Educator Evaluation in Tennessee: Preliminary Findings from the 2012 First to the Top Survey

TENNESSEE CONSORTIUM on Research, Evaluation & Development

Matthew J. Pepper
Susan Freeman Burns
Matthew G. Springer

Introduction and Overview

One of only two states to be awarded a grant in the first round of the 2010 Race to the Top competition, Tennessee is now midway through the implementation of its four-year, $501 million award. The proposals in Tennessee's grant application are becoming reality across the state as Tennesseans transition to new curricular standards and assessments, a new accountability model for student achievement, and a new educator evaluation system.

As reforms progress, the Tennessee Consortium on Research, Evaluation, and Development continues its examination of educator perceptions of these efforts through the First to the Top Survey, administered for the first time in the spring of 2011, and again in April and May 2012. The initial survey gathered baseline data on key reform areas such as teacher and administrator evaluation, school leadership, teacher compensation, and professional development. The 2012 survey was redesigned to include a more in-depth focus on teacher evaluation in light of the Tennessee Educator Acceleration Model (TEAM), the state's new evaluation system implemented statewide during the 2011-2012 school year. The survey's sampling strategy was also modified in order to minimize the survey length for educators.

TEAM is the most widely used of the four evaluation systems across Tennessee that satisfies requirements of Tennessee's First to the Top Act of 2010 by providing a mechanism for

More About the Tennessee Consortium and Its Work

The Tennessee Consortium on Research, Evaluation, and Development was established in 2010 through Tennessee's Race to the Top grant, and is responsible for carrying out a detailed, focused program around key grant initiatives. As part of that effort, the Consortium is conducting 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.

While permission to reprint is not necessary, the recommended citation for these preliminary findings is: Tennessee Consortium on Research, Evaluation, and Development. (2012, July). *Educator evaluation in Tennessee: preliminary findings from the 2012 First to the Top Survey.* Nashville, TN: Pepper, M.J., Burns, S.F., & Springer, M.G.

A copy of the survey can be found on the Consortium's website through this link www.tnconsortium.org/projects-publications/projects-publications/first-to-top-survey/

Any errors within this report remain the sole responsibility of the authors.

annual evaluation of all school personnel. By law, results of these evaluations must be used in personnel decisions regarding tenure, promotion, and compensation. A teacher's overall evaluation score is based on data from three sources: 35 percent on student growth data (e.g., TVAAS), 15 percent on student achievement data (e.g., TCAP scores, graduation rates), and the remaining 50 percent on qualitative measures such as a review of prior evaluations, personal conferences, and position observations.

Position observations are scored using a rubric based on the Teacher Advancement Program (TAP) rubric, grounded in the work of Charlotte Danielson. Much public discussion has surrounded TEAM, efforts required to implement the new observation system, and the efficacy of the rubric itself. The 2012 survey solicited experiences and perceptions of TEAM and other state-approved models from the perspective of those conducting observations, such as principals and assistant principals, as well as teachers and other school-based personnel being observed and evaluated. Preliminary findings from the survey are presented in this brief. More in-depth analysis of survey findings will be presented to the Tennessee Department of Education (TDOE) at a later time, which will also include a literature review and summary of the national policy context.

All certified school staff listed in the TDOE TEAM database were invited to participate in the 2012 First to the Top Survey, which was administered online.[1] Of those invited to participate, 27.3 percent of administrators (N=905) and 24.8 percent of non-administrators (N=16,705) responded to the survey. Respondents were presented with one of six versions which, in addition to soliciting responses to a set of core items regarding teacher evaluation, also contained a distinct module that covered one of the following topics: Great Teachers and Leaders, Professional Development, Data Systems & Resources to Support Instruction, Standards and Assessment & Knowledge of and Attitudes Towards Reform, Instructional Practices and Testing, and Teacher Compensation. For further information regarding the sampling process and sample representativeness, please see Appendix A.

The preliminary findings that follow are organized into five broad areas, each of which is defined by a key research question. Each of these broad questions is further defined by questions that focus more directly and specifically on a portion of the area being addressed. Both broad and more focused questions are outlined below.

1. With what level of fidelity were TEAM and other state-approved models implemented?

    - Research Question 1.1: Were the expected number of short observations conducted within each evaluation model? Were the expected number of lesson-length observations conducted?
    - Research Question 1.2: What was the typical duration of short observations? Of lesson-length observations?
    - Research Question 1.3: Who conducted short observations? Who conducted lesson-length observations?

1 Although the database shares its name with the TEAM evaluation model, it includes data on certified staff from all state-approved evaluation models used within Tennessee during the 2011-12 school year.

2. What was the timeliness, duration, and substance of observation and evaluation feedback, and to what extent did it inform professional development?

