

EDUCATOR EVALUATION IN TENNESSEE: FINDINGS FROM THE 2014 FIRST TO THE TOP SURVEY

MARCH, 2015

Matthew J. Pepper, Mark W. Ehlert, Eric S. Parsons, Seth W. Stahlheber, Susan F. Burns

Tennessee's Consortium on Research, Evaluation, and Development, established in 2010 as part of Tennessee's Race to the Top grant, is responsible for carrying out a detailed, focused program of research around key grant initiatives.

This research report provides a summary of *Findings from the 2014 First to the Top Survey*.

The views expressed herein do not necessarily reflect those of sponsoring agencies or individuals acknowledged. Any errors remain the sole responsibility of the authors.

Please visit www.tnconsortium.org to learn more about our program of research and recent publications.

Tennessee Consortium on Research, Evaluation, & Development

Peabody #44 • 230 Appleton Place • Nashville, Tennessee 37203

Phone: 615.322.5538 • Fax: 615.322.6018

www.tnconsortium.org



TABLE OF CONTENTS

Executive Summary	i
I. Introduction and Methodology	1
II. Implementation	8
III. Feedback	16
IV. Efforts to Improve	23
V. Teacher and Observer Perceptions of the Teacher Evaluation System	44
VI. Concluding Observations	51
Endnotes	55
Appendices	57

List of Tables

Table 1.1	Survey Module Assignments by Position Type	4
Table 1.2	Number of Responses and Response Rates, 2014 First to the Top Survey	5
Table 1.3	Summary of Sample Representativeness Tests, Teachers and Administrators	6
Table 3.1	Extent Teachers Asked to Share Teaching Strengths.	20
Table 4.1	Distribution of Teacher Self-Reported Improvement, by Evaluation Model	28

List of Figures

Figure ES1	Logic Model of How Teacher Evaluations Contribute to Teacher Development and Improve Student Outcomes	i
Figure 1.1	Logic Model of How Teacher Evaluations Contribute to Teacher Development and Improve Student Outcomes	2
Figure 2.1	Average Number of Times Teachers Were Observed in 2013-14 by Evaluation Model and Prior Year Evaluation Rating	9
Figure 2.2	Teacher-Reported Time Spent on Observation Activities: 2012-13 and 2013-14	10
Figure 2.3	Time Teachers Spent Throughout the Year Receiving/Reviewing Feedback, by Prior Year Evaluation Rating	11
Figure 2.4	Change in Principals' Allocation of Time for Various Job Duties	13
Figure 2.5	Who Observed Teaching and Teacher Perceptions of the Degree Feedback Was Useful.	14
Figure 3.1	Teacher and Observer Perceptions of the Primary Focus of Feedback from Evaluators.	17

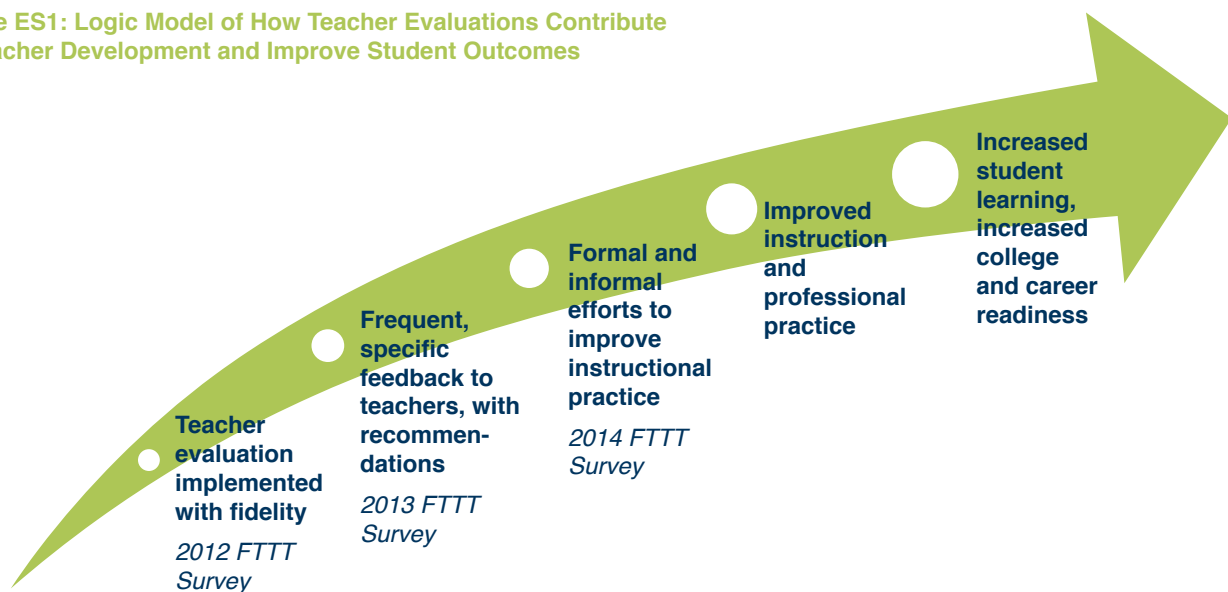
Figure 3.2	Frequency of Follow-Up by Observers on Areas Identified as Needing Improvement, by Prior Year Evaluation Rating	18
Figure 3.3	Teacher Reported Follow-Up by Observers on Areas Identified as Needing Improvement, by Evaluation Model	19
Figure 3.4	How Teachers Were Encouraged to Share Teaching Strengths	21
Figure 4.1	Number of Teachers Reporting They Worked to Improve Their Instructional Practice.	24
Figure 4.2	Reasons for Not Attempting to Improve Their Instructional Practice.	25
Figure 4.3	Number of Rubric Indicators Teachers Targeted for Improvement, by Evaluation Model.	26
Figure 4.4	Number and Percent of Rubric Areas Teachers Sought to Improve that were Identified as Needing Improvement Through the Teacher Evaluation Model	27
Figure 4.5	Areas of Instructional Practice TEAM Teachers Sought to Improve	29
Figure 4.6	Frequency TEAM Teachers Selected Each Rubric Indicator as their Most Improved Area of Instructional Practice	30
Figure 4.7	Primary Reason for Attempting To Improve in Area that Saw Most Improvement	31
Figure 4.8	Activities Pursued and Their Perceived Contributions to Improvement.	32
Figure 4.9	TEAM Teachers: Most Improved Area of Practice, Most Commonly Identified Improvement Activity, and Average Contribution of that Activity to Perceived Improvements	34
Figure 4.10	Written Goals for Improvement and Person Monitoring Progress Toward Them.	35
Figure 4.11	Sources of Evidence Used to Gauge Improvement and Degree Each Supported Improved Practice	36
Figure 4.12	Number of Teachers Who Used Various Sources of Evidence to Gauge Improvement, Number Who Reviewed Evidence with Observer(s), & Average Degree Evidence Supported Improvement, by Reviewed Status	38
Figure 4.13	Most Difficult Areas to Improve Reported by TEAM Teachers	39
Figure 4.14	Agreement with Reasons Areas Were Difficult to Improve, by Evaluation Model	40
Figure 4.15	Areas TEAM Teachers Found Most Difficult to Improve, Average Amount of Improvement in those Areas, and Reason Teachers Most Commonly Agreed or Strongly Agreed Why Area was Difficult to Improve	41
Figure 5.1	Teacher and Observer Responses to 2014 First to the Top Perceptual Questions	45
Figure 5.2	Disaggregated Results for Teacher Level of Satisfaction with the Evaluation Process Used in Their School.	47
Figure 5.3	Disaggregated Results by Model for Teacher Satisfaction with the Evaluation Process Used in Their School.	49

EXECUTIVE SUMMARY

The 2013-14 school year was the fourth and final year of Tennessee’s \$501 million Race to the Top grant. The Tennessee Consortium on Research, Evaluation and Development (the Consortium) serves as the lead external evaluator of the initiatives funded by the federal grant and has implemented an annual statewide First to the Top (FTTT) Survey to examine educator experiences with and perceptions of these improvement efforts.¹ Since 2012 Consortium researchers have focused annual surveys on changes to teacher evaluations that have been implemented as part of the FTTT improvement effort, as all of the state’s professional educators have been affected by this policy initiative.²

Figure 1 presents a logic model of teacher evaluation as a method for increasing student learning through the development of teaching skills. Consortium researchers have used this model to inform the content of FTTT surveys, with each survey including a new section of questions focused on one step of this theory of action. The 2012 survey focused on fidelity of implementation and the 2013 survey focused on the content, quality, and frequency of evaluation feedback. The 2014 version of the FTTT survey included new questions on teacher efforts to improve instructional practice. Teachers provided information about the practices they attempted to improve, their motivations for trying to improve, activities undertaken as improvement efforts, the extent they believe improvement was accomplished, and the evidence used to gauge improvement. Also, each FTTT survey has included items measuring educator perceptions of the quality, impacts, and value of the evaluation system in their school.³

Figure ES1: Logic Model of How Teacher Evaluations Contribute to Teacher Development and Improve Student Outcomes



Results from the 2014 FTTT survey identify reasons to celebrate and areas for improvement. Observation and scoring processes continue to be implemented with fidelity, and 50% of teacher respondents perceive the feedback received from evaluations to be primarily focused on helping them improve professional practice.⁴ In addition, nearly all teachers indicated that they attempted to improve their instructional practice during the 2013-14 school year – often in response to feedback from evaluations – and believed those efforts led to improvements. Teachers most frequently attempted to improve instruction through activities related to developing and delivering instruction such as reviewing and revising lessons and implementing new learning activities in classrooms. These efforts, often motivated by factors other than teaching observations and evaluations, frequently were not guided by written goals or systemically monitored by evaluators. These findings suggest that while teacher evaluations are producing measures of teaching performance, teachers are often on their own as they attempt to improve their instructional practice. Finally, teacher support for reformed evaluation processes did not increase in 2014 as it did in 2013, with 47% of teachers expressing satisfaction with the system overall and 54% believing that the evaluation process will improve their teaching. Key findings supporting this examination of teacher evaluation in Tennessee are presented below.

While teacher evaluations are producing measures of teaching performance, teachers are often on their own as they attempt to improve their instructional practice.

THE IMPLEMENTATION OF TENNESSEE'S EVALUATION PROCESS

Tennessee's evaluation systems continue to be implemented with fidelity, and evaluation processes are becoming routinized.

A consistent finding from the 2012 and 2013 FTTT surveys was that the evaluation process has been implemented with a high level of fidelity, and this trend continued in 2014.⁵ Teachers were observed the prescribed number of times by trained observers, primarily principals and assistant principals. Evaluation scores were calculated as outlined by statute and state policy, results were logged in a statewide database, and outcomes were shared with teachers. Survey responses reveal that teachers and administrators spent similar amounts of time on evaluation tasks in 2014 as in 2013. On average in 2014, teacher respondents indicated they dedicated eight hours **per year** and administrator respondents indicated they dedicated nine hours **per week** on evaluation activities. Administrators also indicated that their allocation of time and effort to various job duties changed little from 2013 to 2014 (i.e., only a small percentage of administrators indicated that they had increased the time and effort devoted to evaluation tasks or that they devoted less time to other administrative duties). These findings suggest that teacher evaluation processes are becoming part of the regular work routine for teachers and administrators.

FEEDBACK FROM TEACHER EVALUATIONS

About half of teachers perceived the feedback they received from teacher evaluations to be primarily focused on helping them improve their teaching. However, teachers continued to report that observers did not consistently follow up on areas identified as needing improvement.

One of the most important findings from the 2013 FTTT survey (described in a report linked here) was that teachers who perceived evaluation feedback to be primarily focused on improvement were more likely to have positive perceptions of the evaluation system. Results from the 2014 FTTT survey found a three percentage point increase in the share of teachers indicating that their feedback was primarily focused on improvement (to 50%) and a small decrease in the percent who perceived feedback as primarily focused on judging their performance (to 19%).

Most teachers indicated that observation feedback identified teaching strengths (96%) and areas needing improvement (86%).⁶ Teachers were more likely to report that feedback from teaching observations was more useful when they were observed multiple times by individuals in the same position (usually their principal or assistant principal). Similar to findings from the 2013 survey, a third of teachers indicated they did not receive any follow-up from observers on areas identified as needing improvement, and nearly another third indicated their observer followed up only once. Only 33% of teachers reported having written goals for improving in the area they believe they improved the most during 2014, and generally, teachers reported that they seldom reviewed evidence of improved practice with evaluators. Finally, more than half of all teachers, and 42% of teachers who scored a 1 on their previous year's overall evaluation, indicated that they had spent less than an hour during the school year receiving and reviewing feedback from observations. These results indicate that observation follow-up and follow-through activities between observers and teachers, which have been shown to contribute to improved student achievement, are not consistently occurring in Tennessee schools.⁷

TEACHER EFFORTS TO IMPROVE INSTRUCTION

The overwhelming majority of teachers worked to improve their instructional practices in 2014 and also reported that they improved in nearly all areas they targeted. The most frequently reported activities teachers pursued reflect ongoing and routine instructional work, independent learning, and informal interactions with peers.

Building on the survey's logic model, the primary focus of the 2014 FTTT survey was on teacher efforts to improve instruction. In short, researchers wanted to know what areas of practice teachers attempted to improve, whether improvement efforts were motivated by feedback from observations, how much teachers believed their efforts improved practice, and the evidence used to gauge success. Nearly all teachers (98%) reported that

they attempted to improve their instructional practice in 2013-2014, and about half focused their improvement efforts on five or fewer specific rubric indicators. On average, each teacher reported that 40% of the indicators they addressed had been identified as needing improvement during the evaluation process. Teachers generally reported improving “some” or “a lot” in the areas they worked on, regardless of the number of rubric indicators selected.

When asked about their most improved area of instructional practice, teachers most frequently reported pursuing improvement efforts that reflect daily work tasks, such as implementing new learning activities and reviewing/ revising lesson plans. These were the same activities observers indicated they recommended most frequently

Teachers most frequently reported pursuing improvement efforts that reflect daily work tasks, such as implementing new learning activities and reviewing/revising lesson plans.

during discussions with teachers about areas needing improvement. Other commonly pursued activities reflect actions that could be completed independently, including self-directed learning and informal discussions with peers. The least frequently pursued activities involve structured interactions with peers or formal professional development sessions. Teachers indicated that all activities they pursued contributed “some” or “a lot” to improved practice. Teachers most frequently cited a lack

of quality training and learning opportunities and a lack of time as barriers to improvement.

Teachers most frequently used evidence that could be gleaned during regular classroom instruction to gauge their improvement (e.g., student interest and classroom performance). Only about one in five teachers reviewed evidence of improvement with an observer, even though teachers perceived evidence as providing more support for improved practice when this occurred.

PERCEPTIONS OF AND OPINIONS ABOUT TEACHER EVALUATION SYSTEMS

Teacher perceptions of Tennessee’s evaluation processes in 2014 were similar to their 2013 perceptions.

Overall, just under half of teachers reported that they were satisfied with the teacher evaluation process used in their school.

Items to gauge teacher and observer perceptions of the teacher evaluation process have been included in all annual FTTT surveys. Comparisons of results from the 2012 and 2013 FTTT surveys showed large positive shifts in teacher perceptions of the quality, value, and potential impacts of teacher evaluations. Levels of agreement with most of these evaluative items in 2014 were within one or two percentage points of the levels observed in 2013 and in most cases indicated slightly less favorable perceptions and opinions. Observers continued to perceive

teacher evaluation processes much more positively than teachers. For example, 79% of observers agreed or strongly agreed that they were satisfied with the teacher evaluation process while only about half of teachers agreed with the same statement.

LOOKING TO THE FUTURE

The logic model underlying Consortium researchers' conceptions of how evaluation can influence the development of teaching skills to improve student learning suggests that teacher efforts to improve instruction lead to changes in their classrooms. In future work, researchers recommend investigating specific changes in the strategies, activities, materials, and assessments students experience in their classrooms, especially those implemented in response to evaluation feedback. This link is focused on changes that occur in the "black box" of classroom instruction and would allow researchers to connect changes in classroom activities to changes in student achievement and other outcomes.

Finally, it is also important to investigate how teacher evaluation processes influence overall teaching quality over a longer time period through teacher selection and retention mechanisms. Consortium researchers have conducted some preliminary work in this area, but further work is recommended to investigate variation by district and to quantify the impact across the state.

I. INTRODUCTION AND METHODOLOGY

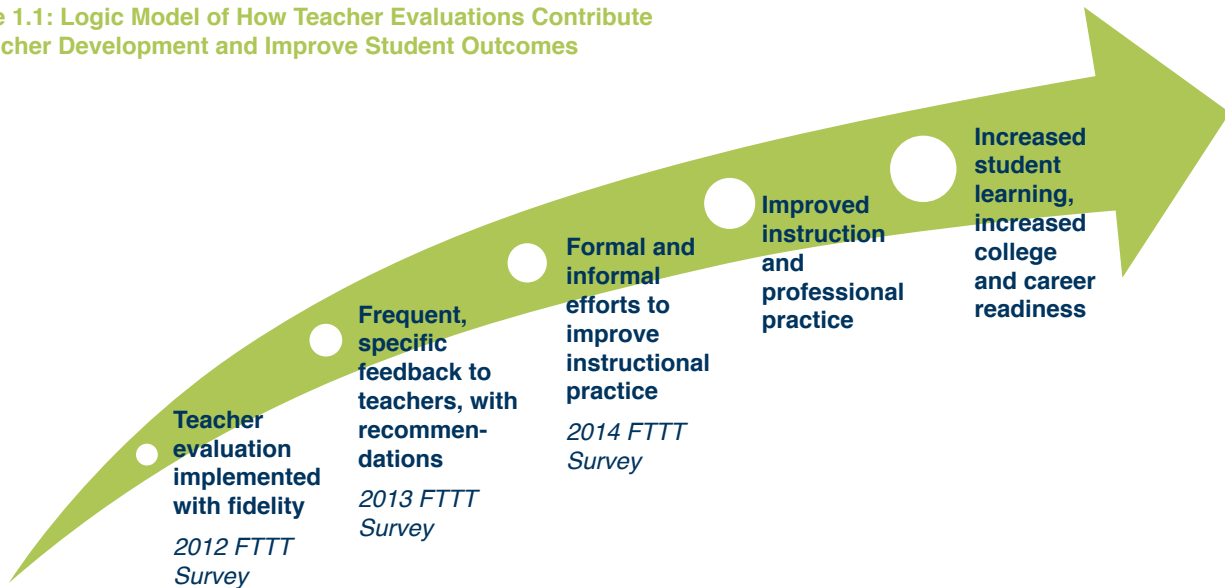
INTRODUCTION

As one of only two states to be awarded a grant in the first round of the United States Department of Education's 2010 Race to the Top competition, Tennessee has just completed the final year of its four-year, \$501 million award. The reward committed the state to implement a variety of reforms, with one of the earliest being the significant revision of Tennessee's educator evaluation processes. In order to increase Tennessee's grant competitiveness, many reform elements were written into statute, increasing the likelihood that the state's major policy shifts around educator evaluation will persist. The Tennessee Consortium on Research, Evaluation, and Development (the Consortium)⁸ continues to examine educator perceptions of these reforms through the First to the Top Survey, administered annually in the spring since 2011.⁹ This annual administration allows for monitoring changing perspectives over time, as well as adjusting to changes to evaluation policies and procedures.

Five teacher evaluation systems are used in Tennessee school districts. While TEAM is the predominant evaluation system across the state (roughly 80% of school districts), the survey also reports findings from districts using four other models: TEM, TIGER, COACH, and AFET.¹⁰ By law, teacher evaluation ratings must be a factor in human capital decisions including, at minimum, tenure, promotion, retention, termination, and compensation. The new evaluations yield a final rating for each teacher stemming from three different sources: student growth data (e.g., TVAAS), student achievement data (e.g., TCAP scores, graduation rates), and qualitative data such as observations of teaching, a review of prior evaluations, and personal conferences.¹¹ Teaching observations are scored using classroom observation instruments or rubrics developed for each approved evaluation model.¹²

The Consortium's initial FTTT survey in 2011 focused broadly on Race to the Top reforms, including teacher evaluation. Evaluation-relevant questions primarily addressed issues of implementation and baseline teacher and observer perceptions. Beginning with the 2012 First to the Top Survey, the first year of statewide implementation, Consortium researchers focused each annual survey on a successive step in an implicit logic model of teacher evaluation. Figure 1.1 details this logic model and the focus of each annual survey.

Figure 1.1: Logic Model of How Teacher Evaluations Contribute to Teacher Development and Improve Student Outcomes



While Tennessee’s evaluation systems can improve student learning by informing decisions on teacher retention and dismissal, the survey focuses on instructional improvement through evaluation, as the vast majority of teachers have been impacted by FTTT through this pathway. The 2012 First to the Top Survey primarily focused on the fidelity of implementation of the evaluation process, investigating the preparedness of observers, the number of observations received by teachers, completion of each step in the observation process, and personnel issues (who conducted observations and how it impacted their other job responsibilities). Broadly speaking, the Consortium found that Tennessee’s evaluation systems were being implemented as designed. With this knowledge, the 2013 FTTT Survey turned to the content, quality, and frequency of evaluation feedback received by teachers. Respondents indicated significant variation across all three elements of feedback (content, quality, and frequency), with very little of this variation explained by the previous year’s evaluation score or experience level of the teacher.

The 2014 First to the Top Survey, and this report, focuses on the next step in the logic model: exploring the ways in which teachers respond to and act on feedback from their evaluation process and attempt to improve their instruction. More specifically, this report will examine the focus and specificity of feedback, the steps teachers take to address areas needing improvement, whether improvement efforts are formal or informal, the frequency and procedures of observer follow-up, the extent that teachers believe efforts resulted in improved instruction, and the evidence teachers used to gauge the success of improvement efforts.

The majority of the findings presented in this report are based on responses from teachers who indicated they were observed as part of the teacher evaluation process used in their schools, and building administrators

and instructional coaches who conducted observations. While a significant number of teacher respondents indicated that they conducted evaluative teaching observations, their responses have been excluded from most presentations of “observer” results.¹³ Additional information regarding the sampling process and sample representativeness can be found in the following subsection, which details methodology.

