fact, report peer observation is a good idea. Most were satisfied with the support they received from their peer observer and agreed their peer observer collaborated with them to improve their teaching. Additionally, the majority said they often considered the feedback they received during the post-observation conferences. Many even reported their students had benefitted from the feedback they

received. Thus, it seems peer observation was a well-received program that many teachers, especially at the elementary and high schools, valued.

The cost of the program, however, was not insignificant. Better ways of assessing the true influence of peer observation on teachers' classroom instruction are imperative so we may rely on more than teachers' perceptions of the program when evaluating its effectiveness. Although evidence indicates peer observation is a sound addition to multiple measures of evaluating teacher effectiveness, scant evidence supports its influence on changing practices.

APPENDIX

Appendix A. Correlations Between Changes in Peer Observation Scores and Changes in Administrators' Appraisal Ratings, 2011–2012 to 2012–2013, by Peer Observation Quartile in 2011–2012

		Correlation with change in peer observation scores from 2011–2012 to 2012–2013		
	Quartile in 2011–2012	Elementary	Middle	High
Change in administrators' PDAS rating from 2011–2012 to 2012–2013	Bottom	.21* (n = 103)	.41* (n = 23)	-.31* (n = 41)
	Middle 50%	.21* (n = 123)	-.17 (n = 28)	-.20 (n = 43)
	Top	.02 (n = 82)	.07 (n = 26)	.19 (n = 19)
Change in administrators' pilot appraisal system formal observation rating from 2011–2012 to 2012–2013	Bottom	—	—	—
	Middle 50%	-.03 (n = 13)	—	.14 (n = 19)
	Top	—	—	-.03 (n = 15)

Source. 2012–2013 Peer Observation Database; 2013 PDAS Database; 2013 Pilot Appraisal Database
Note. PDAS = Professional Development and Appraisal System; Many teachers at 3 high schools, 1 middle school, and 5 elementary schools were not included in the analyses above because they were appraised with different systems each year (i.e., PDAS in 2011–2012 and the pilot appraisal system in 2012–2013).
*$p < .05$

Appendix B. Sample Sizes for Correlations Between Peer Observation Total and Other Observational Measures, 2012–2013

		Elementary	Middle	High
Original appraisal	Percentage of possible PDAS points earned	766	154	242
Pilot appraisal	Administrator walkthrough rating total	129	43	166
	Administrator formal observation rating	129	43	166
	Professional expectations rating	129	43	166
	Student ratings	77	38	147
All teachers	ECS reflective teaching	274	57	175
	ECS PLC	273	57	171
	TELL PLC	907	187	458
	TELL data use	904	185	460

Source. 2012–2013 Peer Observation Database; 2013 Pilot Appraisal Database; 2013 TELL AISD Survey; 2013 Employee Coordinated Survey

Appendix C. Data Use, Professional Learning Community (PLC), and Reflective Teaching Survey Items

Scale	Item stem and response options	Item
Data use (2013 TELL)	How frequently do you use data in the following ways? (Once a year, Once a semester, Once every two months, Once a month, Twice a month, Once a week)	Comparing test scores for your class across academic years (e.g., how 5th grade class as a whole performed in 3rd grade and 4th grade).
		Examining current benchmark scores to create classroom instructional groups.
		Examining data to identify students in need of intervention.
		Collaborating with other educators about data and how it relates to the learning needs of students.
PLCs (2013 TELL)	Indicate how much you agree or disagree with each statement. I participate with a group of my campus colleagues to: (Strongly agree, Agree, Disagree, Strongly disagree, Don't know)	Analyze student performance data
		Discuss ways to meet objectives for specific students
		Plan lessons and units together
		Develop common student assessments
PLCs (2013 Employee Coordinated Survey)	How often does your department/team: (Frequently, Often, Sometimes, Rarely, Unsure/N/A)	Discuss your department/team's professional development needs and goals
		Discuss assessment data for individual students
		Set learning goals for groups of students
		Group students across classes based on learning needs
		Provide support for new teachers
		Provide support for struggling teachers
		Share instructional strategies
Reflective teaching (2013 Employee Coordinated Survey)	How frequently do: (Frequently, Often, Sometimes, Rarely, Unsure/N/A)	Reflections on your past teaching experiences influence your lesson plans?
		You seek out collaboration with other teachers to improve a lesson plan that did not go well?
		You work with other teachers to improve your teaching even when it is going well?
		You adjust your instructional strategies based on student assessment results?

Appendix D. Sample Sizes for Peer Observation Items, by Appraisal Context, 2012–2013

	Elementary		Middle		High	
	Appraisal + stipend	Stipend	Appraisal + stipend	Stipend	Appraisal + stipend	Stipend
I am confident in the accuracy of my peer observer's ratings.	23	187	10	23	57	41
I am satisfied with the support I receive from the peer observer on my campus.	9	107	7	13	33	20
My peer observer collaborates with me to improve my teaching.	23	182	9	23	57	38
I often consider the feedback that I received during my post-observation conference when planning and conducting my daily work.	22	186	10	23	56	38
My students have benefitted from the feedback that I received during my post-observation conference.	22	183	9	20	52	37

Source. 2012–2013 Peer Observation Database; 2013 Pilot Appraisal Database; 2013 Employee Coordinated Survey

REFERENCES

Lamb, L., & Schmitt, L. (2011). AISD Reach program update, 2011-2012: The peer observation program (DRE Publication No. 11.64). Austin, TX: Austin Independent School District.

Lamb, L., & Schmitt, L. (2012). AISD pilot teacher appraisal system, 2011-2012: Multiple measures of teacher evaluation (DRE Publication No. 11.80). Austin, TX: Austin Independent School District.

Oliver, R. M., & Reschly, D. J. (2007). Effective classroom management: Teacher preparation and professional development. TQ Connection Issue Paper, National Comprehensive Center for Teacher Quality.