- Research Question 2.1: How much time was spent reviewing observation feedback by teachers, and how quickly was it delivered?
- Research Question 2.2: What guided evaluation feedback?
- Research Question 2.3: What topics did feedback cover, and what was its perceived value?
- Research Question 2.4: To what extent did evaluation feedback inform professional development activities?

3. Who served as observers, and to what extent did they feel prepared to conduct observations?

- Research Question 3.1: Who served as observers?
- Research Question 3.2: How many hours of training were provided to observers?
- Research Question 3.3: How prepared did observers feel to carry out specific components of the teacher evaluation process?

4. How much time did various evaluation components require, and to what extent did observers and those being observed report their evaluation model is burdensome?

- Research Question 4.1: How much time did observers and those being observed report they spend on evaluation components?
- Research Question 4.2: To what extent did observers report that the evaluation model used in their school is a burden?

5. What is the level of teacher and administrator understanding of and support for aspects of the team evaluation model, and the level of support for utilizing its results for policy decisions related to teachers?

- Research Question 5.1: To what extent did teachers and administrators feel that the TEAM evaluation rubric promotes attainable goals, is comprehensive, and is adequately descriptive?
- Research Question 5.2: To what extent did teachers and administrators understand and support the components of the TEAM evaluation model?
- Research Question 5.3: How did teachers and administrators believe results from the TEAM evaluation model should be utilized?
- Research Question 5.4: To what extent did teachers and administrators understand and support the TEAM evaluation processes and model?

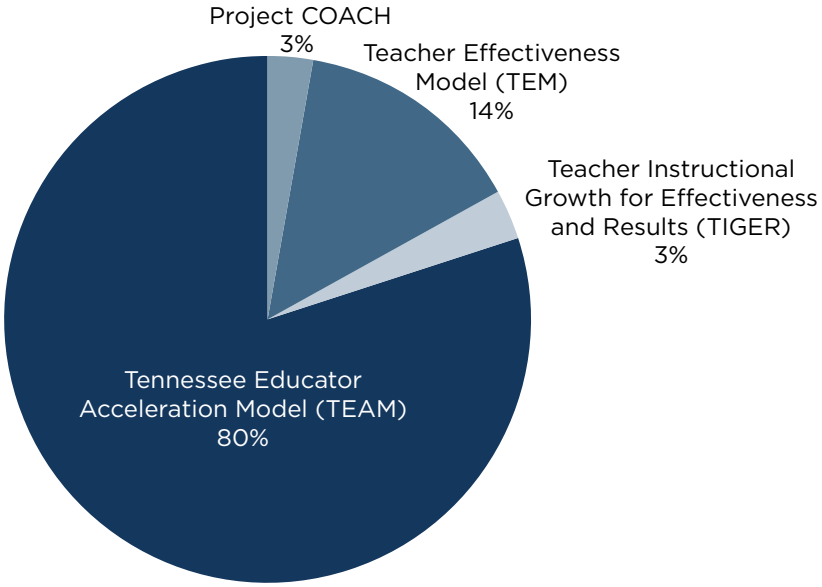
I. The Fidelity of Implementation

These analyses begin with an investigation into the fidelity of implementation of TEAM and other state-approved evaluation systems. Results are presented from teacher responses due to the high degree of procedural fidelity contained within. This section specifically addresses the following three research questions:

- Research Question 1.1: Were the expected number of short observations conducted within each evaluation model? Were the expected number of lesson-length observations conducted?
- Research Question 1.2: What was the typical duration of short observations? Of lesson-length observations?
- Research Question 1.3: Who conducted short observations? Who conducted lesson-length observations?

The sample within this section was first limited to the 14,902 respondents who were not administrators, and not specialists (e.g., counselors, instructional coaches, librarians). An additional 553 teachers who indicated that they served as an official evaluator of their peers did not answer questions concerning the fidelity of their individual teaching observations. Finally, an additional 198 respondents failed to adequately indicate which evaluation model was being utilized at their school, and are discarded from the analysis. The remaining 14,151 teachers serve as the basis of the analyses within this section, and are divided by model as shown in chart 1.0.[2]

Chart 1.0: Percent of Teachers by Model (n = 14,151)



2 A table showing a comprehensive analysis of the representativeness of respondents is included within Appendix A. Note that this analysis compares administrators and non-administrators. A comparison of teacher survey respondents with the universe of Tennessee teachers is hindered by inaccuracies within the variable Assignment Code tracked within the TDOE Educator Information System (EIS).

Research Question 1.1: The number of short and lesson-length observations[3]

Table 1.1 presents the total number of short and lesson-length observations experienced by survey respondents, reported by evaluation model. Over two-thirds of teachers in TEAM districts experienced short observations between one and three times; slightly less than two-thirds of them were observed with lesson-length observations exactly twice. One also notes that the observation pattern of teachers within TIGER and TEM schools varied greatly, while teachers in schools utilizing the COACH model report patterns consistent with this model's design.