Sections two through five of the report present the core of the report’s findings. Revisiting the areas of focus in the 2012 and 2013 First to the Top Survey Reports, section two briefly investigates the fidelity of implementation of the evaluation processes in 2013-14, and section three examines the content, quality, and frequency of feedback received by teachers. Section four presents our findings focused on the mechanisms of teacher improvement efforts. Section five then presents the teacher and observer longitudinal perceptions of the teacher evaluation process. The report closes in section six, with concluding observations and implications.

METHODOLOGY

The 2014 FTTT survey was administered following procedures developed for the 2012 and 2013 FTTT surveys. The process requires developing a sampling frame, randomly assigning members of the sampling frame to receive one of six different modules with survey questions focused on other important reform initiatives included in Tennessee’s Race to the Top grant proposal, and emailing survey invitations to members of the sampling frame. Those procedures, a summary of response rates, and the representativeness of the sample of respondents are summarized in this section of the report. A more thorough description of procedures and sample representativeness can be found in Appendix B.

Survey Administration

As was the case in 2012 and 2013, all certified school staff members listed in the Tennessee Department of Education’s (TDOE) student management system, the Education Information System (EIS) were considered potential survey takers. Data elements from EIS, the CODE evaluation database, and other state level data sources maintained by TDOE were combined to build a file containing a single record for nearly every teacher, administrator, and other professional staff member in the state’s public schools. This file included 77,802 Tennessee educators and was the sampling frame for survey administration. However, only educators in EIS for whom a position title and email address could be determined were invited to take the 2014 survey, which was again administered online. Each complete sampling frame record included the educator’s email address, the school and district where the educator worked in 2013-14, the evaluation system used in that district, and selected demographic, experience, and education level variables. 63,472 teachers, 3,535 administrators, and 2,189 certified support staff (Library Media Specialists, Counselors, and Psychologists) received survey invitations. Of

that number, 26,589 teachers (41.9 percent), 1,529 administrators (43.3 percent), and 838 certified support staff (38.3 percent) responded to the survey.

Sampling frame records were randomly assigned to receive one of six different versions of the FTTT survey where each version contained a different “module” or set of questions focused on a key reform initiative. The six 2014 modules were: Great Teachers and Leaders; Professional Development; Assessment; Standards & Assessment and Knowledge of & Attitudes Toward Reform; Instructional Practices and Response to Intervention; and Teacher Compensation. The results of this assignment are shown in Table 1.1.

Randomization worked as intended, assigning roughly an equal number of invitations for each version, with the exception of the compensation module (the compensation module was not assigned to educators in schools involved in strategic compensation initiatives such as the Innovation Acceleration Fund (IAF) and the Tennessee Teacher Incentive Fund (TIF) grant, and so this difference was by design). Based on these randomization results and a detailed investigation into the characteristics of each subsample (presented in detail in Appendix B), Consortium researchers conclude that module subsamples are equivalent and generally representative of the underlying populations of teachers in Tennessee.

Table 1.1: Survey Module Assignments by Position Type

POSITION	Module						Total
	Leadership	Prof. Dev.	Assessment	Standards	RTI	Compensation	
TEACHERS	10,831	10,865	10,847	10,891	10,794	9,244	63,472
ADMINISTRATORS	581	594	612	582	635	531	3,535
CERTIFIED SUPPORT STAFF (COUNSELOR/PSYCHOLOGIST)	384	382	383	358	374	308	2,189
OTHER	437	397	401	386	408	318	2,347
TOTAL	12,233	12,238	12,243	12,217	12,211	10,401	71,543

Each email invitation contained a unique link for accessing the survey administration site. Invitations were sent on April 2, 2014, and reminder emails were sent on a weekly basis to those who had not yet responded or opted out. The survey closed at midnight on May 31. Educators who used their email link to respond to the survey were directed to the appropriate version of the 2014 FTTT survey.

The first few questions on the FTTT survey asked survey takers to identify their position in the school and most were then asked to indicate whether they had conducted teaching observations during 2013-14 as part of the teacher evaluation process used in their school. Any respondent who answered this question was considered a survey respondent. Table 1.2 presents response rates by position category and shows a relatively robust response rate for all groups.

Table 1.2: Number of Responses and Response Rates, 2014 First to the Top Survey

POSITION	Number Invited	Number of Respondents	Response Rate
TEACHERS	63,472	26,589	41.9%
ADMINISTRATORS	3,535	1,529	43.3%
CERTIFIED SUPPORT STAFF (COUNSELOR/PSYCHOLOGIST)	2,189	838	38.3%
OTHER	2,347	990	42.2%
TOTAL	71,543	29,946	41.9%

Representativeness of Respondent Sample

The large sample of survey respondents lends confidence to the reliability of survey results. However, the degree to which results reflect the opinions of the population of educators in Tennessee depends in part on whether the characteristics of survey respondents are similar to the characteristics of the broader educator population. Table 1.3 summarizes the results of tests of the representativeness of teachers and building administrators who responded to the 2014 FTTT survey. There is a slight under-representation of novice teachers and those holding only a bachelor’s degree (groups that contain a great deal of overlap). Female and elementary school teachers are also somewhat over-represented in the sample, but only by a few percentage points.

Table 1.3: Summary of Sample Representativeness Tests, Teachers and Administrators

GROUP / CHARACTERISTIC	Too Many (Overrepresentation)	Too Few (Underrepresentation)	Implications
TEACHERS			
EXPERIENCE	Late career (15+ years)	Novice (< 5 years)	Responses more critical
TIER (School type)	Elementary	Secondary	Responses more positive
DISTRICT SIZE	Smallest districts	Largest districts	No clear patterns
GENDER	Females	Males	Responses more positive (Possibly a function of tier)
ADMINISTRATORS			
EXPERIENCE	Late career	Early career	No clear patterns
TIER (SCHOOL TYPE)	Elementary	Secondary	No clear patterns
DISTRICT SIZE	Smallest districts	Largest districts	No clear patterns

Most characteristics of administrators who responded to the survey were different from the characteristics of administrators in the sampling frame, although these differences were relatively small. The only exception is the variable that categorizes the size of the district in which the administrator works, with administrators from smaller districts more likely to respond than administrators from larger districts. However, the implications of this are unclear, as there are no clear patterns based on district size in FTTT survey results from prior years. A full discussion of representativeness results, including examination of representativeness within each evaluation model and CORE Region, can be found in Appendix B.

Limitations

Readers should note that survey responses only include the perceptions of two out of every five Tennessee teachers. The number of teacher responses and the relatively small differences between the sample and teacher universe, however, leads researchers to conclude with reasonable confidence that reported results are representative of Tennessee teachers.

Results presented in this preliminary report are descriptive in nature - primarily frequency counts and related percentages. Some simple relationships are examined through cross-tabulations, but these, by definition, only examine two variables at a time, which hinders trying to determine the unique contribution of one predictor variable on outcomes of interest. These descriptive statistics can suggest relationships between key variables, but, in isolation, cannot fully explain variation in responses. Finally, readers are encouraged to remember that descriptive statistics cannot prove causation.

II. IMPLEMENTATION

INTRODUCTION

One of the consistent findings from the 2012 and 2013 First to the Top (FTTT) surveys was the high level of fidelity with which evaluation systems were implemented. Survey results indicated that during the 2011-12 and 2012-13 school years teachers were observed the expected number of times by trained evaluators. Nearly all observers indicated that they had attended evaluator training, and over ninety percent of observers reported on the 2013 survey that they were well prepared to conduct evaluation tasks.¹⁵ This is important because one must first verify that evaluation systems are being implemented as designed before investigating their impact.

Implementation was not a primary focus of the 2014 FTTT survey; however, some implementation questions were included in order to continue tracking longitudinal trends. As in earlier years, results from the 2014 survey show a high level of implementation fidelity of Tennessee's teacher evaluation systems. Teachers are being observed the expected number of times, primarily by principals and assistant principals. Additional analyses reveal that teachers reported spending about eight hours **per year** on observation-related tasks in 2013-14, with patterns of time spent across observation tasks very similar to findings from 2012-13. The amount of time administrators spent on various tasks also appears to have stabilized at about nine hours **per week**, again very similar to results observed in 2012-13. In summary, findings from the 2014 FTTT survey suggest that Tennessee's evaluation systems continue to be implemented with fidelity, and longitudinal results suggest that the time commitment of the evaluation processes on teacher and administrator schedules has stabilized, with administrators spending much more time on evaluation than teachers.

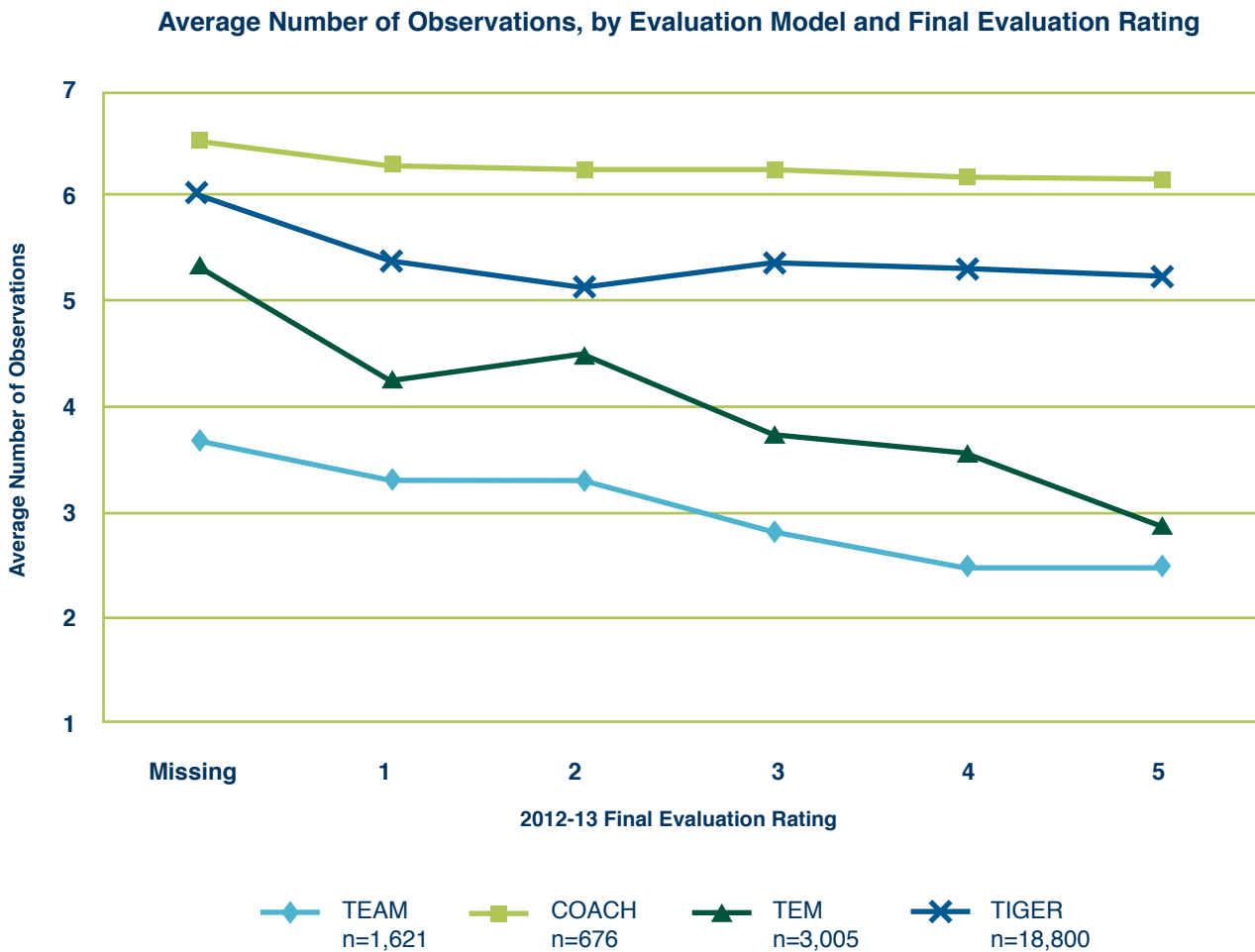
NUMBER OF EVALUATIONS

Procedures concerning the prescribed frequency of teacher observations vary by evaluation model and for TEAM – Tennessee's default evaluation system – also by licensure status.¹⁶ One of the more significant policy changes to TEAM since its 2011-12 statewide roll-out was a modification that decreased the required number of observations for teachers who met specific scoring and license criteria. Shelby County's alternative teacher evaluation system, the Teacher Effectiveness Measure (TEM), also varies the required number of observations based on previous year evaluation scores, as do districts using the Teacher Instructional Growth for Effectiveness and Results (TIGER) model. Districts using COACH require the same numbers of observations for all teachers.

The 2014 FTTT survey asked teachers to indicate the number of times they were observed as part of the teacher evaluation process, and their responses are presented by the model used and previous year evaluation score in

Figure 2.1.¹⁷ As expected, results from both the TEAM and TEM model show a negative relationship between the number of observations teachers received and their previous year evaluation scores, while results from teachers in districts that use the COACH model do not.¹⁸ Respondents using the TIGER model are difficult to interpret, but show a slight negative relationship. Note that teachers in the “Missing” evaluation score category were likely new teachers in 2014, which helps explain why they generally had more observations than teachers with prior year evaluation scores. Overall, it appears that teachers are being observed at the expected frequency consistent with the previous year evaluation score.

Figure 2.1: Average Number of Times Teachers Were Observed in 2013-14, by Evaluation Model and Prior Year Evaluation Rating



TIME SPENT ON OBSERVATION ACTIVITIES, TEACHERS

The 2014 FTTT survey asked respondents to indicate the amount of time they spent throughout the entire school year on these four tasks: preparing for observations, participating in pre-conferences, being observed, and receiving and/or reviewing feedback from observations. Note that this last category is broadly described so that teachers include both the time spent reviewing feedback during observation post-conferences and time spent reviewing feedback on their own. Figure 2.2 displays the results from these questions from both the 2014 and 2013 FTTT surveys.

Figure 2.2: Teacher-Reported Time Spent on Observation Activities: 2012-13 and 2013-14

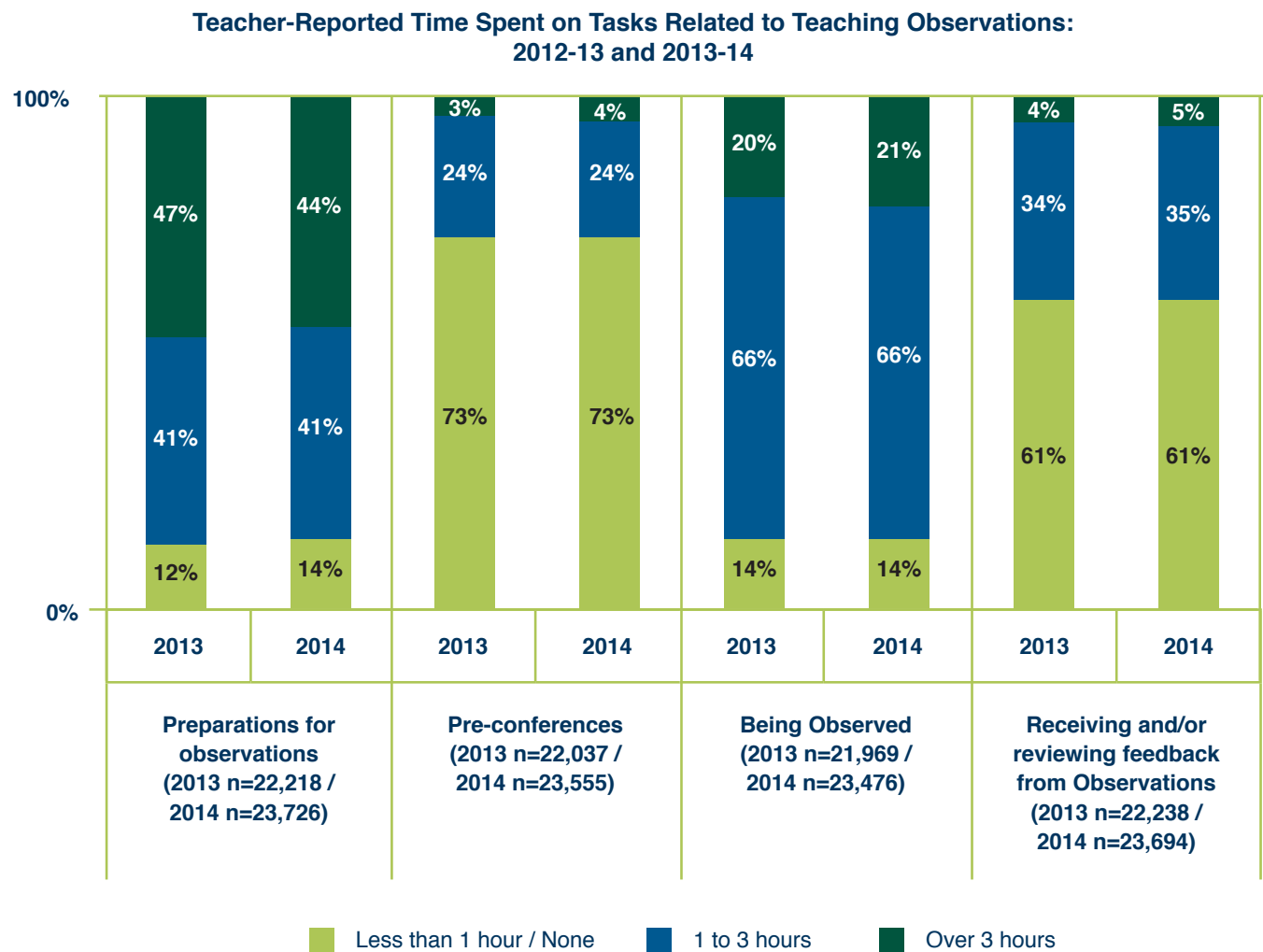
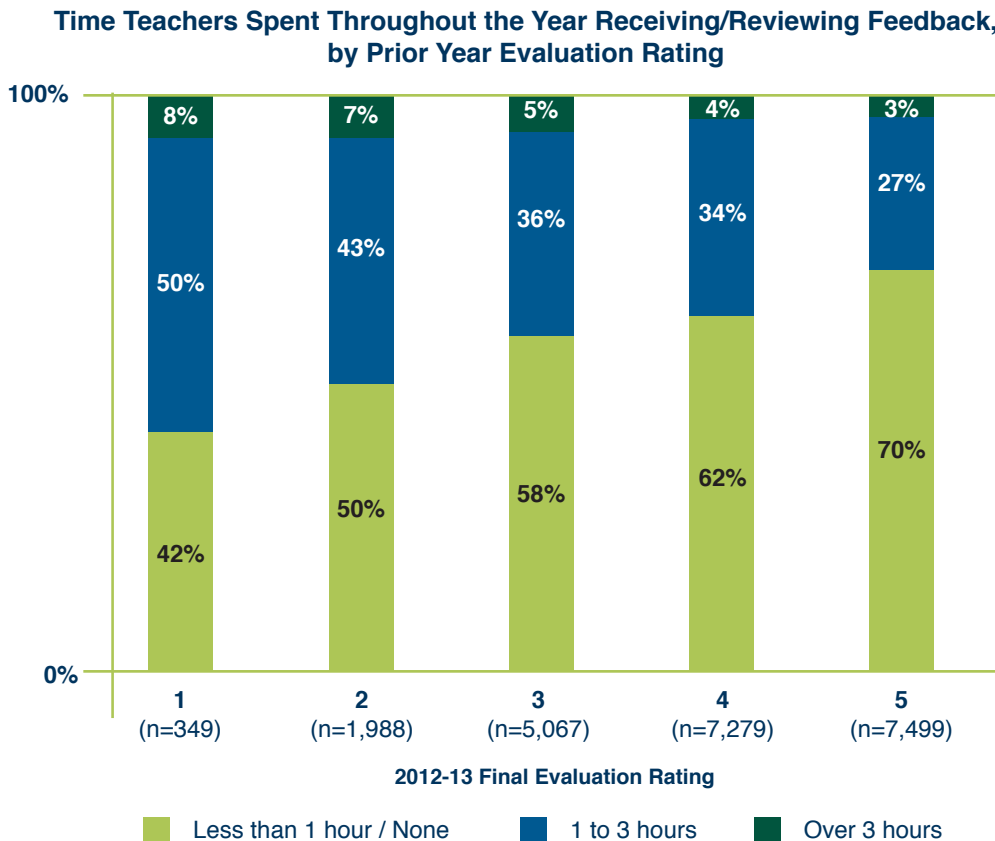


Figure 2.2 reveals similar results from the two years of survey data, with only a few percentage points difference for any observation task across the years. Teachers continued to indicate that preparing for observations took the largest amount of time, with 44% of respondents indicating that they spent over three hours during the 2013-14 school year preparing for observations. Pre-conferences took the least amount of time, followed by the time spent receiving and/or reviewing feedback.

Consortium researchers explored the relationship between the time teachers reported spending on receiving and reviewing feedback and their previous year evaluation score. One might surmise that teachers who had struggled in the previous year would have been provided more attention from evaluators. Indeed, this should occur by design, with low-scoring teachers in the two most common evaluation models – TEAM and TEM – required to receive more observations than higher-scoring peers. Figure 2.3 shows teacher-reported time spent receiving and reviewing observation feedback by their previous year evaluation score. Generally, teachers with lower prior year evaluation scores spent more time with observation feedback than teachers with higher prior year evaluation ratings. Note, however, that 50% of teachers that scored a two in the previous year and 42% of teachers who scored a one in the previous year reported that they spent less than an hour receiving and/or reviewing feedback.