Table 1.1: Number of Short & Lesson-Length Observations by Model

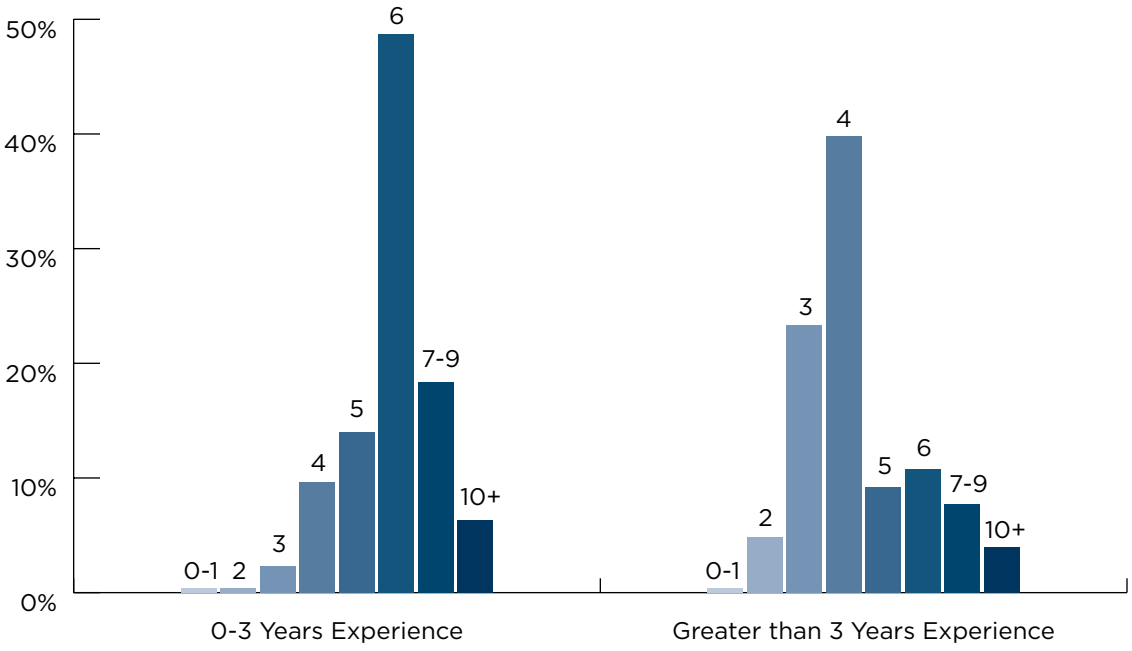
Number of Short Observations	TEAM	TIGER	TEM	COACH
0	7.2%	7.2%	20.2%	0.5%
1	23.5%	5.7%	9.7%	1.0%
2	37.8%	12.7%	20.0%	2.0%
3	16.2%	10.7%	9.8%	2.3%
4	9.1%	39.5%	26.2%	3.8%
5-9	4.4%	19.2%	11.1%	80.1%
10+	1.9%	5.0%	2.9%	10.2%
Number of Lesson-Length Observations	TEAM	TIGER	TEM	COACH
0	0.7%	27.3%	11.8%	75.5%
1	9.1%	14.9%	7.0%	7.6%
2	64.8%	33.8%	28.0%	7.1%
3	18.4%	10.2%	12.9%	5.1%
4	5.9%	10.4%	32.5%	0.8%
5+	1.1%	3.5%	7.8%	4.0%

Notes: Lesson-Length Observations n = 13,998; Short Observations n = 13,917

3 As defined on the survey: A short observation is one that is part of the teacher evaluation process during which one or more evaluators observes for less than 20 minutes what is generally NOT a complete lesson. A lesson length observation is one that is part of the teacher evaluation process, during which one or more evaluators observes what is generally a complete lesson.

Given that each of the evaluation models treats apprentice teachers differently than those who have a professional license, a secondary question within this subsection delves into the number of observations by teacher years of experience. The teacher-reported years of experience were first recoded into two values: 0-3 years and greater than 3 years. This variable was then compared with another new variable, which added the number of short observations and lesson-length observations for each teacher. The results, shown in Chart 1.1, reveal patterns that are generally consistent with expectations—with a modal value of six observations for apprentice teachers and modal value of four observations for teachers with a professional license.

Chart 1.1: Number of Total Observations, by Teacher Years of Experience (n = 13,404)



Research Question 1.2: What is the typical duration of short and lesson-length observations?