Figure 2.3: Time Teachers Spent Throughout the Year Receiving/Reviewing Feedback, by Prior Year Evaluation Rating



TIME SPENT ON OBSERVATION ACTIVITIES, ADMINISTRATORS

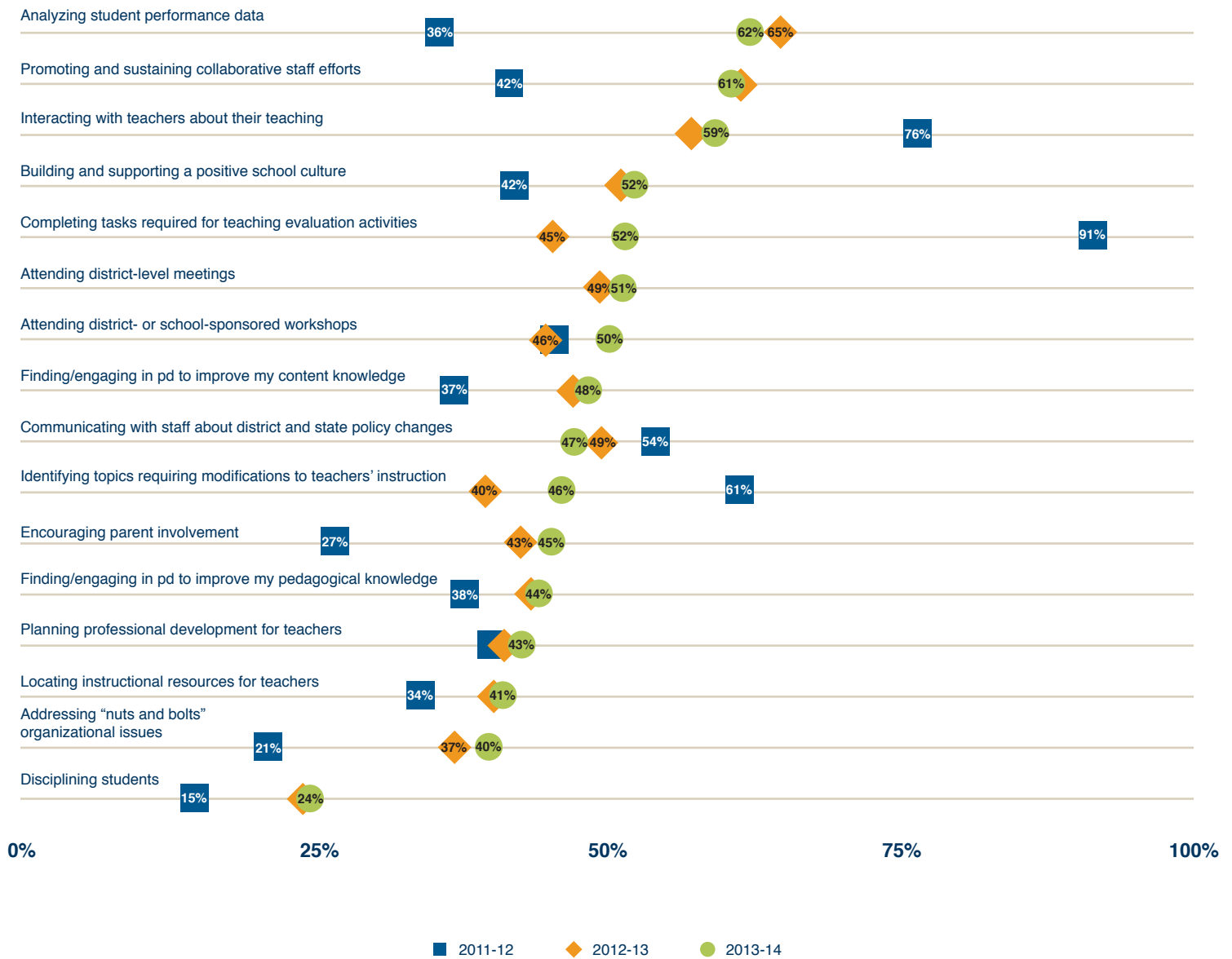
Principals and assistant principals were also asked how much time they spent each week on teacher evaluation activities. As with teacher responses, there was little change in this measure from 2013 to 2014. On the 2014 survey, 38% of administrators indicated that they had spent, on average, 6 hours or less weekly on evaluation activities, 43% indicated spending 7 to 12 hours, and 18% indicated that they spent 13 or more hours weekly. These percentages are within two points of 2013 findings.

The 2014 FTTT survey also continued to look more broadly into how teacher evaluation requirements might change the time administrators spent on various responsibilities. The first year of revisions to the teacher evaluation process was 2011-12, and, not surprisingly, survey results from that year revealed that almost all administrators claimed that they spent more time on teacher evaluation compared to the previous year. In each successive year the FTTT survey repeated this question, asking administrators to report whether they spent more or less time than in the prior year on a broad list of typical job duties. Results from 2011-12 through 2013-14 are shown in Figure 2.4. Each data point plotted in the figure indicates the percent of administrators who claimed that they spent more time on the listed activity in the current year compared to the previous year. One of the most apparent patterns is the difference between results in 2011-12 (shown with square markers) and results from 2012-13 and 2013-14 (shown with diamond and circle markers, respectively). The first year of the revised teacher evaluation system saw significant changes in the time administrators spent on various administrative tasks, with 91% stating that they spent more time on teacher evaluation activities compared to the previous year. Interacting with teachers about their teaching – an item related to evaluation – was also an outlier.

Since 2011-12, however, the changes in time and effort devoted to various administrator job responsibilities appears to have lessened, with fairly consistent values for the amount of change in 2012-13 and 2013-14. Additionally, administrators reported that they continued to spend more of their time analyzing student performance data, promoting and sustaining collaborative staff efforts, and interacting with teachers about their teaching.

Figure 2.4: Change in Principals' Allocation of Time for Various Job Duties

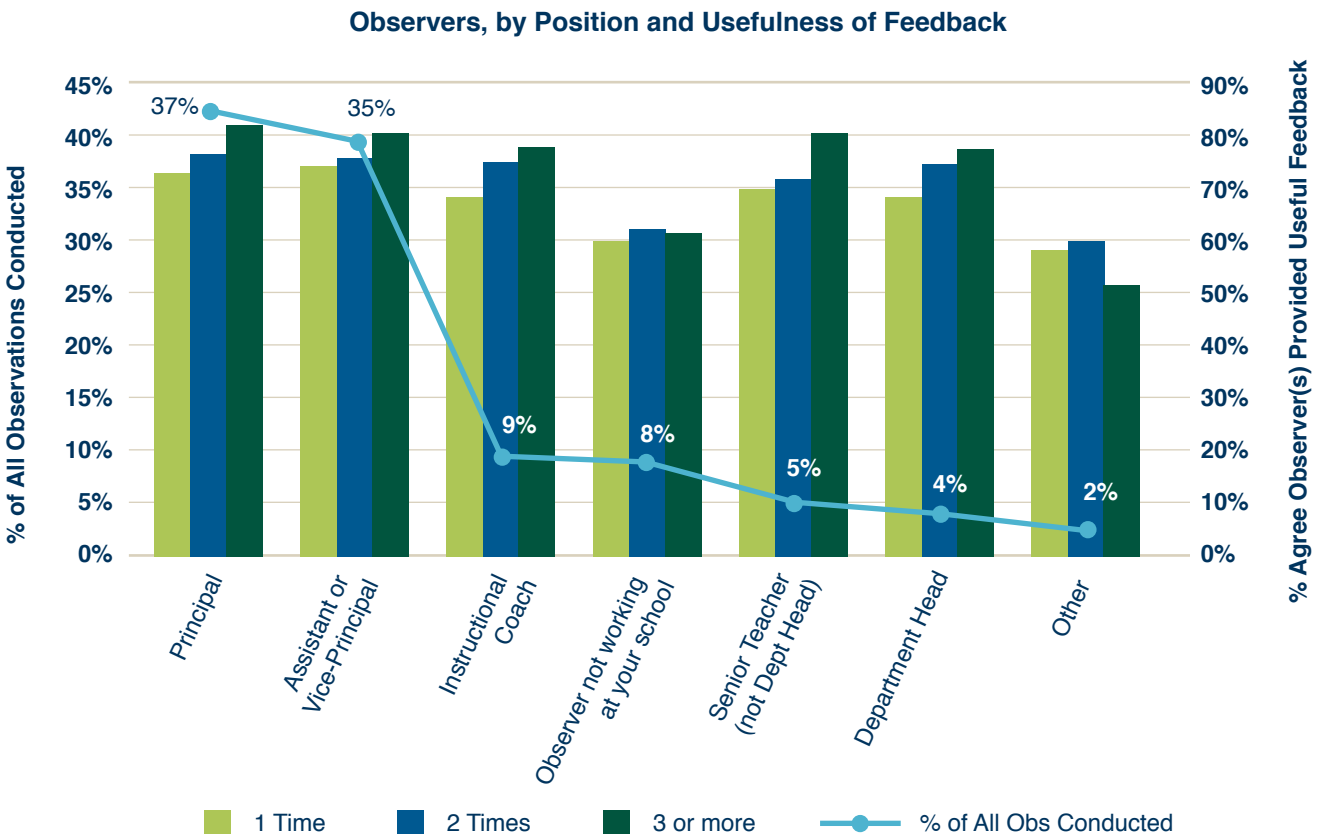
Percent of Principals Indicating They Spent MORE TIME on the Listed Activity Compared to the Previous Year: 2011-12 through 2013-14



OBSERVER POSITIONS AND THE PERCEIVED VALUE OF THEIR FEEDBACK

Finally, researchers examined who conducted teacher observations and the extent to which teachers perceived feedback from those observers to be helpful. Districts use individuals in a variety of positions to conduct observations, including administrators, coaches, mentor teachers, central office employees, and department heads. Teachers were asked to indicate how many times individuals in various positions had conducted observations of their teaching, and the results from this question were used to estimate the total share of teaching observations conducted by people in those positions. This is shown by the line in Figure 2.5 using the vertical scale on the left-hand side. According to teacher respondents, just over 37% of the total observations were conducted by a principal and nearly 35% were conducted by an assistant principal. Less than ten percent of observations were conducted by an instructional coach or observer not working at the same school as the teacher, and less than five percent were conducted by a senior teacher and department head. As in previous years, administrators continued to conduct most of the observations in 2013-14 with some assistance provided by instructional coaches and others.

Figure 2.5: Who Observed Teaching and Teacher Perceptions of the Degree Feedback Was Useful



The bars in Figure 2.5 present two additional pieces of data: the percent of teacher respondents who agreed or strongly agreed that the observer(s) in the listed position provided useful feedback, broken down by the number of times they were observed by someone in that position. Those percentages are scaled against the right-hand vertical scale. The first takeaway, from the height of the bars, is that teachers perceived feedback from observers not working at their school and individuals in the “Other” category to be slightly less useful than feedback from administrators, coaches, senior teachers, and department heads. Perhaps surprisingly, feedback from instructional coaches was perceived to be as useful as feedback from administrators.

Second, teachers perceived feedback to be more useful when they were observed more frequently by people in the same job category. For example, 7% fewer teachers who had been observed by a principal only once perceived the feedback to be useful compared with teachers who had been observed by a principal three or more times. This may be due to a greater level of comfort and/or familiarity with a specific observer, or that multiple observations allow for greater continuity in feedback and discussions about teaching practice. Disentangling the root cause of this relationship is not possible with these survey questions, but researchers note that this relationship exists for observers from all job categories except for those not at a teacher’s school and the catch-all “Other” category.

SUMMARY

The 2014 FTTT survey found that Tennessee’s teacher evaluation systems continued to be implemented with fidelity across the state, and that the time devoted to this process by both teachers and administrators has stabilized. As expected, teachers with a lower evaluation score from the previous year reported spending more time receiving and reviewing evaluation, although a significant percentage of low-scoring teachers reported spending less than an hour reviewing feedback throughout the entire year. Finally, principals and assistant principals continued to conduct the majority of observations, with some assistance provided by instructional coaches and teachers/department heads. Taken as a whole, these results paint a picture of an evaluation system that is becoming institutionalized within Tennessee.

III. FEEDBACK

Observation post-conferences are a designated forum for sharing feedback, providing opportunity for observers to review scoring, identify strengths and weaknesses, and suggest resources to address areas of concern. Earlier First to the Top Surveys reveal that the perceived focus of evaluator feedback is related to teacher perceptions of and attitudes about the evaluation system. Two important predictors of how teachers perceive the primary focus of their evaluation feedback are whether teachers were asked to share their teaching strengths with others and the extent to which observers follow up with teachers concerning areas identified as needing improvement. This chapter explores the characteristics of feedback teachers received from their observers and whether that feedback was perceived as primarily focused on instructional improvement, making a judgment, or both goals equally.

PRIMARY FOCUS OF FEEDBACK

In teacher evaluation, there is an inherent tension between judging performance and improving future practice. Like many states, Tennessee's evaluation systems build in elements relevant to both goals. For example, the TEAM model requires observers to identify after each teaching observation an area that the observer identifies as needing improvement. Simultaneously, per state statute, evaluation results can also be used to remove teacher tenure and as a cause of dismissal.¹⁹

Like its predecessors, the 2014 FTTT survey asked teachers if they perceived evaluator feedback to be focused more on improvement, more on judgment, or equally focused on both goals. In earlier surveys, responses to this question were found to be significant predictors of teacher perceptions of the evaluation system. For example, on the 2013 FTTT survey 67% of teachers who perceived feedback to be primarily focused on improvement agreed or strongly agreed with the statement *Overall, I am satisfied with the teacher evaluation process used in my school*. Just over half of teachers who perceived feedback to be equally focused on improvement and judgment agreed with this statement, while only 18% of teachers who perceived feedback to be more focused on judgment did so. Additionally, Consortium researchers found strong relationships between teacher responses to this question and the extent that observers extended the post-conference conversation to later, follow-up interactions.

Figure 3.1 summarizes teacher and principal perceptions of the primary focus of feedback in each year of the survey. There was a significant shift in the perceptions of teachers toward a focus on improvement from 2012 to 2013. This trend continued in 2014 but to a much smaller extent, with half of teachers reporting they perceived the primary focus of the feedback to be on improving teaching. An additional third perceived feedback to be equally

focused on improvement and judgment, and about one out of five perceived feedback to be focused primarily on making a judgment about their performance. Compared to the significant shift between 2012 and 2013, it appears that teacher perceptions of the focus of feedback have stabilized.

Figure 3.1: Teacher and Observer Perceptions of the Primary Focus of Feedback from Evaluators

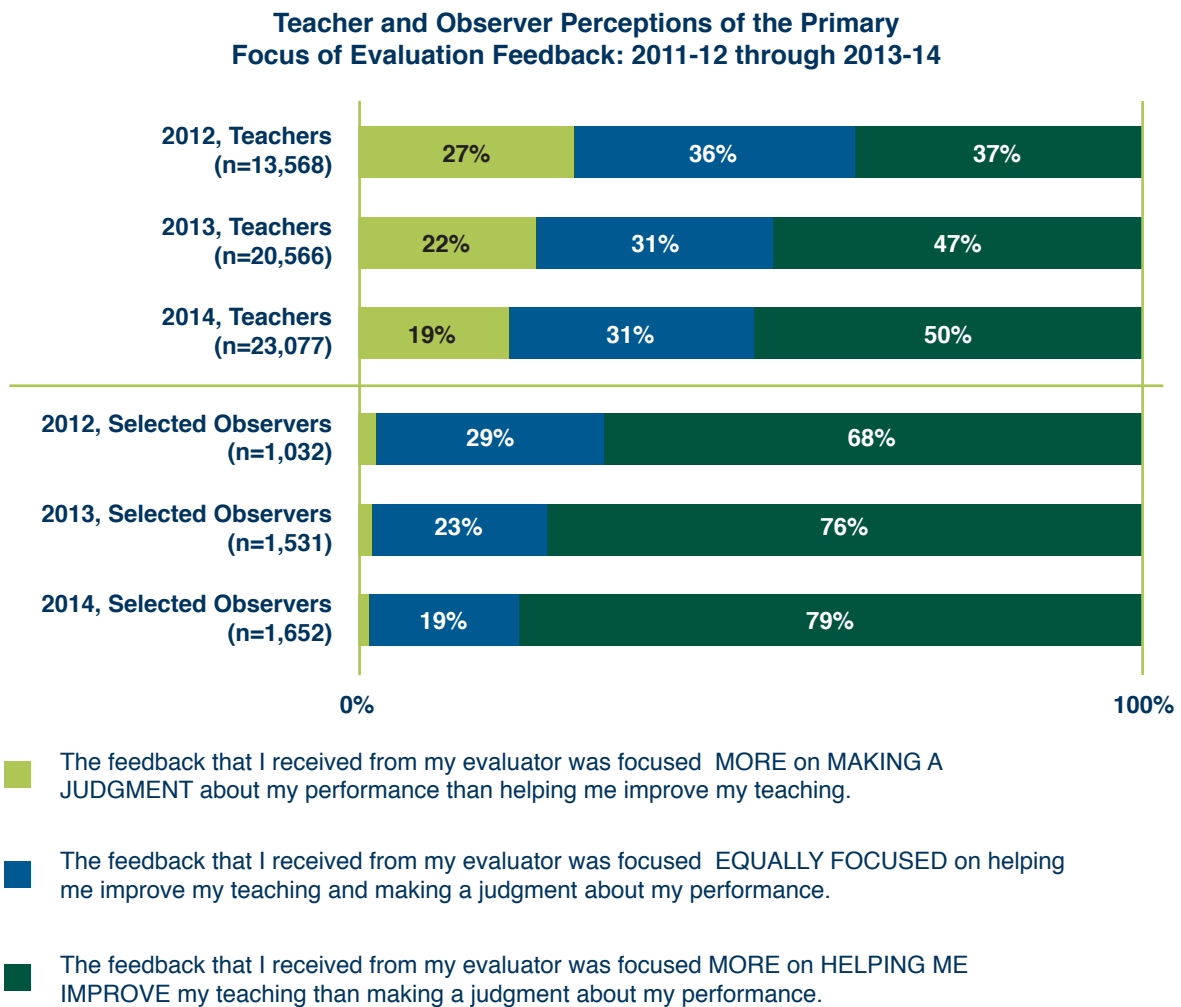


Figure 3.1 also shows that selected observers (principals, assistant principals, and instructional coaches who conduct observations) consistently reported that their feedback focused on improvement at significantly higher levels than teachers. This trend continued in 2014, with four out of five selected observers perceiving that their feedback was more improvement focused.

OBSERVER FOLLOW-UP

Two additional questions have been related to teacher perceptions of the focus of evaluation feedback. The first asked teachers to indicate the number of times their observer(s) “follow[ed] up with [them] concerning indicator(s) identified as ones needing improvement.” Figure 3.2 presents the distribution of responses to this question by 2012-13 final evaluation score. Very few teachers reported that evaluators followed up with them five times or more, though the percent is non-negligible for teachers with prior year summative evaluation scores of 1 or 2. Results also show little overall difference in the extent of follow-up by the previous year evaluation score. In general, at least a third of teachers, regardless of prior year evaluation rating, indicated that their observers never followed up on areas identified as needing improvement. There is only a small relationship between prior year score and frequency of evaluator follow-up, with struggling teachers only slightly more likely than high scoring teachers to receive more follow-up from observers.