Teachers who indicated being observed at least once were also asked to indicate the duration of their "typical short observation" and "typical lesson-length observation". The results below reveal patterns that are consistent with expectations of the model.[4]

Chart 1.2.1: Length of Typical Short Observation (n = 12,587)

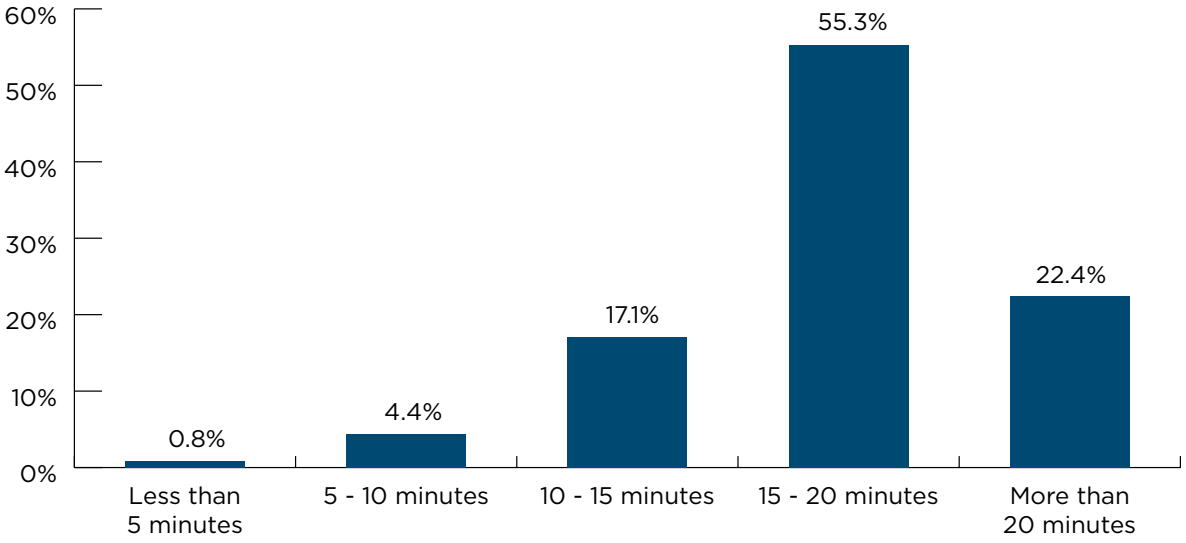
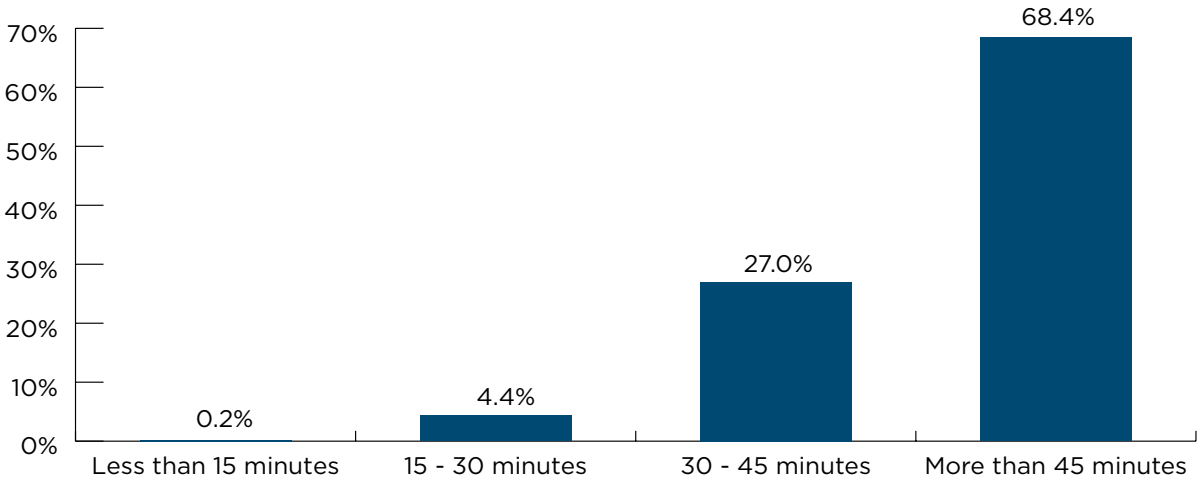


Chart 1.2.2: Length of Typical Lesson-Length Observation (n=13,122)



4 Consortium researchers also investigated whether or not the duration of observations varied by teacher years of experience. Teachers with less than three years of experience do report longer lesson-length observations, but not to a significant extent—71% report observations more than 45 minutes, compared with 66% of teachers with more than twenty years of experience. In contrast, teachers with more years of experience were more likely to report longer short observations than teachers with fewer years of experience. In summary, there is very little evidence that, on a per-observation basis, administrators are spending more time on a newer teacher's observation than a more experienced teacher's observation.

Research Question 1.3: Who conducted short observations? Who conducted lesson-length observations?

One of the more important outcomes from this report is an investigation into the degree to which the new evaluation models have placed an increased or decreased burden on educators. Data within the two charts below reveal who conducted observations of survey respondents. Note that respondents were asked to indicate all of the individuals who conducted an observation, and could indicate more than one answer choice. The task of evaluating appears to fall primarily to principals and assistant principals, with relatively few observations conducted by specialists and observers external to the school.