Figure 3.2: Frequency of Follow-Up by Observers on Areas Identified as Needing Improvement, by Prior Year Evaluation Rating

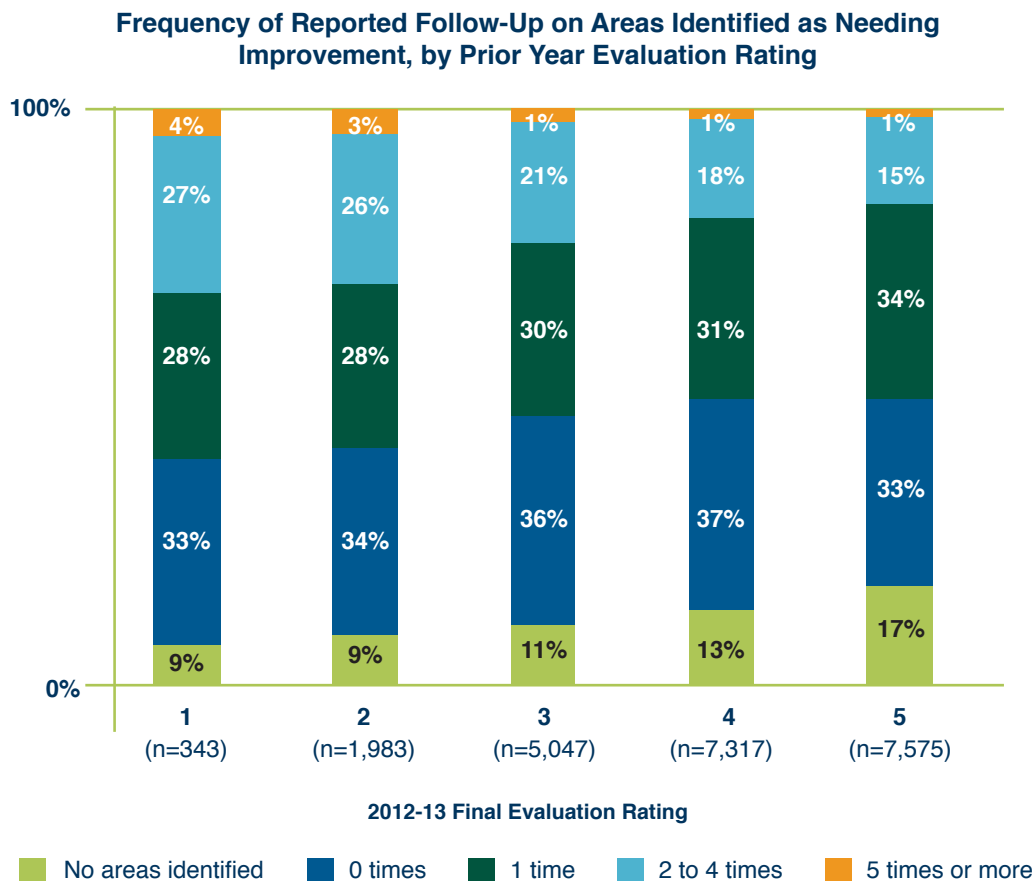
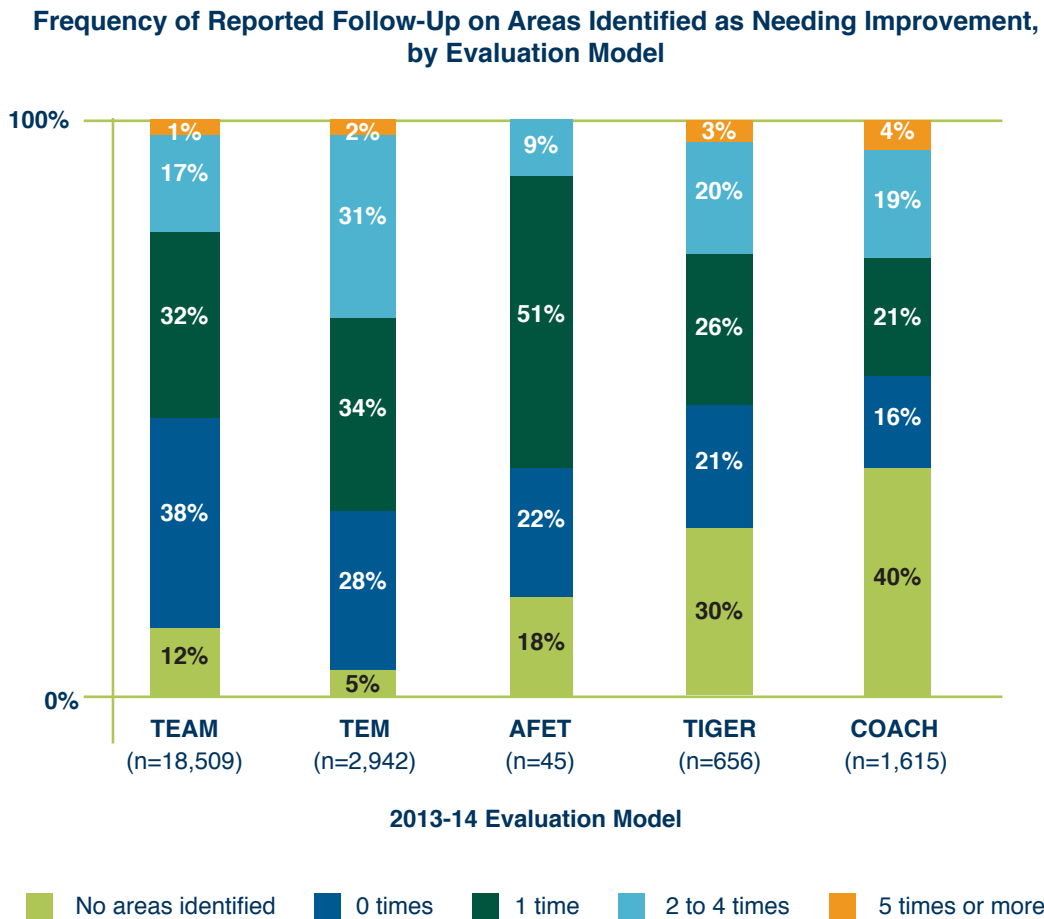


Figure 3.3 presents the frequency of reported evaluation follow-up by evaluation model. Results indicate more variation across evaluation models than across evaluation ratings. AFET, the model used in the Achievement School District (ASD) appears to encourage the most observer follow-up, with sixty percent of teachers receiving follow-up more than twice. Note also that all teacher respondents in AFET schools indicated that they had an area identified for improvement. Just under half of responding teachers evaluated under TEAM and more than two-thirds of teachers under TEM received some follow-up by observers with a third indicating their observers followed up with them once. The final two models, TIGER and COACH, show the lowest levels of follow-up by observers, a finding that may be driven by the models' particular design. Note that 30% of TIGER teachers and 40% of COACH teachers indicated that no areas had been identified for improvement.

Figure 3.3: Teacher Reported Follow-Up by Observers on Areas Identified as Needing Improvement, by Evaluation Model



SHARING OBSERVATION STRENGTHS WITH OTHER TEACHERS

Finally, researchers investigated the extent to which observers encouraged teachers to share areas identified as strengths with their peers. Results from both 2013 and 2014 are presented in Table 3.1. Fewer than 4% of respondents indicated that their observers identified no teaching strengths. Of the remaining teachers, 57% indicated that they were not encouraged to share teaching strengths with their peers. One might hypothesize that this is due to high scoring teachers being treated differently than low scoring teachers. Researchers examined responses to this question by 2012-13 teacher evaluation scores, which revealed little variation based on previous year evaluation scores with differences smaller than 10 percentage points.²⁰ As with the number of times observers follow-up with teachers about an area needing improvement, suggestions for a teacher to share areas of strength do not appear to be related to previous year final evaluation scores.

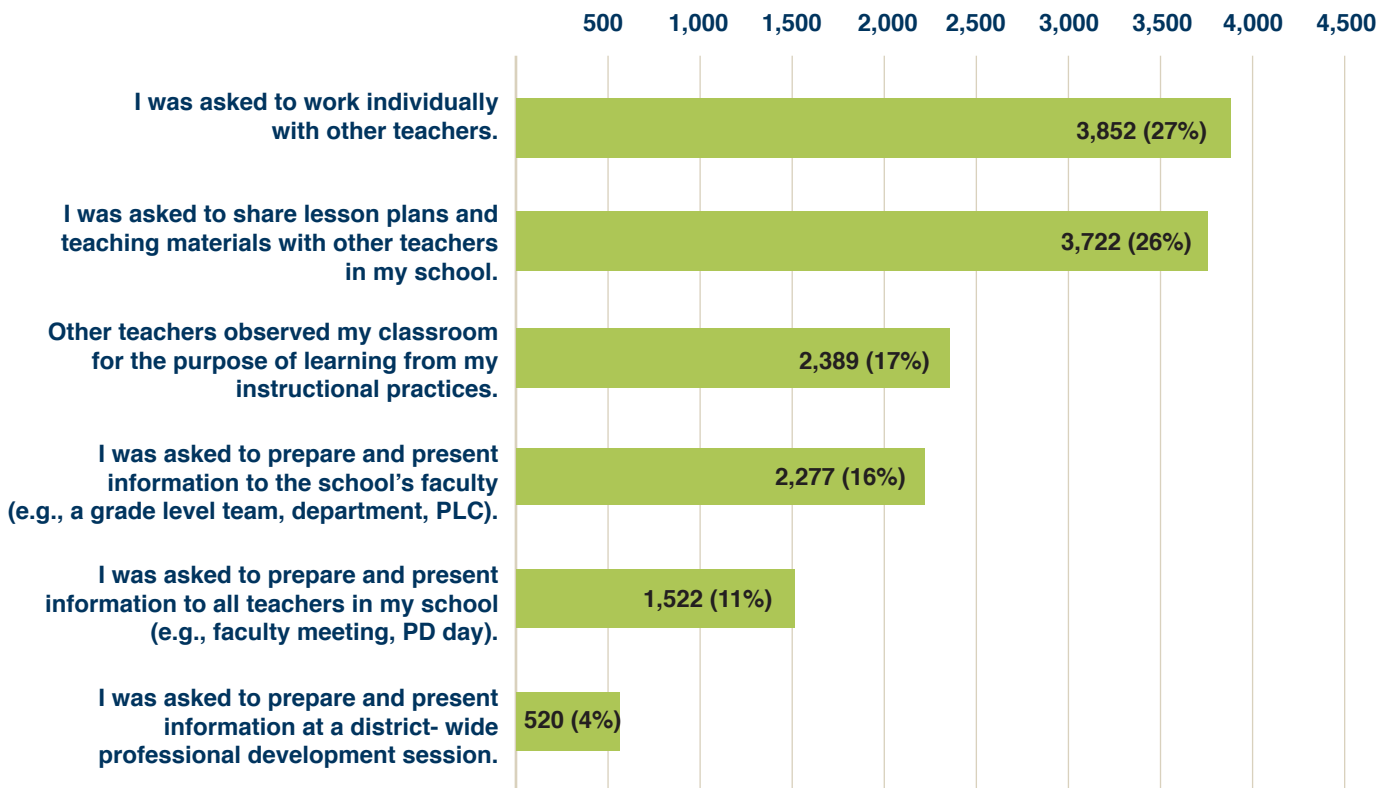
Table 3.1: Extent Teachers Asked to Share Teaching Strengths

	2013 Teachers (n=22,513)	2014 Teachers (n=24,148)
% Indicating No Teaching Strengths were Identified	3.9%	3.8%
% Indicating They Were Not Asked to Share Teaching Strengths	51.1%	56.9%

Teacher respondents asked to share their strengths with peers did so in various ways, as can be seen in Figure 3.4. Teachers were asked to indicate all methods of sharing their teaching strengths, so some teachers marked multiple options. Of the fourteen thousand respondents asked to share their teaching strengths, teachers were most commonly encouraged to work individually with other teachers, followed closely by sharing lesson plans and teaching materials.

Figure 3.4: How Teachers Were Encouraged to Share Teaching Strengths

**How Did Your Observers Encourage You to Share Your Teaching Strengths With Other Teachers in Your School?
 (2014 Responses Only, n=14,282)**



SUMMARY

This chapter reported results from FTTT survey questions that explored the extent and tone of evaluation feedback in Tennessee. Not surprisingly, teachers who perceive feedback to be more focused on improvement than judgment are more likely to perceive their evaluation system positively on a variety of measures. Results from the 2014 FTTT survey show a small increase in the percent of teachers who perceived the primary focus of their feedback to be on improving practice. In 2013-14 half of teachers perceived their feedback to be more focused on improvement, a third indicated that their feedback was equally focused on improvement and judgment, and one out of five indicated that their feedback was more focused on making a judgment. Selected observers perceive their feedback to be improvement-focused at much higher rates than teachers.

In 2012-13 teacher perceptions of the purpose of feedback was shown to be related to observer post-observation activities, specifically, how often they followed up on areas identified as needing improvement and whether observers asked teachers to share their teaching strengths with other teachers. Fourteen percent of respondents indicated that they did not have an area identified as needing improvement, a third had an area identified but did not receive any follow-up from observers, an additional third received one follow-up exchange from their observer, and twenty percent received two or more follow-up engagements. Variation in responses on this question is more related to evaluation model than prior year evaluation score. Indeed, there was not much difference in the frequency at which observers followed-up with high- and low-scoring teachers, nor with whether teachers were encouraged to share a teaching strength with their peers. Overall, perceptions of the primary focus of teacher observation feedback were similar to the prior year. Observation follow-up with teachers also continued to occur for only some teachers, and the frequency of follow-up was lower in schools using alternative models than in schools using the TEAM and TEM models.

IV. EFFORTS TO IMPROVE

INTRODUCTION

A fundamental assumption of the logic for educator evaluation systems in Tennessee schools is that teachers will use feedback from their teaching observations and other evaluation processes to improve their professional practice. According to the Tennessee Department of Education’s TEAM website,²¹ the evaluation system “serves as an excellent tool to help evaluators provide educators with constructive feedback to improve their instructional practice.” Survey results reported in Chapter 3 indicate that observers are providing relevant feedback to teachers. However, measuring teaching practice and providing feedback cannot lead to changes in practice if teachers don’t use the information. This section focuses on the “enactment phase” of the logic model. We first examine the extent of teacher efforts to improve instructional practices, the reasons they selected specific improvement targets, and their own perceptions of how much they improved. Next, we report activities teachers pursued to improve instruction and the relative contributions of those activities to perceived improvements. The final two sections report the types of evidence teachers reviewed to gauge their efforts to improve and information about areas of practice that teachers found most difficult to improve.

EXTENT OF TEACHER EFFORTS TO IMPROVE INSTRUCTION

Figure 4.1 illustrates that almost all teachers who completed the survey said they attempted to improve their instructional practice during the 2013-14 school year. Less than 2% of teacher respondents said they did not attempt to improve their instructional practices during the school year. Among teachers in schools using the TEAM rubric (the vast majority of teachers in Tennessee and the survey sample), 2.1% indicated they did not work to improve their instructional practice in 2013-14. This percentage is slightly lower for the other evaluation models: 1.1% for COACH teachers, 1.2% for TEM teachers, and 1.7% for TIGER teachers. This small subset of teachers is not significantly different from the sample as a whole in terms of teacher characteristics such as experience, tier, or final evaluation rating.

Figure 4.1: Number of Teachers Reporting They Worked to Improve Their Instructional Practice

Number of Respondents Who Indicated They Worked On Improving Their Instructional Practice During the 2013-2014 School Year, By Model

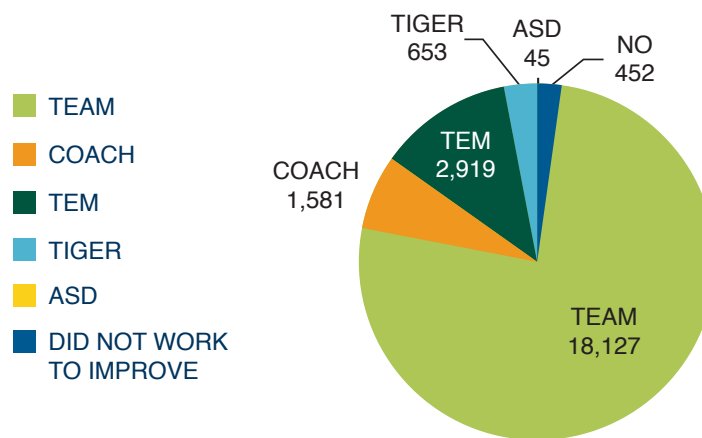
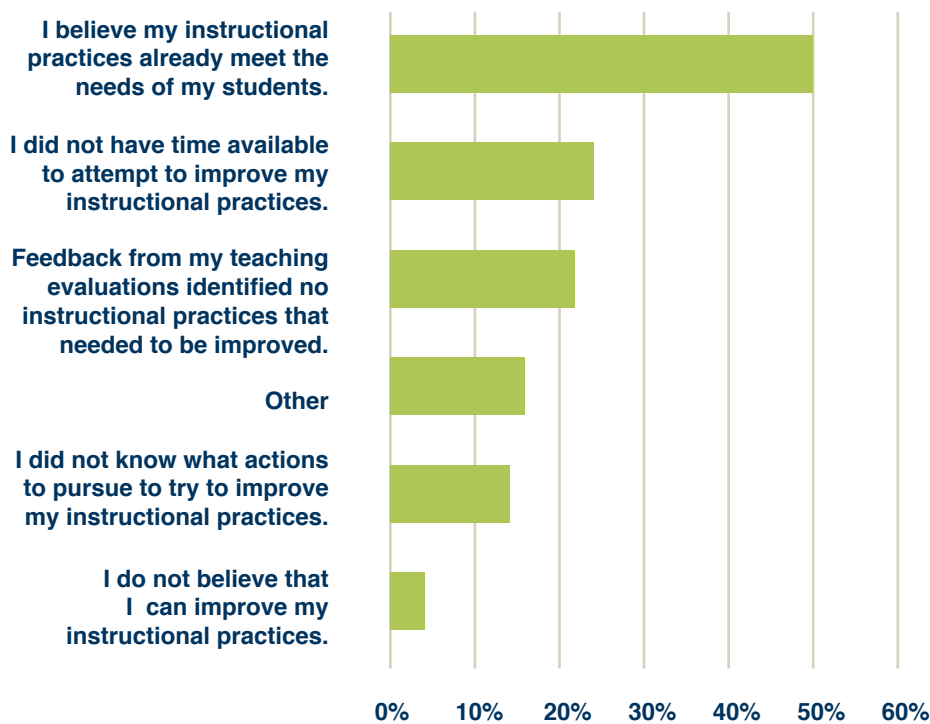


Figure 4.2 presents reasons teachers selected for not attempting to improve instructional practices. This group represents about 1% of the survey sample. Teachers could select more than one response to this item. The reason selected most frequently was that they believed their instruction already met the needs of their students. A lack of time and having no feedback indicating they needed to improve were the next most frequently selected reasons. The “other” category was selected by about 1 in 6 of these teachers, and the reasons they provided included chronic illness, pending retirement, and other life/family circumstances. It seems likely that the share of teachers who said they attempted to improve instructional practice, as well as the reasons a few did not, are not related to changes in teacher evaluation systems and that if the same question had been asked of teachers ten years ago, the responses would have been very similar (although there is no way of verifying this directly).

Figure 4.2: Reasons for Not Attempting to Improve Their Instructional Practice

**Reasons Teachers Did Not Work to Improve Their Instructional Practice
 (n=452)**

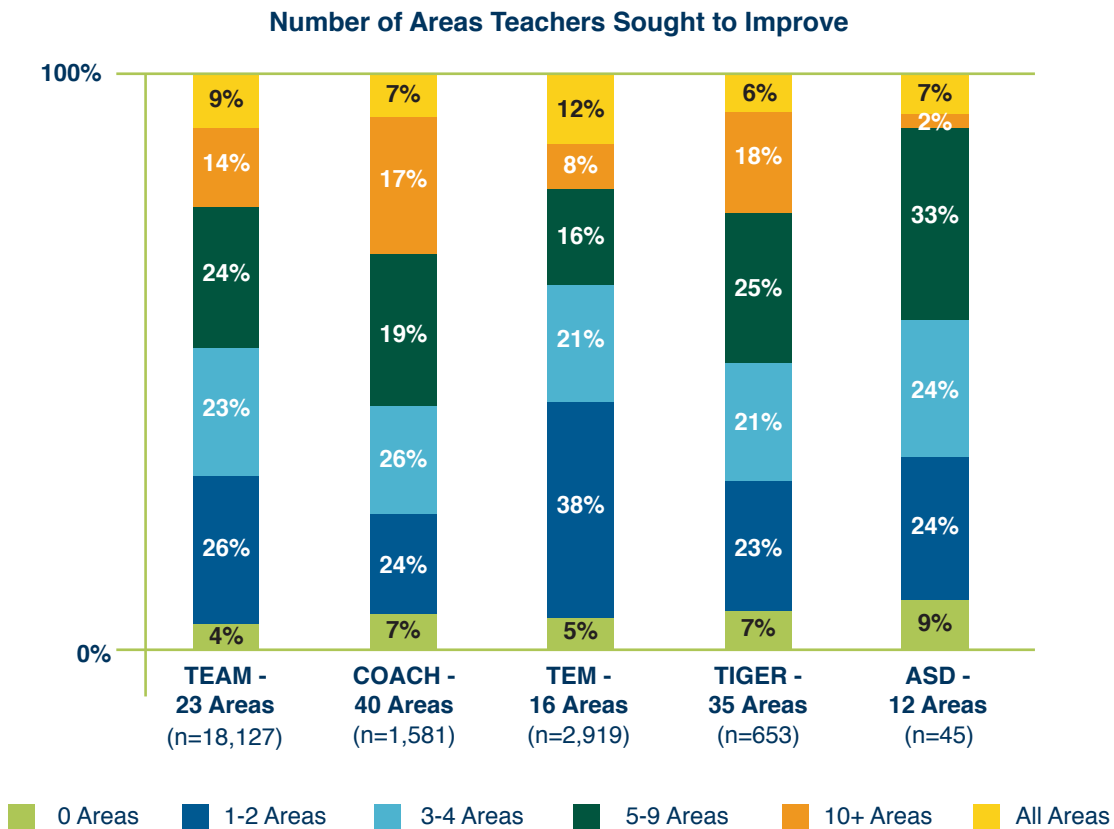


Breadth of Improvement Efforts

The remaining 98% of teacher respondents who indicated that they attempted to improve their instruction were asked to select every rubric indicator they attempted to improve over the school year and, for each selected indicator, to indicate whether it had been identified as needing improvement as part of their teaching evaluations. They were also asked how much they believed their practice had improved on each indicator using a four-point “None” to “A Lot” scale. The first issue of interest is the number of indicators teachers attempted to improve.

Figure 4.3 presents, by evaluation model, the number of rubric indicators teachers said they attempted to improve. A small percentage of teachers who indicated that they worked to improve their instructional practices during the year selected no rubric indicators (ranging from 3% of TEAM teachers to 9% of ASD teachers).²² Of the remaining teachers, between 35% (TEAM) and 58% (TEM) selected 1 to 4 rubric indicators, a number that would facilitate more focused and concentrated effort.²³ Just as interesting, between 4% (TIGER) and 12% (TEM) of teachers indicated they tried to improve on ALL rubric indicators.

Figure 4.3: Number of Rubric Indicators Teachers Targeted for Improvement, by Evaluation Model



Researchers examined whether teacher characteristics were related to the number of rubric indicators teachers worked on improving during 2013-14. There does not appear to be any relationship between the number of items selected and the type of school in which teachers worked (tier variable). Teachers with fewer years of experience were less likely to select just one item and those in the most experienced group were more likely to identify just one area. Similarly, less experienced teachers were more likely than more experienced teachers to select every rubric indicator. Final 2012-13 evaluation ratings had no correlation with the number of rubric indicators teachers said they attempted to improve ($r = -.04$), though the sign on the correlation coefficient indicates that those with higher ratings selected fewer rubric indicators. Teachers receiving an individual growth score did not appear to target more or fewer indicators for improvement.

For each area identified as one they tried to improve, teachers were also asked whether that area was identified as needing improvement during the evaluation process, and how much they believed their practice had improved in that area. Figure 4.4 shows the number of the rubric indicators teachers selected as areas they tried to improve that were also identified as areas needing improvement during the teacher evaluation process. Of all the indicators teachers targeted for improvement ($n=158,856$), 26.5% were identified for improvement as part of the evaluation process. In addition, the proportion of selected rubric indicators identified during teacher evaluation processes was also examined at the individual teacher level. This “average of averages” reveals that the typical TEAM teacher had about 40% of the rubric indicators they attempted to improve identified for improvement as part of the teacher evaluation process.