Chart 1.3.1: Short Observation Observers (n = 13,121)

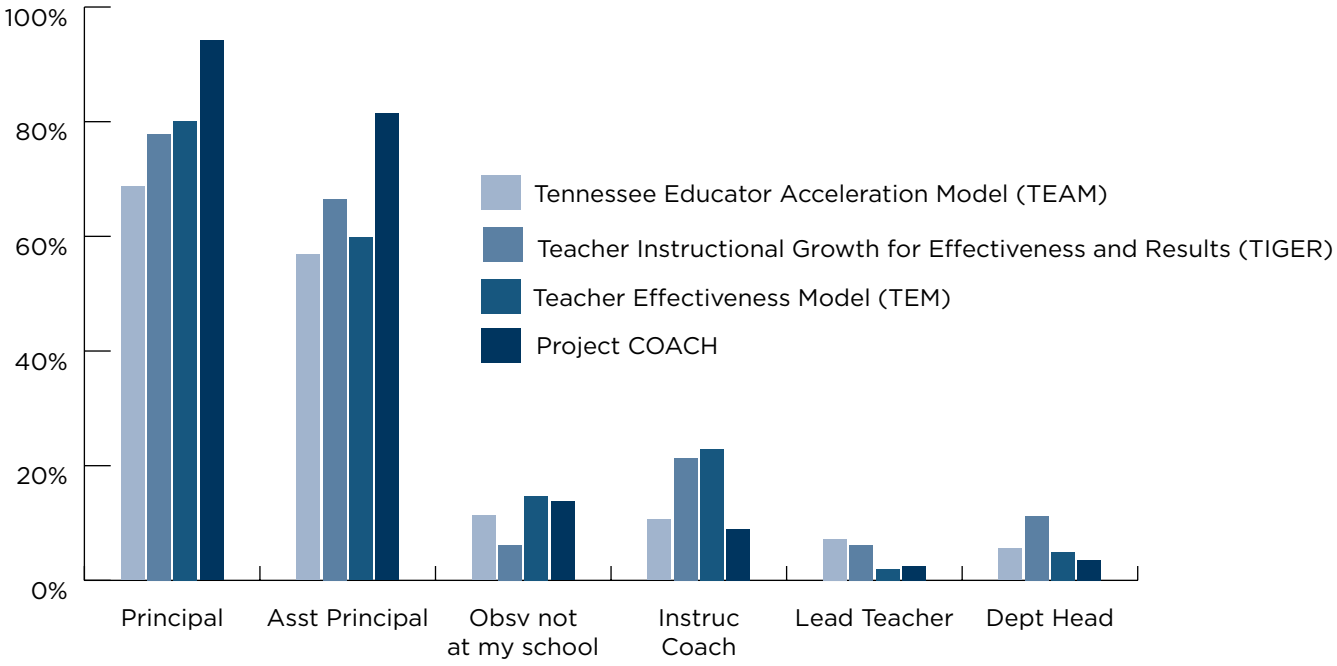
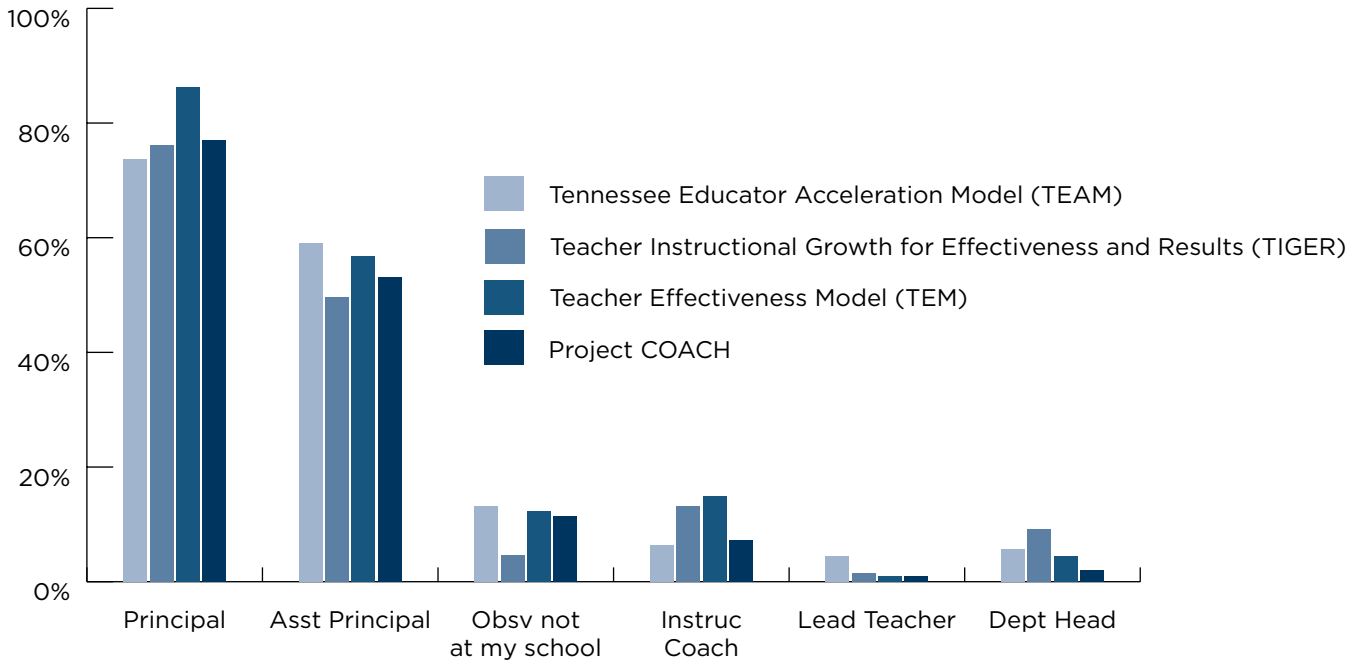


Chart 1.3.2: Lesson-Length Observation Observers (n = 13,121)



Notes: A small number of respondents who marked 'Other' are excluded, as are those who failed to indicate any observer position.

Taken together, results from Section 1 reveal that teachers were observed the expected number of times and for the expected duration. Additionally, principals and assistant principals conducted the vast majority of observations.

II. Characteristics of Feedback and Its Use

Analyses continue with an investigation into questions surrounding issues of observation and evaluation feedback. This section specifically addresses the following four research questions:

- Research Question 2.1: How much time was spent reviewing observation feedback by teachers, and how quickly was it delivered?
- Research Question 2.2: What guided evaluation feedback?
- Research Question 2.3: What topics did feedback cover, and what was its perceived value?
- Research Question 2.4: To what extent did evaluation feedback inform professional development activities?

Research Question 2.1: How much time was spent reviewing feedback by teachers, and how quickly was it delivered?

Research question 2.1 investigates the time with which teachers received and reviewed feedback both from short and lesson-length observations.[5] Feedback within the COACH model, by design, appears shorter than other models, and most COACH teachers report no lesson-length observations. Otherwise there is a fairly even distribution of time spent receiving/reviewing feedback throughout the utilized scale.

5 The definition of a short and lesson-length observation, as utilized on the survey, is outlined in Research Question 1.1.

Chart 2.1.1: Typical time spent receiving and/or reviewing feedback following each SHORT observation, by model