Figure 4.4: Number and Percent of Rubric Areas Teachers Sought to Improve that were Identified as Needing Improvement Through the Teacher Evaluation Model

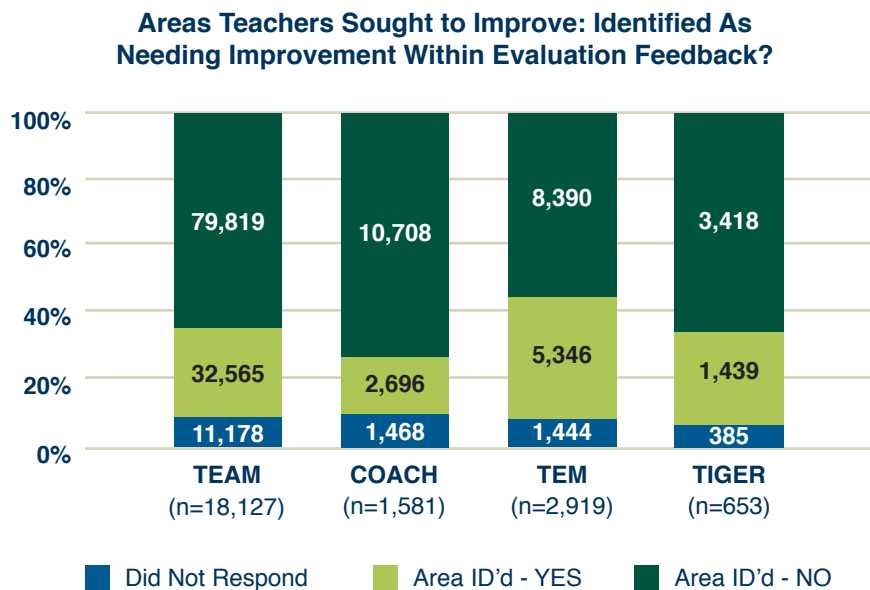


Table 4.1 presents teacher responses to the question regarding how much they believed their practice improved on each selected indicator. Between 80% and 90% of teachers reported improving “some” or “a lot” on the indicators they attempted to improve, regardless of the area of instructional practice or whether that area was identified as needing improvement in the evaluation process. In fact, there are no meaningful differences in the average improvement for rubric indicators that had been identified during teacher evaluations and those that had not been identified.

Table 4.1: Distribution of Teacher Self-Reported Improvement, by Evaluation Model

Evaluation Model	# of Areas Selected	Teachers' Self-Reported Improvement				Average Improvement - Areas ID'd	Average Improvement - Areas Not ID'd
		<i>None</i>	<i>A little</i>	<i>Some</i>	<i>A lot</i>		
TEAM	111,102	2.1%	16.3%	53.1%	28.5%	2.09	2.08
COACH	13,427	2.0%	15.4%	51.6%	31.0%	2.10	2.11
TEM	13,754	2.2%	10.4%	44.0%	43.4%	2.28	2.27
TIGER	4,862	1.7%	16.8%	51.9%	29.7%	2.14	2.07

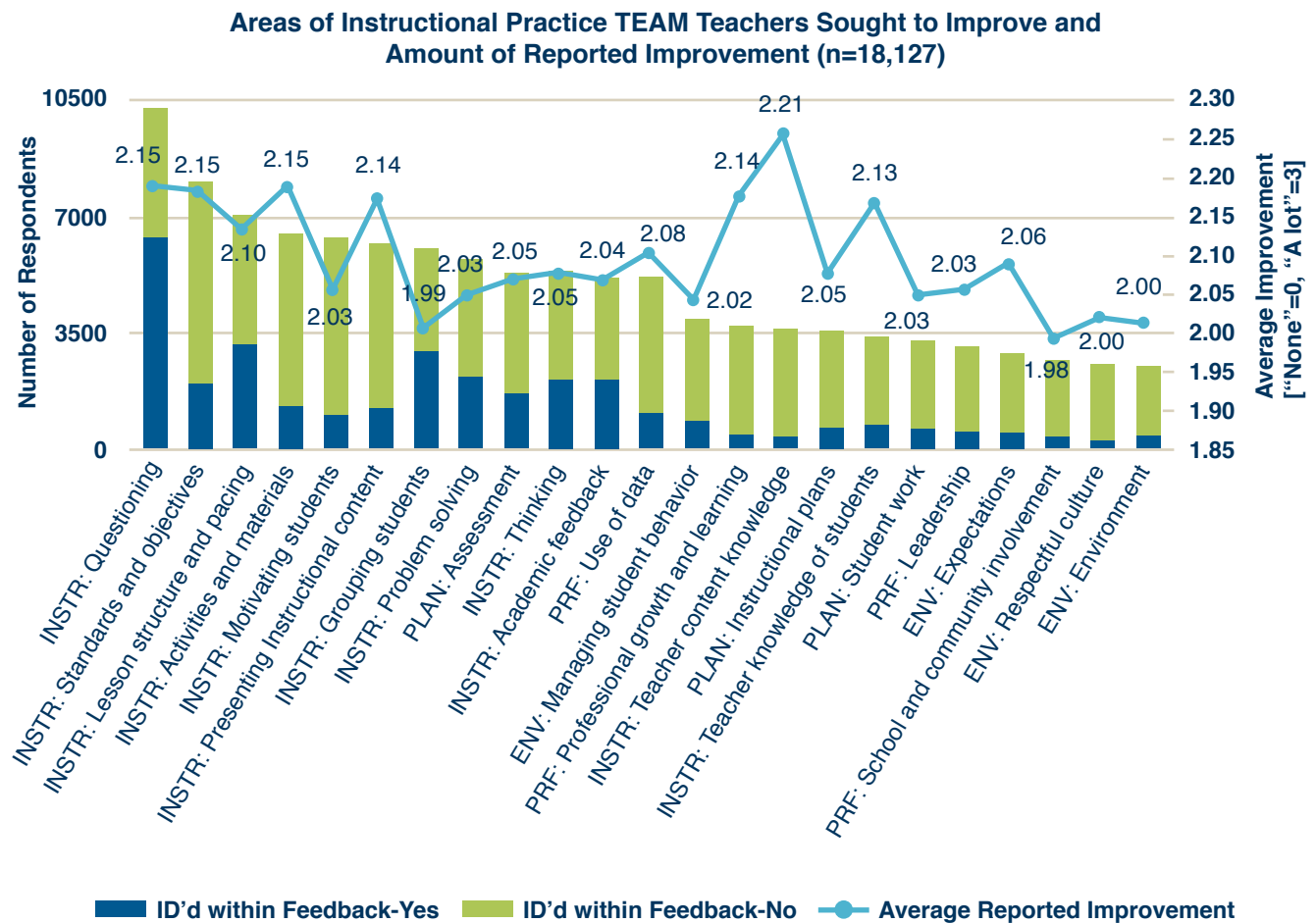
There are, however, some differences in reported levels of improvement across groups of teachers. For example, relatively new teachers reported slightly higher levels of improvement, on average, than more experienced teachers, and the same is true for elementary teachers compared to middle or high school teachers. Interestingly, there was no systematic relationship between the number of rubric indicators selected and teachers' self-reported levels of improvement.

Focus of Improvement Efforts

It also is interesting to examine which indicators teachers targeted for improvement. Figure 4.5 shows the number of TEAM teachers who said they sought to improve each rubric indicator and the number who said that area had been identified as needing improvement. Teachers tended to focus their improvement efforts on indicators from the Instruction domain; the six most frequently selected indicators, and eight of the top ten, are from the Instruction domain. Among the 10 indicators teachers most frequently sought to improve, “Questioning” was the only indicator which more than half of teachers indicated had been identified as needing improvement during

observation feedback. Two other indicators had nearly 50% of teachers indicating they had been identified for improvement during teacher evaluations: “Grouping Students” and “Lesson structure and pacing”. Teachers reported that all other indicators they worked to improve were identified during teaching evaluations at much lower rates.

Figure 4.5: Areas of Instructional Practice TEAM Teachers Sought to Improve

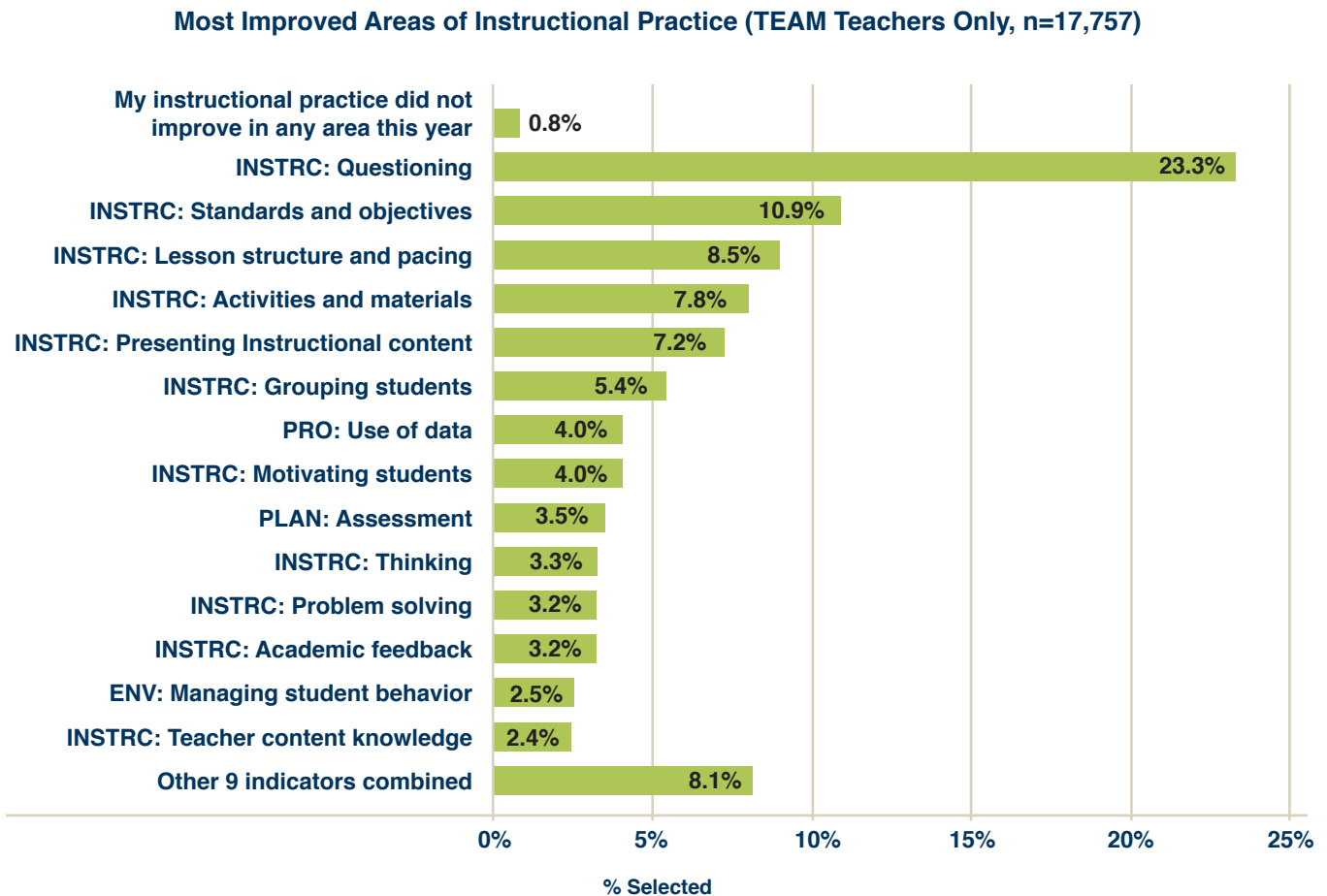


Similar data for other evaluation models are presented in Appendix C. A clear pattern emerges across all evaluation models: areas most commonly targeted for improvement generally are related to planning, preparing, or delivering instruction, i.e., the primary daily work of teachers. Areas most commonly targeted were generally identified as needing improvement in evaluation feedback. These were also the areas most commonly selected as “most improved” later in the survey. From these findings, it appears that teachers DO respond to feedback, in that what they choose to work on is determined by evaluation feedback.

Most Improved Area of Instructional Practice

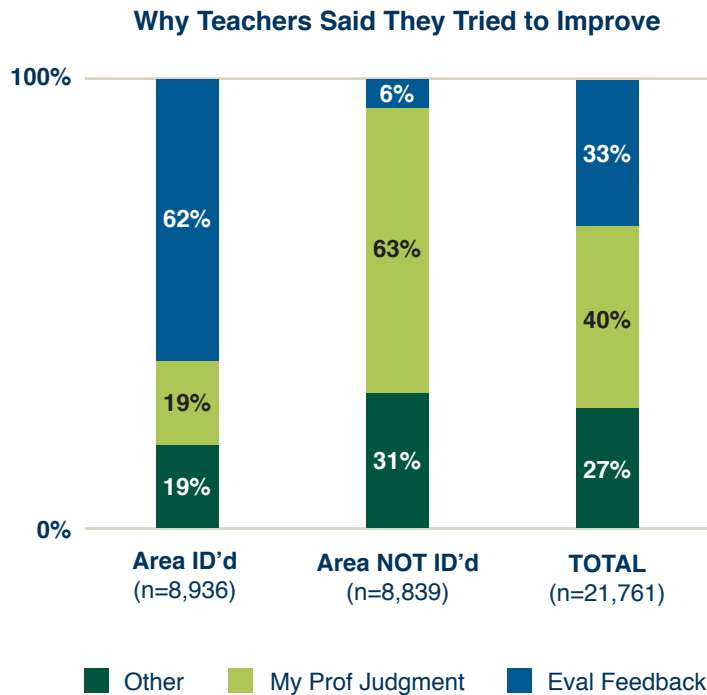
Teachers were also asked to select the single rubric indicator on which they believed they improved the most during the year. Teachers who selected an indicator were then asked about the activities they pursued to improve practice. Figure 4.6 presents the frequency that each TEAM indicator was identified as the one teachers improved most.²⁴ The first option allowed teachers to indicate that their practice had not improved in any area in which case a respondent was not asked about improvement activities they pursued. Only .8% of TEAM teacher respondents indicated that their instructional practices did not improve in any area during the year (.7% across all models). Conversely, “Questioning” was by far the indicator selected most frequently, which mirrors earlier results from the question that allowed teachers to indicate all areas they tried to improve. Note also that the top six indicators are from the INSTRUCTION domain. All other indicators were selected by less than 5% of respondents.

Figure 4.6: Frequency TEAM Teachers Selected Each Rubric Indicator as their Most Improved Area of Instructional Practice



Teachers were asked to indicate the primary reason they attempted to improve in their most improved area. The two most frequent responses to this question were “My Professional Judgment” and “Feedback from the teacher evaluation process indicated this as an area needing improvement”, accounting for nearly 72% of responses. However, responses to this question were highly related to whether the teacher’s most improved area was identified as needing improvement through the observation process. Figure 4.7 presents the responses for three groups of teachers: those whose “most improved” area was one they attempted to improve and had been identified for improvement as part of the teacher evaluation process; those whose area had been targeted for improvement but not flagged during evaluation; and all teachers.²⁵ As expected, for teachers whose most improved area was identified as needing improvement, nearly two-thirds reported their primary reason for attempting to improve in that area was feedback from the teacher evaluation process. For teachers whose most improved area had not been identified as needing improvement, approximately two-thirds reported that their professional judgment was the primary reason for attempting to improve in that area.

Figure 4.7: Primary Reason for Attempting To Improve in Area that Saw Most Improvement



WHAT TEACHERS DID TO IMPROVE INSTRUCTIONAL PRACTICE

This section focuses on the activities teachers pursued to develop their most improved area of practice. Teachers were presented a list of common improvement strategies and asked to mark all they had pursued. For each strategy selected, teachers were then asked how much that activity contributed to improved practice. Figure 4.8 presents the frequency with which each activity was selected and corresponding responses to the follow-up question on how much the activity contributed to improved practice. The two most frequently identified improvement activities are directly related to teachers' daily work of designing and delivering instruction: trying new learning activities and lesson planning. The next two improvement activities reflect strategies that can be pursued independently, with "Self-directed reading/learning" not bound by daily work schedules and "Informal consulting with peers" more likely to happen at school. The percentages shown beside the bars represent the proportion of respondents who indicated the activity contributed "a lot" to improved practice. About 40% of teachers reported that these top four strategies improved practice "a lot". While other improvement strategies also had relatively high percentages of teachers indicating they contributed "a lot" to improvements (e.g., one-on-one work with mentors or coaches, and college course work), they were pursued much less frequently. It is worth noting that there is much more variation in the contribution to improved practice within activities than between activities. In other words, every selected activity was perceived as helping improve practice some or a lot by most teachers.

Figure 4.8: Activities Pursued and Their Perceived Contributions to Improvement

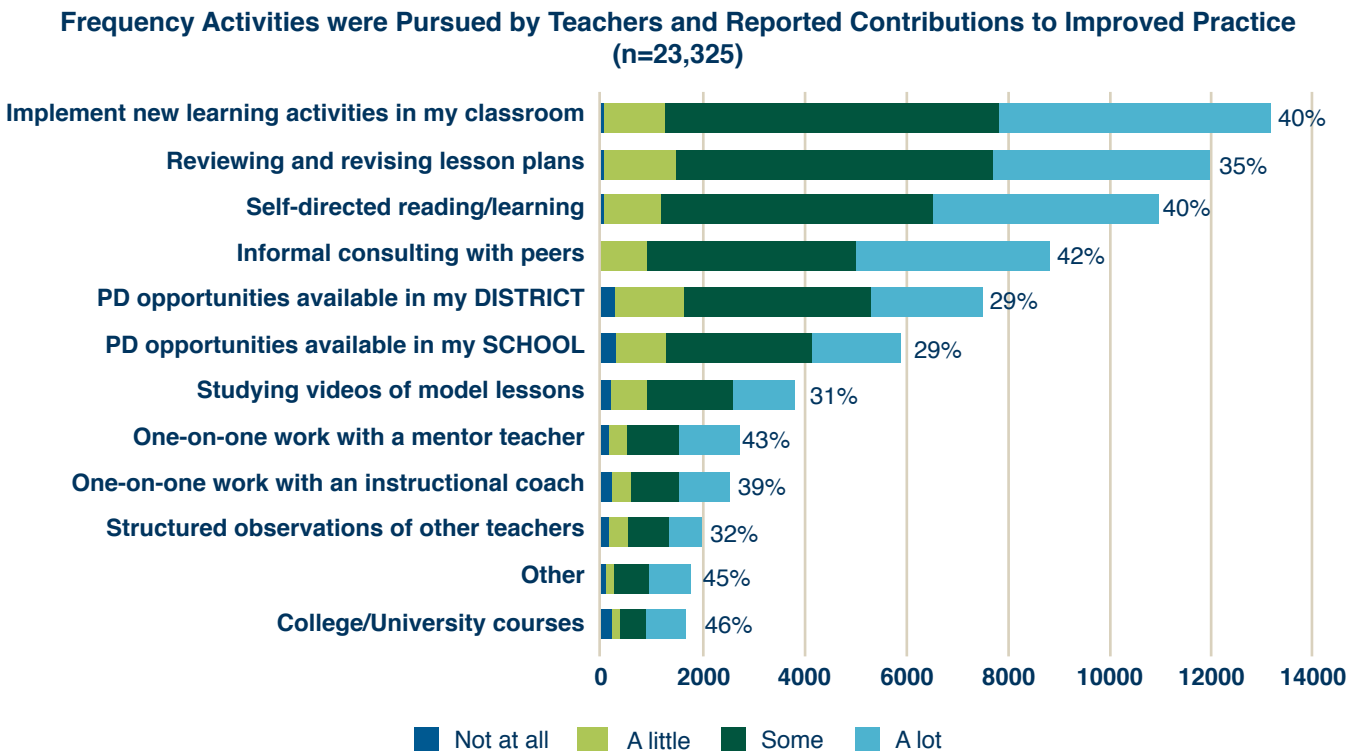
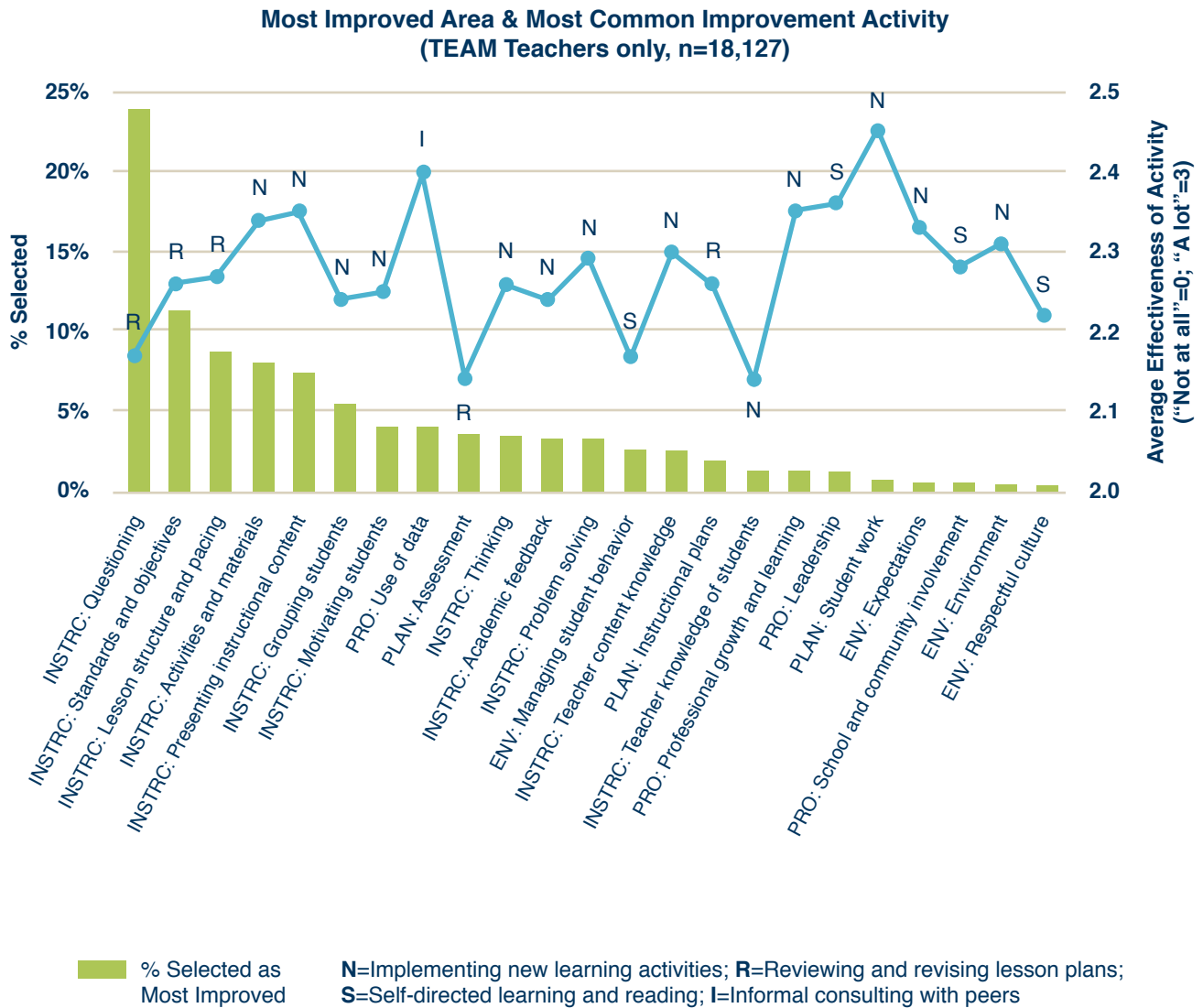


Figure 4.9 combines data from TEAM teachers about the indicator they believed was most improved and the activities they pursued to accomplish those improvements.²⁶ For each rubric indicator, the chart provides the percent of teachers who indicated it was their most improved area of practice, the activity most often pursued to make improvements, and the average contribution of that activity to improved practice. The indicator most frequently selected as most improved was “Questioning,” selected by nearly 25% of TEAM teacher respondents. The activity teachers selected most often for improving this indicator is coded as “R” for Reviewing and Revising Lesson Plans (the specific activity associated with each letter is shown at the bottom of the chart). The average contribution to improvement for this activity is shown by the line and dots with the corresponding value on the right hand axis – between 2.1 and 2.2 for this activity. “Implementing new learning activities” was the most frequently selected improvement activity for 13 of the 23 rubric indicators. Reviewing and revising lesson plans was a distant second (most frequently selected for only 5 indicators), followed by Self-directed learning and reading (4 indicators). Just as interesting are the activities that were never most frequently selected by teachers as an improvement activity. These included school and/or district professional development opportunities and one-on-one interactions with mentor teachers or instructional coaches. It is likely that schedule limitations and availability of instructional coaches inhibited the ability to work with others in one-on-one settings to improve teaching practices. Similarly, the limited opportunity for professional development sessions during the school year may have reduced opportunities to address improvement targets via this more traditional mode of teacher knowledge and skill development.

Figure 4.9: TEAM Teachers: Most Improved Area of Practice, Most Commonly Identified Improvement Activity, and Average Contribution of that Activity to Perceived Improvements



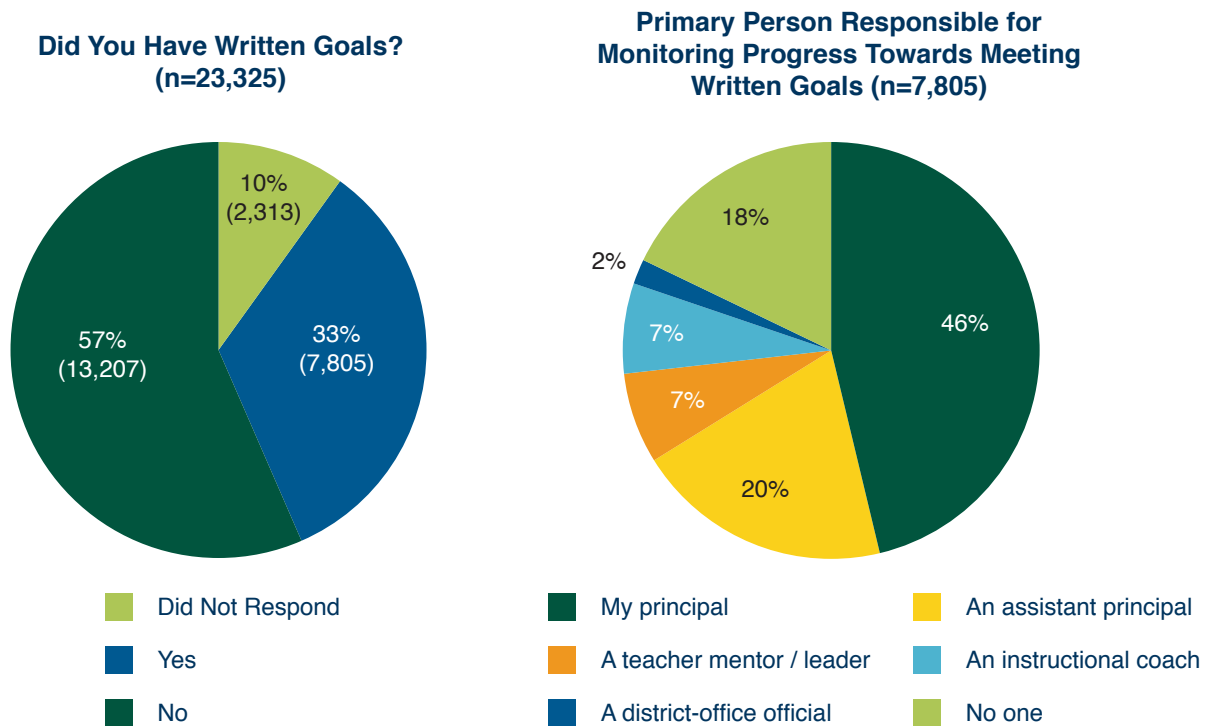
While it was possible to combine data from all teachers on the overall share who selected each potential improvement activity with a summary of their perceptions about relevant contributions to instructional improvement, teachers answered these questions in the context of a specific rubric indicator that they believed they improved the most. Some activities might have been more effective for improving some areas of instructional practice, but less effective for others. A summary presentation would miss those contextual differences. Appendix

C contains tables showing, by model, the share of teachers who indicated each indicator was their most improved area of practice and the percent of those teachers who pursued each listed activity – along with their average ratings of the contribution to improvement. These tables provide useful, detailed information about how teachers perceived the value of improvement activities for specific areas of instructional practice.

HOW TEACHERS GAUGED THEIR EFFORTS TO IMPROVE

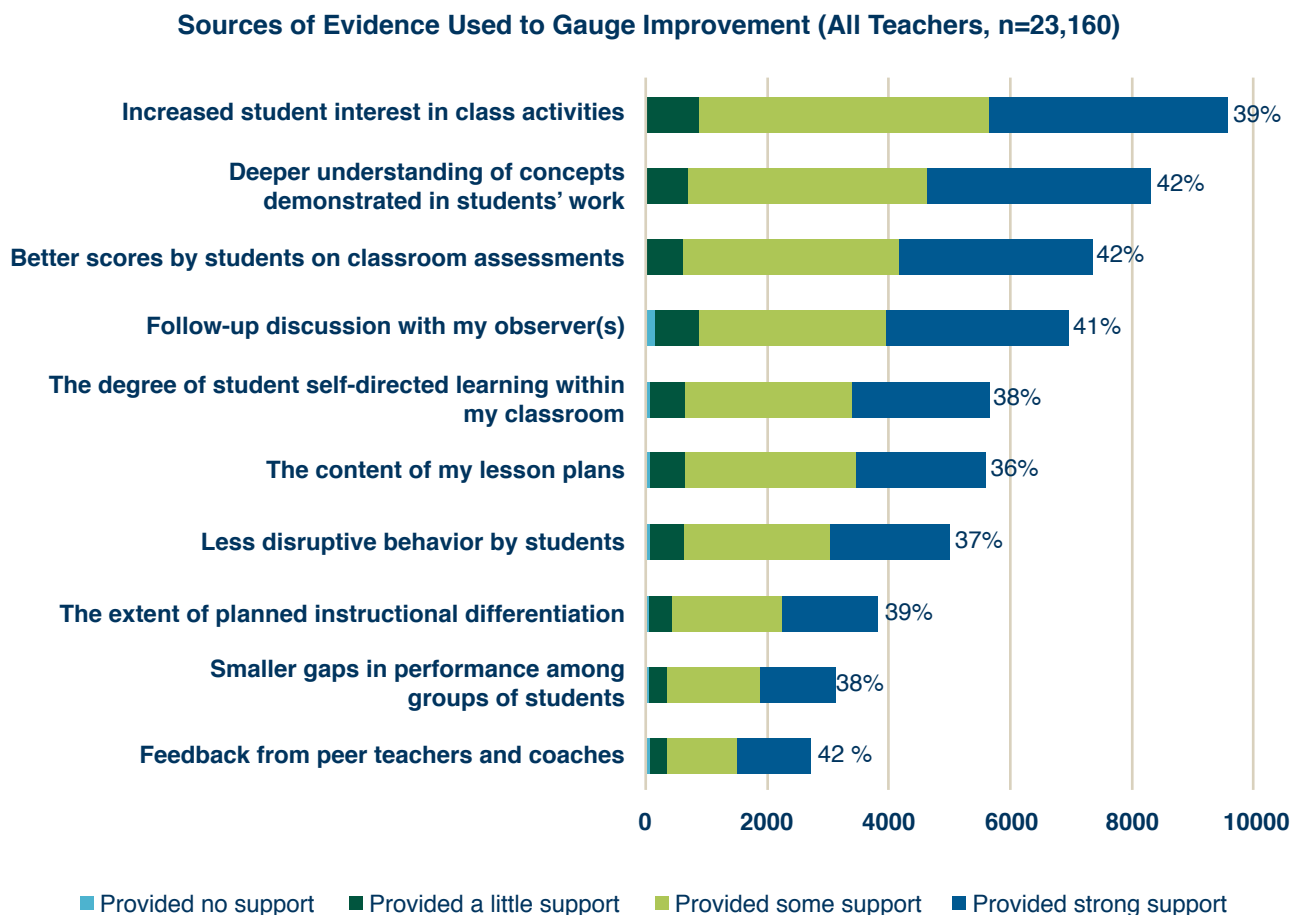
Results presented earlier in this report indicate that teachers believe their efforts to improve led to improvements in practice. This section investigates how teachers gauged the success of their efforts to improve. Judging improvements in instructional practice can be facilitated by developing written goals for the aspects of teaching performance identified as needing improvement. Teachers who identified a most-improved rubric area were also asked if they had written goals for improving practice in the selected area. Figure 4.10 presents results for this question. Only 33% teachers indicated that they had written goals for improving their practice in the area they identified as most improved. Of these 33%, almost half (46%) were monitored on their progress toward these written goals by their principal. However, almost one out of five teachers with written goals (18%) reported that their progress toward the goals was not monitored.

Figure 4.10: Written Goals for Improvement and Person Monitoring Progress Toward Meeting Them



Teachers were asked to indicate what evidence they used to determine how much their practice had improved, how much the evidence supported their perceptions of improved practice, and whether they reviewed the evidence with an observer. Figure 4.11 shows the number of teachers who selected each of the listed sources of evidence (teachers could mark more than one option) and a summary of related responses about the level of support provided by each source of evidence. The percentages shown beside the bars are the percent of teachers who indicated the evidence provided strong support for improved practice. Around 40% of teachers indicated the evidence provided strong support, and 75% to 85% indicated that each selected type of evidence provided “some” or “a lot” of support.

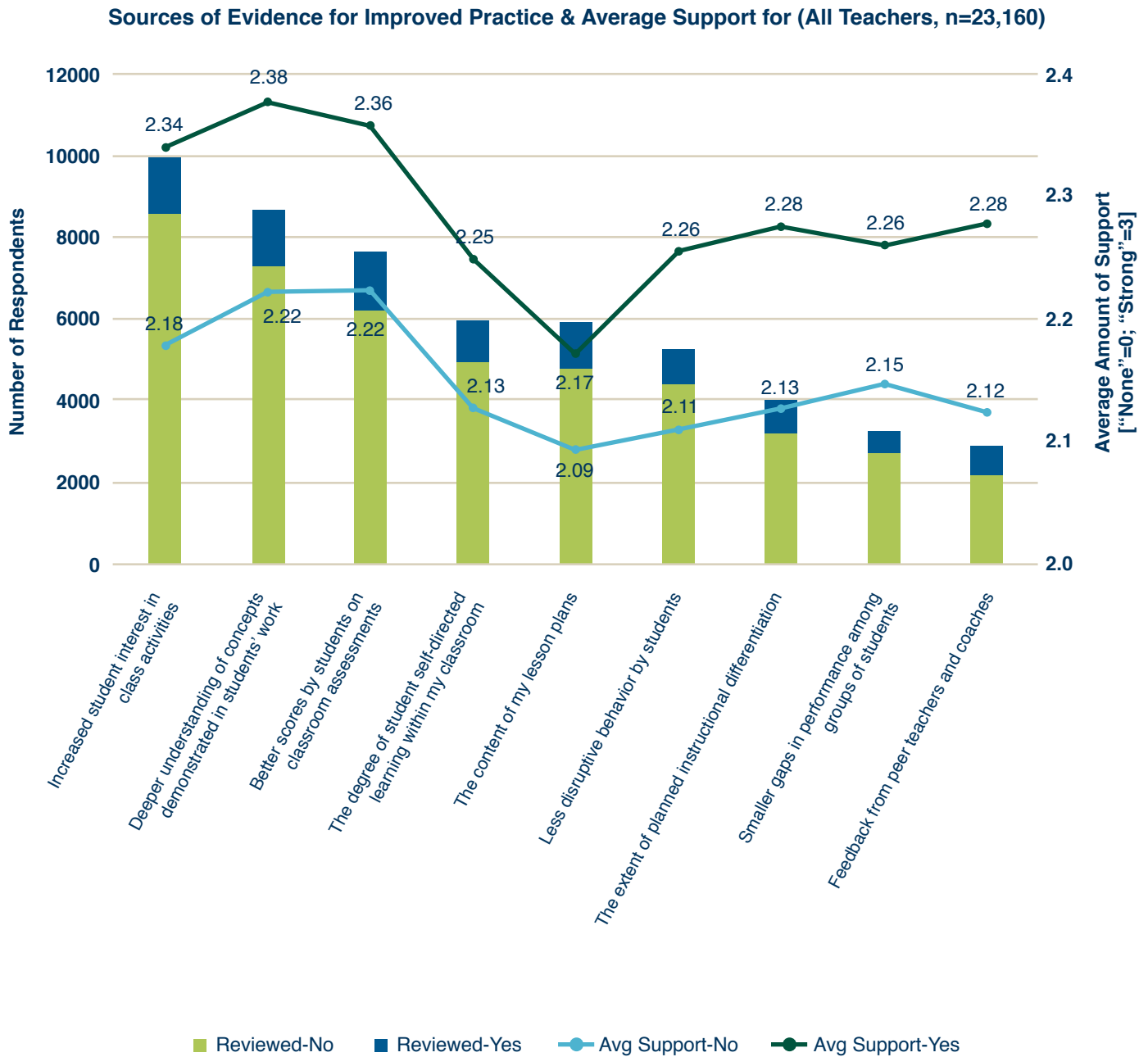
Figure 4.11: Sources of Evidence Used to Gauge Improvement and Degree Each Supported Improved Practice



Teachers tended to rely on evidence which could be gauged in daily practice, with the three sources most often selected focusing on student interest and classroom measures of learning. It is interesting that validation through discussions with observers is the fourth most frequently cited source of evidence, and it was no more supportive of improvements than classroom performance of students.

For each improvement activity, Figure 4.12 presents the degree to which evidence was reviewed with an observer and the level of perceived support for improvement when an observer was and was not involved. The vertical bars in Figure 4.12 show the proportion of teachers who indicated they had discussed each source of evidence with an observer. The lower portion of the vertical bars show the number did NOT review the source of evidence with their observer, and the upper portion shows the number who reviewed the source of evidence with an observer. Note that, according to teachers, most of the time evidence was not reviewed with observers. The dots on the two lines on the graph show the average ratings of support for improvement for each source of evidence. (The scale for this measure is on the right vertical axis and ranges from 0 to 3.) The upper line shows averages when teachers indicated the evidence was reviewed with an observer, and the lower line shows averages when teachers indicated the evidence was not reviewed. For every source of evidence, sources reviewed with an observer were seen as providing a substantially higher level of support than sources not reviewed with an observer. Differences between the two averages ranged between .1 and .2 scale points for all sources of evidence. There are two possible explanations for this finding. First, teachers who review a source of evidence for improved performance with an observer might be more likely to see the evidence as supporting improved performance. Second, teachers might choose to review evidence with an observer only when it supports improvement; evidence that does not suggest improved practice is not shared with observers

Figure 4.12: Number of Teachers Who Used Various Sources of Evidence to Gauge Improvement, Number Who Reviewed Evidence with Observer(s), & Average Degree Evidence Supported Improvement, by Reviewed Status



AREAS MOST DIFFICULT TO IMPROVE

The final questions in this section asked teachers to identify the rubric indicator they found most difficult to improve and why they found it hard to improve in that area. Figure 4.13 presents the rubric indicators selected as most difficult to improve by teachers working in TEAM schools.²⁷ The indicator most often selected as most difficult to improve was “Questioning,” selected by about 1 in 6 responding teachers, with “Problem solving” next, selected by about 1 in 10 teachers. Other indicators selected frequently by TEAM teachers reflect areas where success requires integrating knowledge about students and how they learn with teaching practices; these include “Motivating students,” helping students develop critical “Thinking” skills, “Grouping students” for instruction, and determining “Lesson structure and pacing.” Conversely, indicators at the bottom of the figure were selected by very few teachers and reflect routine teaching activities, such as “Presenting instructional content”, or having knowledge, such as “Teacher content knowledge” and “Teacher knowledge of students.”

Figure 4.13: Most Difficult Areas to Improve Reported by TEAM Teachers

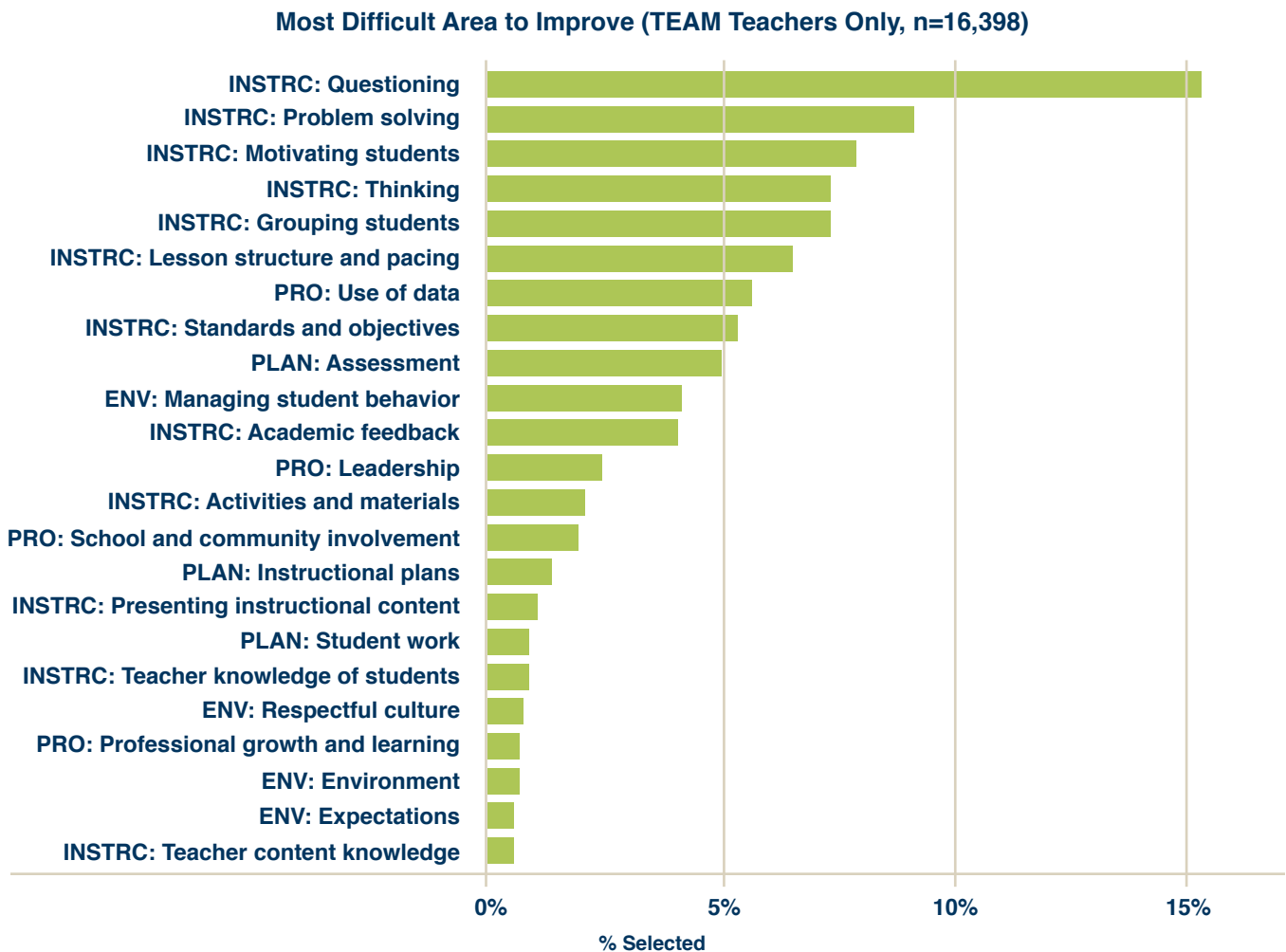


Figure 4.14 presents an analysis, by model, of how much teachers agreed with each of several listed reasons that improvements were difficult. Results reveal that a lack of professional learning opportunities and a lack of time were barriers to improvement.