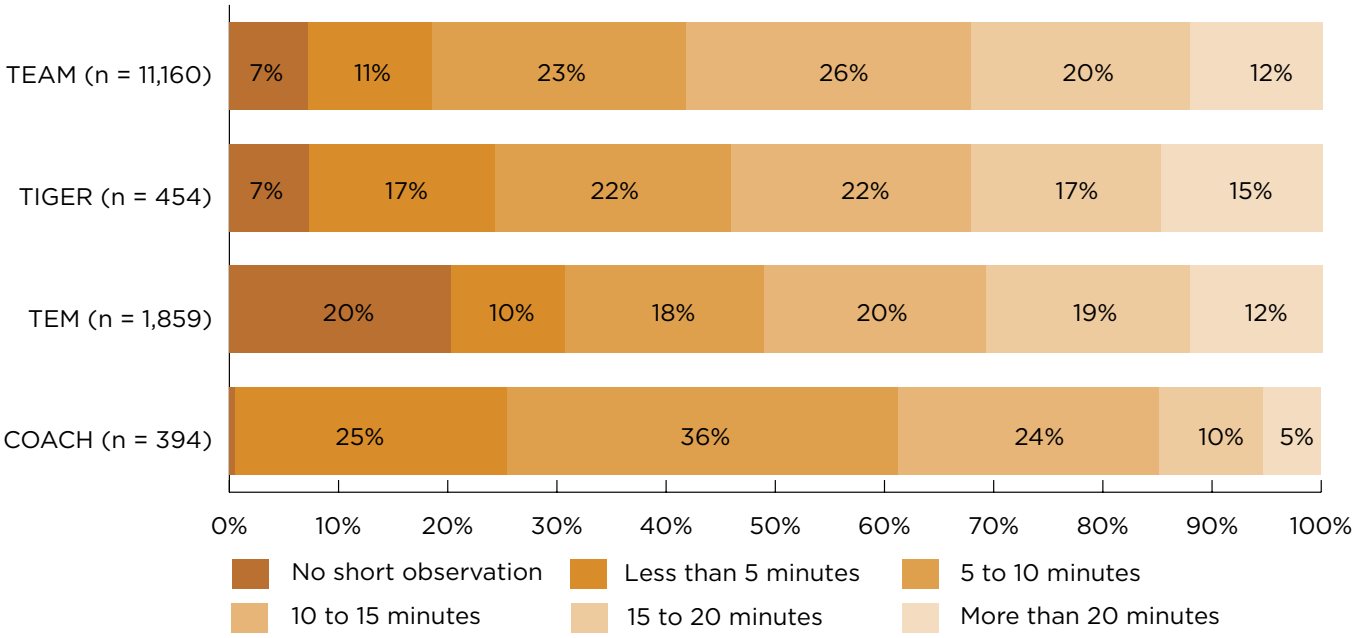
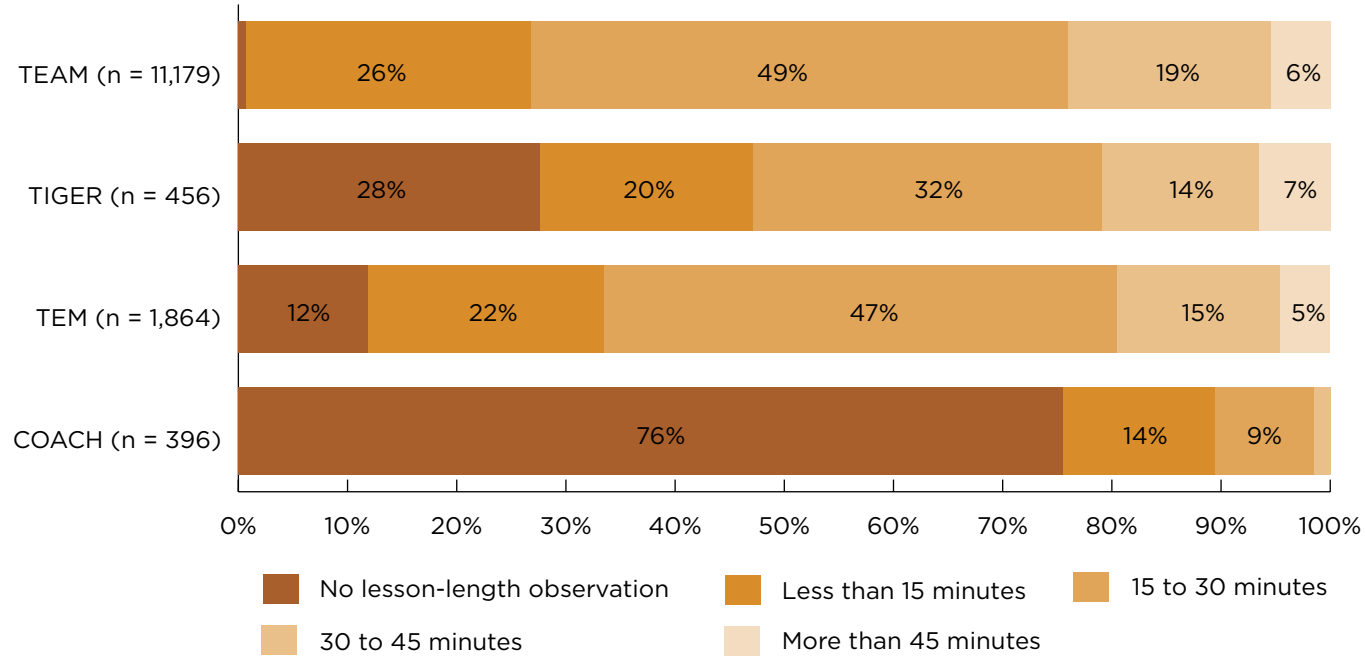


Chart 2.1.2: Typical time spent receiving and/or reviewing feedback following each LESSON-LENGTH observation, by model



As shown in the Table 2.1 below, both verbal and written feedback is typically provided to the vast majority of teachers within ten days of short and lesson-length observations.

Table 2.1: Time between observation and feedback in both SHORT and LESSON-LENGTH observations.

How much time passed before you received the following types of feedback following SHORT observations?			
	0-10 days	Over 11 days	I typically did not receive this type of feedback
Verbal feedback (n = 12,477)	86%	9%	5%
Written feedback (n = 12,371)	83%	12%	6%
How much time passed before you received the following types of feedback following LESSON-LENGTH observations?			
	0-10 days	Over 11 days	I typically did not receive this type of feedback
Verbal feedback (n = 13,064)	87%	11%	2%
Written feedback (n = 12,967)	83%	15%	3%

Research Question 2.2: What guides evaluation feedback?

Generally, designed feedback processes are being followed. Chart 2.2.1 presents teacher responses to the question "Has your numerical score for each of this year's observations been shared with you?", and reveals that the vast majority of teachers do receive their scores. Chart 2.2.2 reveals that the model rubric is generally utilized to guide the conversation between the observer and the teacher.