Figure 4.14: Agreement with Reasons Areas Were Difficult to Improve, by Evaluation Model

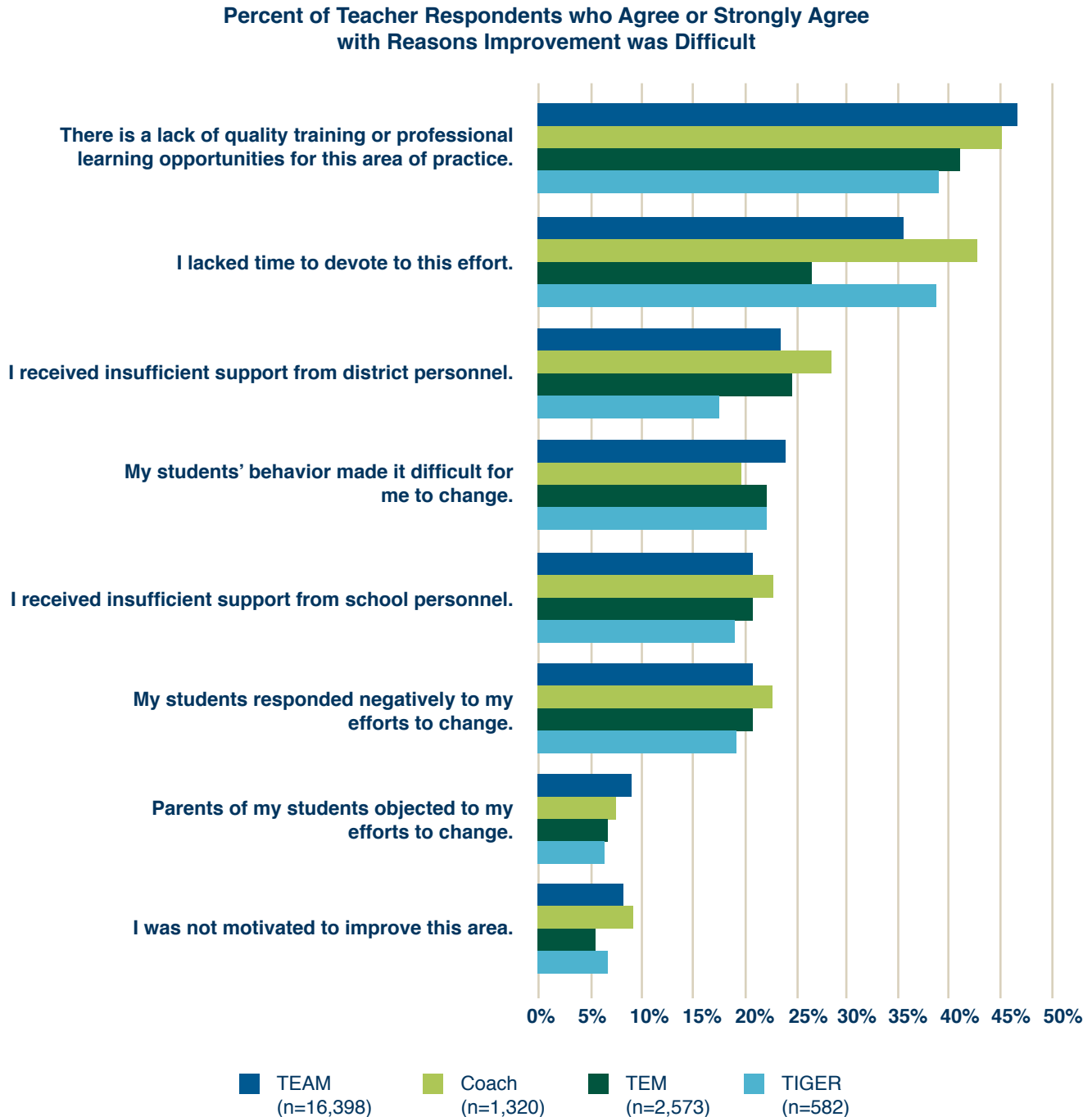
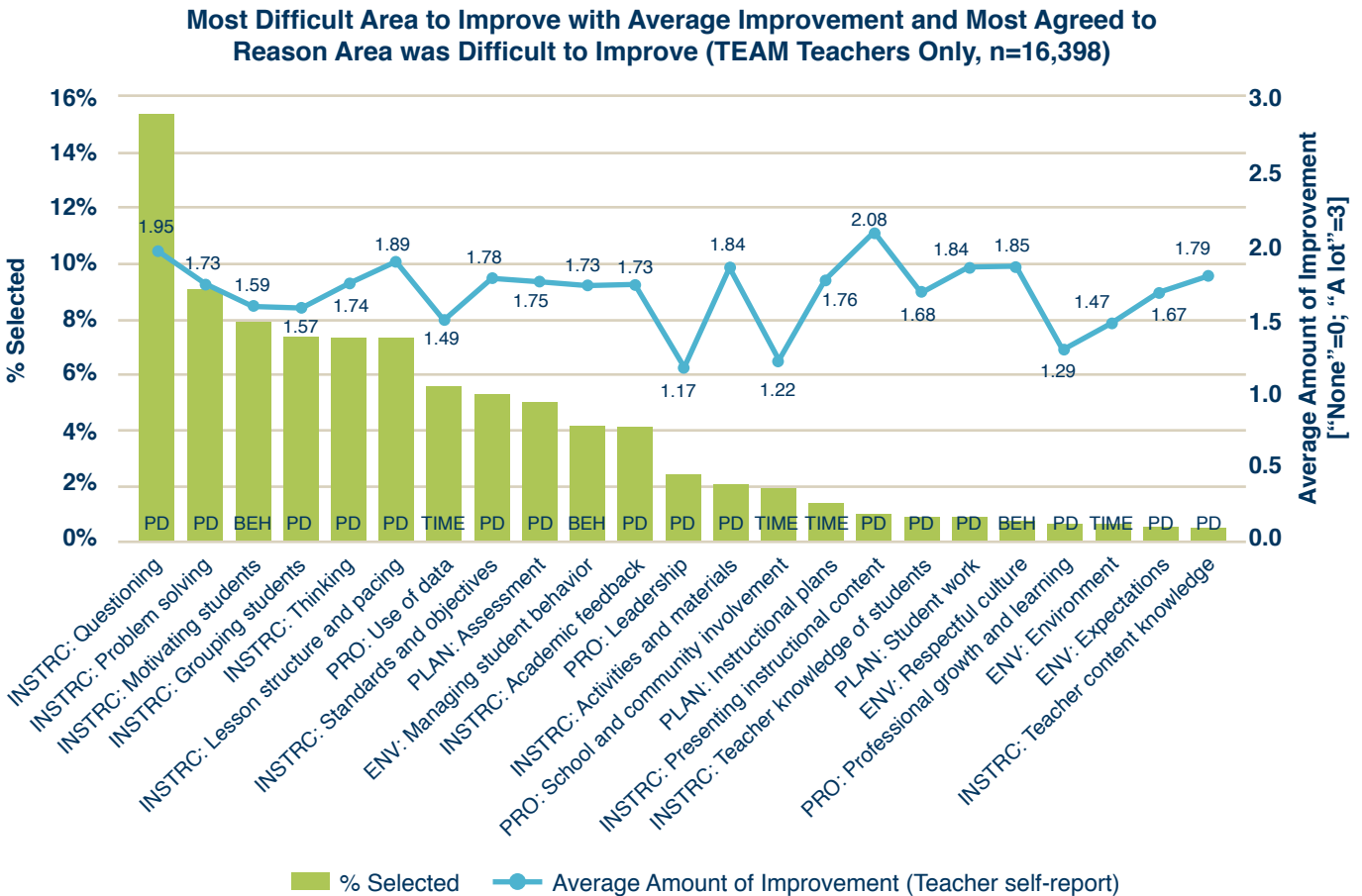


Figure 4.15 presents the percent of TEAM teachers who indicated each TEAM rubric indicator was the most difficult to improve (height of vertical bars), and the average level of self-reported improvement for that indicator (line and dots with scale on right hand axis). The barrier most frequently cited for each TEAM rubric indicator is identified by a label at the base of the bar. For example, teachers who responded that “Motivating students” was the most difficult indicator to improve (3rd indicator from the left) were most likely to cite students behavior as a barrier to improving practice (BEH is the label at the base of the bar). Teachers who selected this indicator as most difficult to improve reported an average level of improvement of 1.59 on the 3 point scale (between “a little” and “some”). Overall, the average extent that teachers reported improving practice on these indicators ranged from just over 1 to just over 2, values reflecting that most teachers selected “A little” = 1 and “Some” = 2.

Figure 4.15: Areas TEAM Teachers Found Most Difficult to Improve, Average Amount of Improvement in those Areas, and Reason Teachers Most Commonly Agreed or Strongly Agreed Why Area was Difficult to Improve



PD = “Lack of quality training or professional learning opportunities for this area”; **TIME**=“I lacked time to devote to this effort”; **BEH**=“My students’ behavior made it difficult for me to change”

SUMMARY

Nearly all of Tennessee’s teachers indicated that they attempted to improve their professional practice in 2013-2014. About half of teachers focused their improvement efforts on a few (less than five) areas of instructional practice. On average, teachers reported that approximately 40% of the indicators they worked on had been identified as needing improvement during the teacher evaluation process and the areas most commonly worked on were related to planning and delivering instruction. The average self-reported level of improvement for each area ranged between 2.1 and 2.5 on a 3-point scale, which reflects that teachers generally reported improving “some” to “a lot” on all areas they worked on improving.

More than 95% of teachers who said they tried to improve their instruction selected an area as the one on which they improved the most. About 40% of those teachers indicated that the selected area had been identified as needing improvement through the teacher evaluation process. In these cases, two-thirds indicated that evaluation feedback was the primary reason they attempted to improve in that area. For teachers whose most improved area had not been identified as needing improvement during teacher evaluations, two-thirds indicated that their professional judgment was the primary reason they attempted to improve.

The improvement activities most frequently pursued reflect daily work tasks (e.g., implementing new learning activities and reviewing/revising lesson plans). The other most commonly pursued activities could be completed independently (e.g., self-directed learning and informal discussions with peers). The least frequently pursued activities involve structured interactions with peers. Teachers indicated a relatively high level of contribution to improved practice for all activities they said they pursued.

A majority of teachers indicated they did not have written goals for improving practice in the area they indicated had improved the most, and among those who did, roughly 1 in 6 reported that no one was monitoring their progress towards achieving the written goals. The evidence teachers used most frequently to gauge their improvement focused on student interest and classroom performance, i.e., evidence that can be gleaned during instruction. Only about one in five teachers indicated that they reviewed evidence with an observer, even though, on average, teachers perceived evidence as providing more support for improved practice when that evidence was reviewed.

Areas that teachers found most difficult to improve were generally related to tailoring instruction to the diverse needs of their students, encouraging and supporting higher order thinking, and using informal assessment as a regular part of instruction. The main reasons teachers found improving in these areas difficult were a lack of high quality training or professional learning opportunities and a lack of time.

These findings suggest that teachers are using evaluation feedback to guide efforts to improve their instructional practice, although their professional judgment still influences what they work on. It appears that the majority of improvement efforts and the corresponding evidence teachers use to judge their improvement are embedded in the daily work of being a teacher.

V. TEACHER AND OBSERVER PERCEPTIONS OF THE TEACHER EVALUATION SYSTEM

As with any widespread reform, the perceptions of those implementing the evaluation system and those subject to it are important to track and understand. While favorable educator perceptions of the evaluation system are not an explicit goal of the Race to the Top, the evaluation system's success is partially based on the levels of educator buy-in and confidence in the process. The 2014 First to the Top Survey asked respondents a series of perceptual questions concerning the fairness, efficacy, and impact of the evaluation process. In addition to looking at changes in these perceptions over time, this section also explores the relationships between these metrics and how teachers perceived the primary focus of their evaluation feedback.

CHANGES IN TEACHER AND OBSERVER PERCEPTIONS

The 2014 First to the Top Survey asked respondents a variety of perceptual questions about the teacher evaluation process, including its fairness, clarity, level of classroom disruption, and impact on professional development. Responses to a selection of these questions are shown in Figure 5.1, which presents results from teachers and, when applicable, selected observers on each horizontal line. The question presented is listed on the left slightly above the line, with the percent of respondents agreeing or strongly agreeing with the listed statement represented by the placement of the marker from left to right. Results from the 2012, 2013, and 2014 FTTT Surveys are displayed using diamond markers, circles, and squares, respectively. Markers for teacher responses are solid while hollow markers are used for selected observers (principals, assistance principals, and instructional coaches who conduct observations). The three negatively worded questions are grouped at the bottom, while all of the other questions are positively worded. Selected observers did not answer all of the questions asked of teachers, so the hollow markers only appear on a subset of questions. For example, the first row shows that, in 2012, 71% of teachers agreed or strongly agreed with the statement *My observers are qualified to evaluate my teaching*. This rose to 77% of teachers on the 2013 First to the Top Survey and 78% on this year's survey.

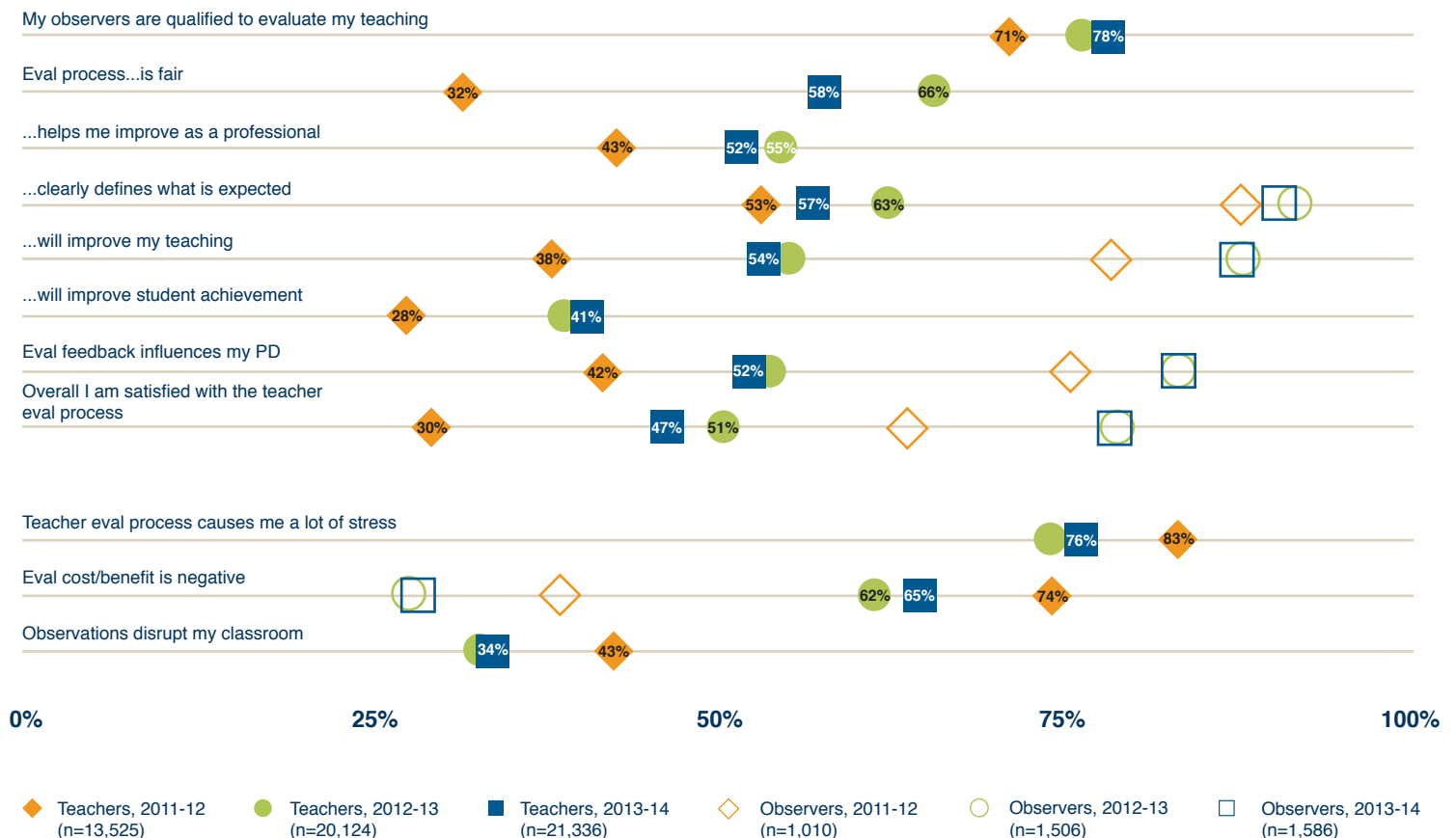
Note first the significant improvement in both teacher and observer perceptions from 2012 to 2013. Teacher perceptions of their evaluation process were originally quite low, but these perceptions significantly improved in 2013. Positively-worded questions showed large increases – of up to 34% on the question pertaining to fairness – and negatively-worded questions showed large decreases. On the summative item *Overall I am satisfied with the teacher evaluation process*, the percent of teachers agreeing with this statement increased from 30% to 51% from 2012 to 2013. Compared to this significant shift from 2012 to 2013, results from the 2014 First to the Top Survey show little movement and are similar to results from 2013. There is a slight regression on many, but not all,

questions in 2014. The percent of teachers satisfied overall decreased from 51% to 47%. The largest decrease was on the question concerning fairness, which shows an 8% decrease in the percent of teacher agreement. Slight improvements in the perceptions of teachers were found in questions concerning the qualifications of evaluators (+1%) and whether the evaluation process will improve student achievement (+2%).

Observer perceptions from the 2012 First to the Top Survey were much more positive than those of teachers and have subsequently increased. In 2012, 64% of observer respondents agreed or strongly agreed with the statement *Overall I am satisfied with the teacher evaluation process*, and this increased to 79% in 2013. Observer responses in 2014 show little difference compared to their perceptions in 2013, with the percentages of agreement within 1% for all questions. Overall, observer perceptions of the teacher evaluation system appear to have stabilized at levels much more positive than teacher perceptions.

Figure 5.1: Teacher and Observer Responses to 2014 First to the Top Perceptual Questions

Percent of Teachers and Selected Observers Responding Agree or Strongly Agree to Perceptual Questions Concerning the Teacher Evaluation System: 2011-12 through 2013-14



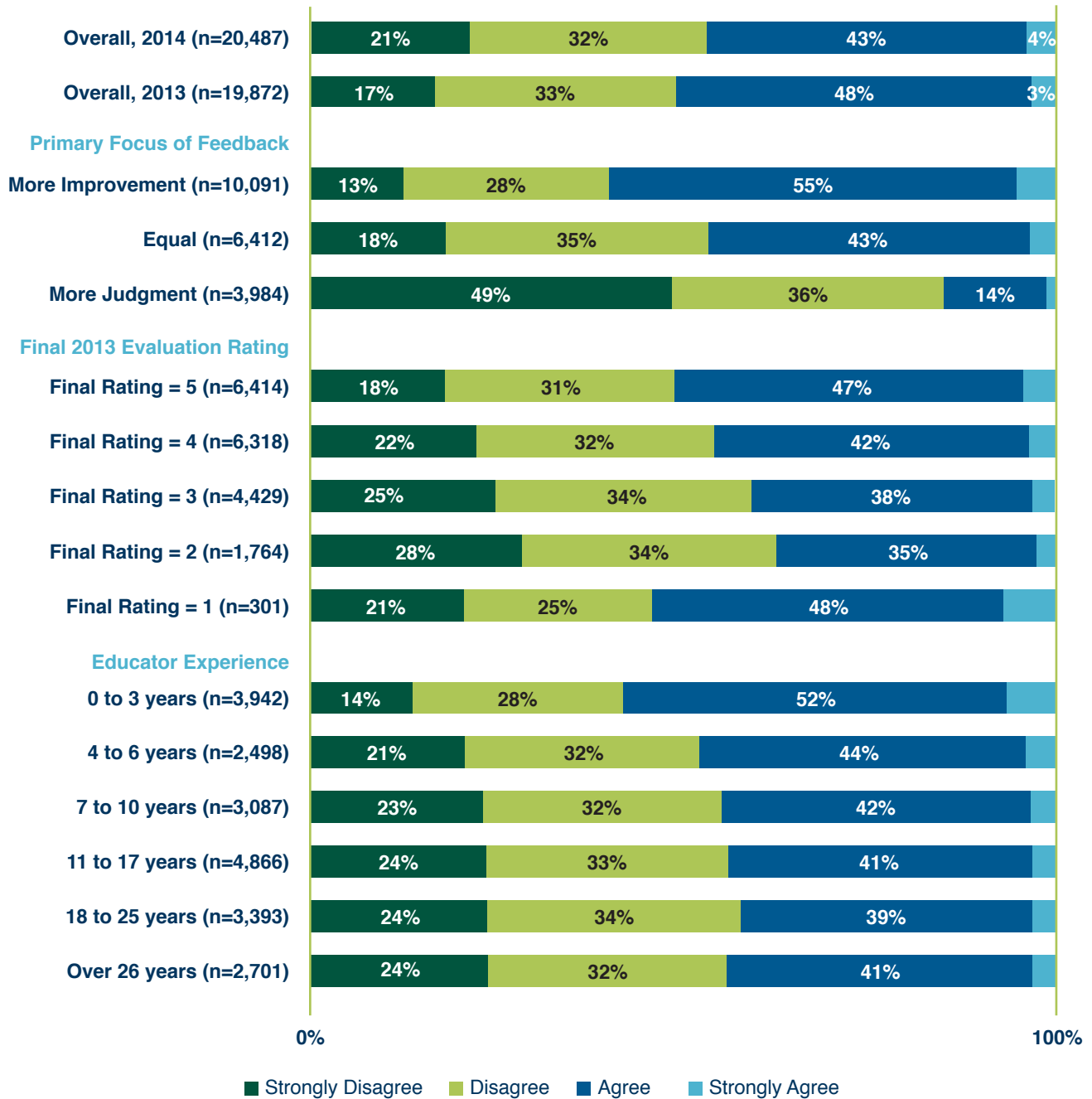
FURTHER EXPLORATION OF TEACHER PERCEPTIONS

Figure 5.2 presents differences in responses to the item *Overall I am satisfied with the teacher evaluation process* across important covariates. As with findings from the 2013 First to the Top Survey, there are significant differences in satisfaction across teachers who perceived their feedback to be more focused on improvement, on judgment, or equally weighted. Clearly, those who perceived feedback to be more focused on judgment are not as satisfied with the evaluation system, with only 15% indicating overall satisfaction. In contrast, 60% of the teachers who perceived their feedback to be more focused on improvement were satisfied with the evaluation process.

Also like last year, teachers with higher scores in the previous year were more satisfied than those with lower scores. Unlike last year, however, this trend does not hold for the 301 respondents who had received a score of 1 in the previous year. Surprisingly, these individuals were actually more satisfied with the evaluation process than teachers with other previous year scores. This may be due to extra attention or assistance received. Finally, the third grouping of responses reveals only small differences in the satisfaction with the evaluation process by teacher experience, except for inexperienced teachers, who are satisfied at a moderately higher rate.

Figure 5.2: Disaggregated Results for Teacher Level of Satisfaction with the Evaluation Process Used in Their School

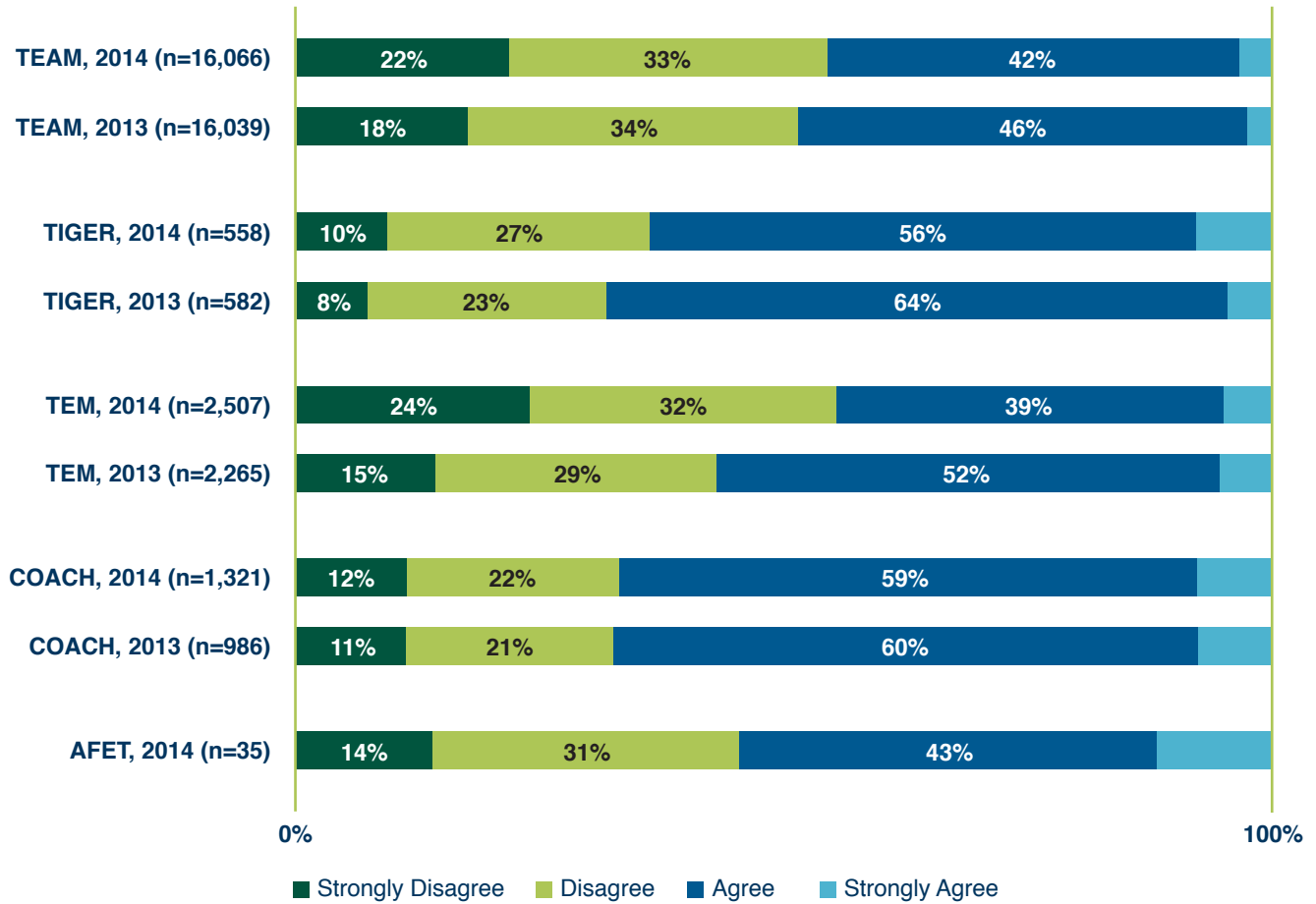
**Disaggregated Results for Teachers' Level of Agreement with the Statement:
 Overall I am Satisfied with the Teacher Evaluation Process Used in My School**



Finally, Figure 5.3 presents results from disaggregated responses to the overall satisfaction question by model from 2013 and 2014. The designs of Tennessee’s evaluation models share some similarities, but important differences exist. Interpreting differences in satisfaction across models is challenging, however, because a district’s leadership can opt into an alternative evaluation model. Differences in perceptions could very well be due to unobserved issues related to the particular district characteristics that made its leadership opt into an alternative model, or unrelated issues. Also, one alternative model, TEM, is used only by the largest school system in the state, Shelby County. Another model, TIGER, is in use mostly in rural districts, so ascribing these differences in perceptions to model characteristics may not be appropriate. It is interesting, however, to see change over time. Two models, TEAM and TIGER, show a slight decrease in the overall satisfaction levels of teachers from 2013 to 2014. TEM shows a significant decrease in satisfaction, but this model was adopted by the entirety of Shelby County in 2014, when in 2013 it was in use in the Memphis City School System only. Many teachers were evaluated under TEM for the first time in 2014. COACH continues to show the highest levels of satisfaction, and the model used in the Achievement School District, AFET, shows a satisfaction level of 54% - relatively high for the first year of a model’s use.

Figure 5.3: Disaggregated Results by Model for Teacher Satisfaction with the Evaluation Process Used in Their School

**Disaggregated results by model for teachers' level of agreement with the statement
*Overall I am satisfied with the teacher evaluation process used in my school***



SUMMARY

The positive trends in teacher perceptions of the evaluation system were one of the major takeaways from the 2013 First to the Top Survey Initial Report. In 2012 teachers expressed fairly negative opinions of the evaluation process, but in 2013 perceptions were more positive. Results from the 2014 First to the Top Survey reveal fairly similar responses compared to 2013 results, with perhaps a slight regression in the perceptions of teachers expressing more positive perceptions. The largest drops in teacher agreement were an 8% decrease on a question pertaining to the fairness of the evaluation system and a 6% decrease on a question asking whether the evaluation system clearly defines what is expected. There were very small increases in teacher agreement on questions related to the evaluation process improving student achievement (2%) and the qualifications of evaluators (1%). Selected observers continue to remain much more positive about the teacher evaluation process than teachers. Finally, the relationships between teacher perceptual questions and the improvement/judgment question and teacher covariates are, with a few minor exceptions, similar to the relationships found in results from the 2013 First to the Top Survey.

CHAPTER VI. CONCLUDING OBSERVATIONS

Consortium researchers designed the FTTT survey and reviewed results with a focus on how educator evaluation results are used to improve student learning through the development of teaching skills and professional practices. While evaluation systems can improve student learning by informing decisions on teacher retention and dismissal, the vast majority of professional educators in Tennessee will be impacted by evaluations through feedback and interactions that recognize their strengths, identify areas that can be improved, and spur efforts to make improvements. The following concluding observations are based on results of the 2014 FTTT survey and focus on the importance of the developmental process inherent in teacher evaluation systems.

IMPLEMENTATION

Tennessee's educator evaluation models prescribe processes, rules, and resources for how teacher evaluations should be conducted, scored, and used. It is important to investigate whether evaluation systems are being implemented with fidelity. All iterations of the FTTT survey capture how frequently observations are conducted and the time required to complete evaluation activities. Based on longitudinal analyses of survey findings, Consortium researchers offer the following concluding observations concerning the implementation of evaluation processes.

Classroom observations continue to be conducted, teacher performance scored and categorized, and results communicated.

One consistent finding from FTTT surveys is that approved teacher evaluation systems have been implemented with fidelity. Observers received training and felt well prepared to conduct observations. Evaluators conducted observations and completed other evaluation tasks as outlined by statute and state policy, logged results in a statewide database, and shared outcomes with teachers. Teachers with lower ratings in the previous year were observed more frequently and spent more time receiving feedback. These are positive findings and demonstrate that personnel in Tennessee's public schools have implemented new evaluation procedures as directed.

The time spent by teachers and administrators on tasks related to teacher evaluation throughout the year has stabilized, with relatively little time required from teachers and significant time required from administrators.

Survey results indicate that principals and assistant principals conducted three-quarters of observations, and that evaluation activities occupied a significant amount of their professional work week. More than 60% of administrators reported they spent more than 9 hours per week on activities related to teacher evaluations. Administrators also indicated that their allocation of time and effort to various job duties did not change much from

2013 to 2014. Teachers, in comparison, spent on average about eight hours per year on tasks related to teaching observations, which was very similar to patterns observed in 2013. These findings suggest that evaluation processes are becoming integrated into the regular work routines of teachers and administrators.

FEEDBACK

Feedback from teacher evaluations provides important information on instructional strengths and areas identified as targets for improvement. The primary focus of the 2013 FTTT survey was on the frequency, tone, content, quality, and use of feedback teachers received through the evaluation system. The following are findings from the 2014 FTTT survey about feedback teachers received from evaluations.

In 2014 half of teachers perceived observer feedback to be more focused on improvement than judgment.

Findings from the 2013 survey indicated that teacher perceptions of the primary focus of evaluation feedback

In 2014 the percentage of teachers who reported that the primary focus of feedback was on improvement increased slightly from 47% to 50%.

were strong predictors of perceptions of and attitudes about their evaluation system. Teachers who perceived that feedback was primarily focused on improving practice (rather than judging performance) were more positive in their perceptions and attitudes than teachers who perceived judgment as an equal or stronger focus. In 2014 the percentage of teachers who reported that the primary focus of feedback was on improvement increased slightly from 47% to 50%. This was coupled with a slight decrease, from 22% to 19%, in the percentage of

teachers who perceived that feedback was primarily focused on judgment.

Follow-up and follow-through activities between observers and teachers did not occur consistently or frequently.

In 2014 more than a third of teachers reported that their observers never followed up on areas needing improvement, and another 31% indicated their observers followed up only once. These findings were consistent across teachers with high and low prior year evaluation scores. Additionally, only 52% of teachers indicated that their feedback influenced the professional development in which they participated and only 33% of teachers indicated that they had written goals for improvement in the area they improved the most.

EFFORTS TO IMPROVE

The 2014 FTTT survey focused on teacher efforts to improve their instructional practices. This is the critical phase in the theory of action that suggests evaluations can lead to improved teaching. The following key findings indicate that teachers work to improve their teaching, though not always in response to evaluations or in partnership with their observers.

Teachers worked to improve instructional practices in 2014 – often in response to feedback from evaluations – and generally reported making improvements.

Virtually all teacher respondents (98%) indicated that they attempted to improve their instructional practice in the 2013-14 school year. Teachers varied in the number of indicators they attempted to improve and, on average, about 40% of the areas each teacher addressed had been identified during evaluation as needing improvement. Teachers indicated that they improved “some” to “a lot” in every area they worked on, and the average extent of improvement did not vary with whether the area had been identified through evaluation. Teachers also identified their most improved instructional area and provided their primary motivation for attempting to improve in this area. Two-thirds of teachers who improved most in an area that was identified during evaluations as needing improvement indicated that the primary motivation for improvement was feedback from evaluations. Conversely, teachers who improved most on an indicator that had not been identified during evaluations indicated that they pursued improvements because of their own professional judgment.

Teacher efforts to improve practice appear to most frequently reflect ongoing and routine instructional work.

Teachers provided information about the activities they pursued to improve the area they believed they improved the most, and these efforts to improve most frequently reflected regular work or informal learning efforts.

Implementing new activities in the classroom was the most frequently selected improvement activity for 13 of the 23 TEAM indicators, while reviewing and revising lesson plans was the most frequently selected for 5 other TEAM indicators. The next most frequently selected activities were self-directed reading/learning and informal discussions with peers. Teachers indicated these activities contributed “Some” or “A lot” to improved practice.

Interestingly, improvement efforts involving school or district-wide professional development and learning from other teachers (e.g., structured observations of other teachers and working one-on-one with master teachers or coaches) were never the most frequently selected improvement

Teachers identified the lack of quality training/learning opportunities and lack of time as the two most important reasons for why they found improvement efforts difficult.

activity for improving on any TEAM indicator. This may be related to another survey finding, which is that teachers identified the lack of quality training/learning opportunities and lack of time as the two most important reasons for why they found improvement efforts difficult.

PERCEPTIONS AND ATTITUDES

Organizational research consistently finds that the likelihood of successful organizational change is positively related to the level of stakeholder buy-in. All FTTT surveys have included items for assessing educator perceptions of the quality of evaluation systems, the potential for those systems to lead to positive outcomes, and educator satisfaction with evaluation processes and measures. Including the same or similar items on every annual survey allows monitoring how perceptions change over time and leads to the key finding discussed below.

Teacher and observer perceptions of the evaluation system in 2014 were similar to 2013, with just under half of teachers indicating in 2014 that, overall, they were satisfied with the teacher evaluation process used in their school.

One of the major findings from the 2013 FTTT Survey was a large positive shift in teacher perceptions of and attitudes towards evaluation. Results from the 2014 FTTT survey reveal similar levels of support for teacher evaluation as in 2013, with a slight overall regression in some areas. A smaller proportion agreed that the evaluation system was fair (from 66% in 2013 to 58% in 2014) and that the evaluation system clearly conveys expectations (from 63% to 57%). While almost four out of five teachers agreed that evaluators were qualified, only two out of five teachers agreed that evaluation processes will improve student achievement, and about half agreed that evaluation will improve their teaching. As in 2013, selected observers (principals, assistant principals, and instructional coaches) continued to be much more positive about the teacher evaluation process than teachers. For example, in 2014, 47% of teachers expressed overall satisfaction with the teacher evaluation process, in contrast with 79% of observers.

SUMMARY

The evaluation system continued to be implemented with fidelity, and teachers indicated that they responded to feedback through efforts to improve their practice. However, teachers were just as likely to pursue improvement in areas of practice that were not targeted by evaluation feedback and reported that improvement efforts typically involved activities embedded in traditional school routines and teacher work, such as revising lesson plans and implementing new activities in classrooms. Responses from teachers indicate that observers generally did not systematically monitor improvement efforts, and teachers cited lack of time and quality learning opportunities as key barriers to improvement. These findings suggest that while teacher processes are producing measures of teaching performance, teachers are often on their own as they attempt to improve their instructional practices.

END NOTES

¹ Established in 2010 as part of Tennessee's Race to the Top grant, the Tennessee Consortium on Research, Evaluation and Development is responsible for carrying out a detailed, focused program of research and evaluation around key grant initiatives. Learn more about the Consortium at www.tnconsortium.org.

² Survey items designed to gauge implementation and impacts of other FTTT improvement initiatives are included as "modules" in the FTTT survey and are answered by randomly assigned subsamples of respondents.

³ The majority of Tennessee districts use the TEAM model, the default, statewide evaluation system. Four other alternative evaluation systems, approved by the State Board of Education, were also in use within Tennessee in 2013-14.

⁴ The other two options that teachers could choose were that feedback was primarily focused on judging performance (about 18% of respondents) or feedback was equally focused on improvement and judgment (about 33% of respondents).

⁵ Analyses based on results from the 2012 and 2013 FTTT surveys can be found at the Consortium website: <http://www.tnconsortium.org>.

⁶ Teachers reporting that no areas had been identified as needing improvement were more likely to be in districts using the COACH or TIGER evaluation models.

⁷ For evidence on the link between observation feedback and increased student achievement see Taylor, E. S., Tyler, J. H. (2012). The Effect of Evaluation on Teacher Performance. *American Economic Review* 102 (7), 3628-3651 and Steinberg, M., Sartain, L. (2015). Does better observation make better teachers? *New evidence from a teacher evaluation pilot in Chicago*. *Education Next* 15(1), 70-76.

⁸ Established in 2010 as part of Tennessee's Race to the Top grant, the Tennessee Consortium on Research, Evaluation, and Development is responsible for carrying out a detailed, focused program of research and evaluation around key grant initiatives. As part of that work, the Consortium conducts the annual First to the Top Survey in an effort to solicit educator experiences of and attitudes towards First to the Top initiatives and reforms. Learn more about the Consortium at www.tnconsortium.org.

⁹ See Appendix A for a copy of the 2014 First to the Top Survey. A report containing frequencies of survey question and crosstabs by selected variables such as evaluation model, years of experience, and tier (elementary/middle/high school) can be found at http://www.tnconsortium.org/data/files/gallery/ContentGallery/2014_TNCRED_First_to_the_Top_Survey_Crosstabs.html.

¹⁰ Four additional evaluation models were used in Tennessee school districts in 2013-2014: Teacher Effectiveness Measure (TEM) in Shelby County, Teacher Instructional Growth for Effectiveness and Results (TIGER) in select districts who are part of the Association of Independent and Municipal Schools (AIMS), Project COACH in Hamilton County and three other districts in East Tennessee, and the Achievement Framework for Excellent Teaching (AFET) in the Achievement School District.

¹¹ Calculation of a teacher's overall score depends upon whether the teacher's growth measure is an individual measure, as is the case for teachers of TCAP tested grades and subjects, or a measure based on school- or system-level growth, as is the case for teachers in non-tested grades and subjects. Teachers with individual growth scores have it account for 35% of their overall score, achievement measures account for 15%, and qualitative measures account for 50% of the overall score. Teachers without an individual growth score have an alternative such as a schoolwide growth score account for 25%, achievement measures account for 15%, and qualitative measures account for 60% of the overall score. Also, in 2013-14 nineteen districts incorporated results from student perception surveys into teacher evaluation results.

¹² More information on Tennessee's evaluation policies and the TEAM model can be found at <http://www.team-tn.org>.

¹³ Of the 26,589 teacher respondents, 1,699 (6.4%) indicated that they had conducted observations in 2013-14.

¹⁴ Approximately 6,000 educators appeared in EIS but not in CODE, and as such did not have an email address to receive an invitation. Several thousand of these educators appeared to be erroneously listed in EIS due to a governance issue with Tennessee's largest district; the remainder were mostly positions other than traditional teachers, such as reading specialists and central office staff.

¹⁵ These findings are presented within two reports, both available from <http://www.tnconsortium.org>: Pepper, M.J., Burns, S.F., & Springer, M.G. (2012). Educator evaluation in Tennessee: preliminary findings from the 2012 First to the Top Survey. and Ehlert, M.W., Pepper, M.J., Parsons, E.S., Burns, S.F. & Springer, M.G. (2013) Educator Evaluation in Tennessee: Initial Findings From the 2013 First to the Top Survey.

ENDNOTES

¹⁶ Tennessee Department of Education. (2014). Suggested Observation Pacing. Retrieved from <http://team-tn.org/suggested-observation-pacing>.

¹⁷ A fifth alternative teacher evaluation model, the Achievement Framework for Excellent Teaching, is used in the Achievement School District (ASD). The number of responses from the ASD, a small district, is too small for their results to be included in the figure.

¹⁸ The COACH model uses a design that incorporates frequent, brief visits by observers. Over half of COACH respondents chose the highest answer option, "More than 6 times". This answer option was recoded by researchers as 7.5 visits for use in calculating the mean number of observations for these teachers.

¹⁹ Tennessee Department of Education. (2014). New Tenure Law, Frequently Asked Questions. Retrieved from <http://team-tn.org/wp-content/uploads/2013/10/New-Tenure-Law-FAQs.pdf>.

²⁰ A document that reports results for all 2014 FTTT survey questions by important covariates, such as previous year evaluation score, total years of educator experience, and tier, can be found here: http://www.tnconsortium.org/data/files/gallery/ContentGallery/2014_TNCRED_First_to_the_Top_Survey_Crosstabs.html

²¹ <http://team-tn.org/resources/best-practices/>

²² It is unclear why teachers would not select rubric indicators. Perhaps their perceptions of improvement efforts were broader than specific indicators or perhaps none of the indicators reflected the knowledge or skills they attempted to improve. Comments from these teachers will be examined to see if they provided a rationale for not selecting rubric indicators.

²³ The number of areas of practice teachers attempted to improve may be related to actual improvements they achieved. Teachers who said they focused on improving only a few areas of practice may have had better success than teachers who indicated they attempted to improve a large number of indicators.

²⁴ Complete results for each approved evaluation model are presented in Appendix C.

²⁵ The number of responses within the third category is more than the sum of the other two because 3,986 respondents identified an area as "most improved" but did not previously identify that area as one they had sought to improve - and so did not have an opportunity to indicate whether the area had been identified as needing improvement within evaluation feedback.

²⁶ Overall data about the number of teachers pursuing each improvement activity and their responses to the follow-up question about contributions to improvement are presented in tables for each evaluation model in Appendix C.

²⁷ Tables showing similar results from teachers in schools using other approved evaluation models and rubrics are presented in Appendix C.

APPENDICES

Appendix A *2014 First to the Top Survey*

Appendix B *The Representativeness of Respondent Samples from the Spring 2014
First to the Top Survey*

Appendix C *Response Frequencies to Survey Items from the “Efforts to Improve”
Section (Chapter 4), by Evaluation Model*