Chart 2.2.1: The sharing of numerical scores, by model

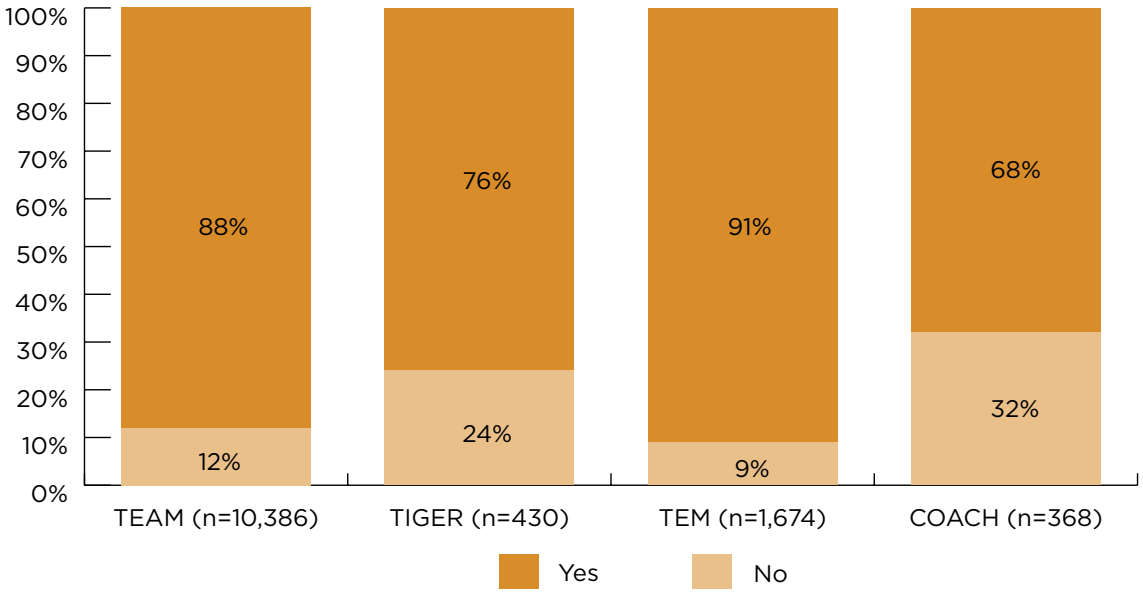
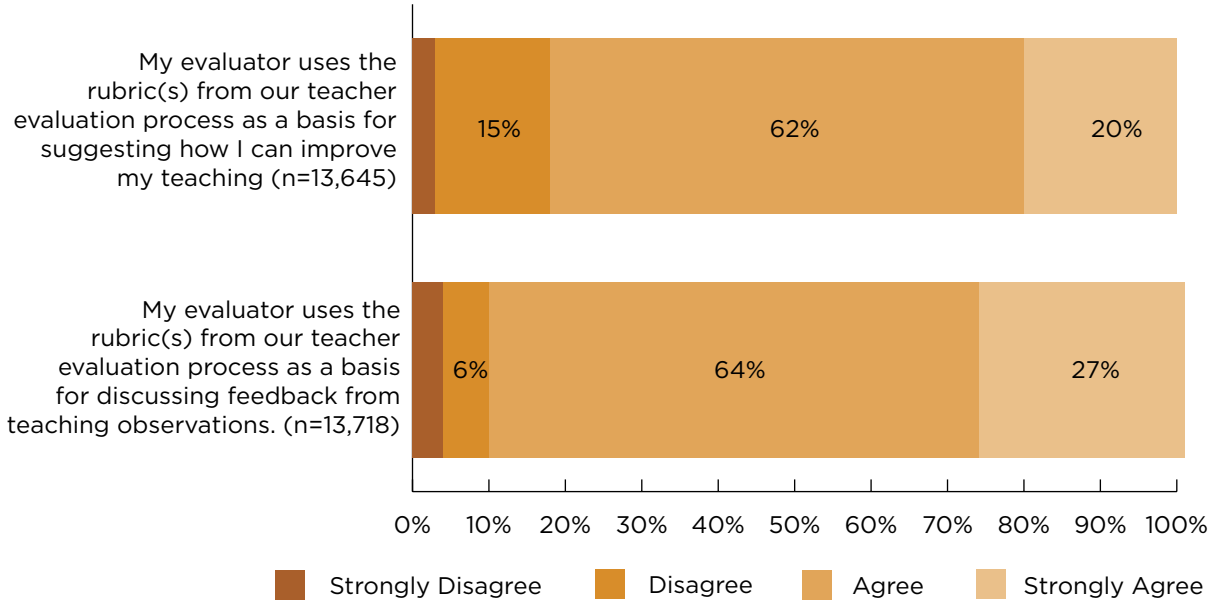


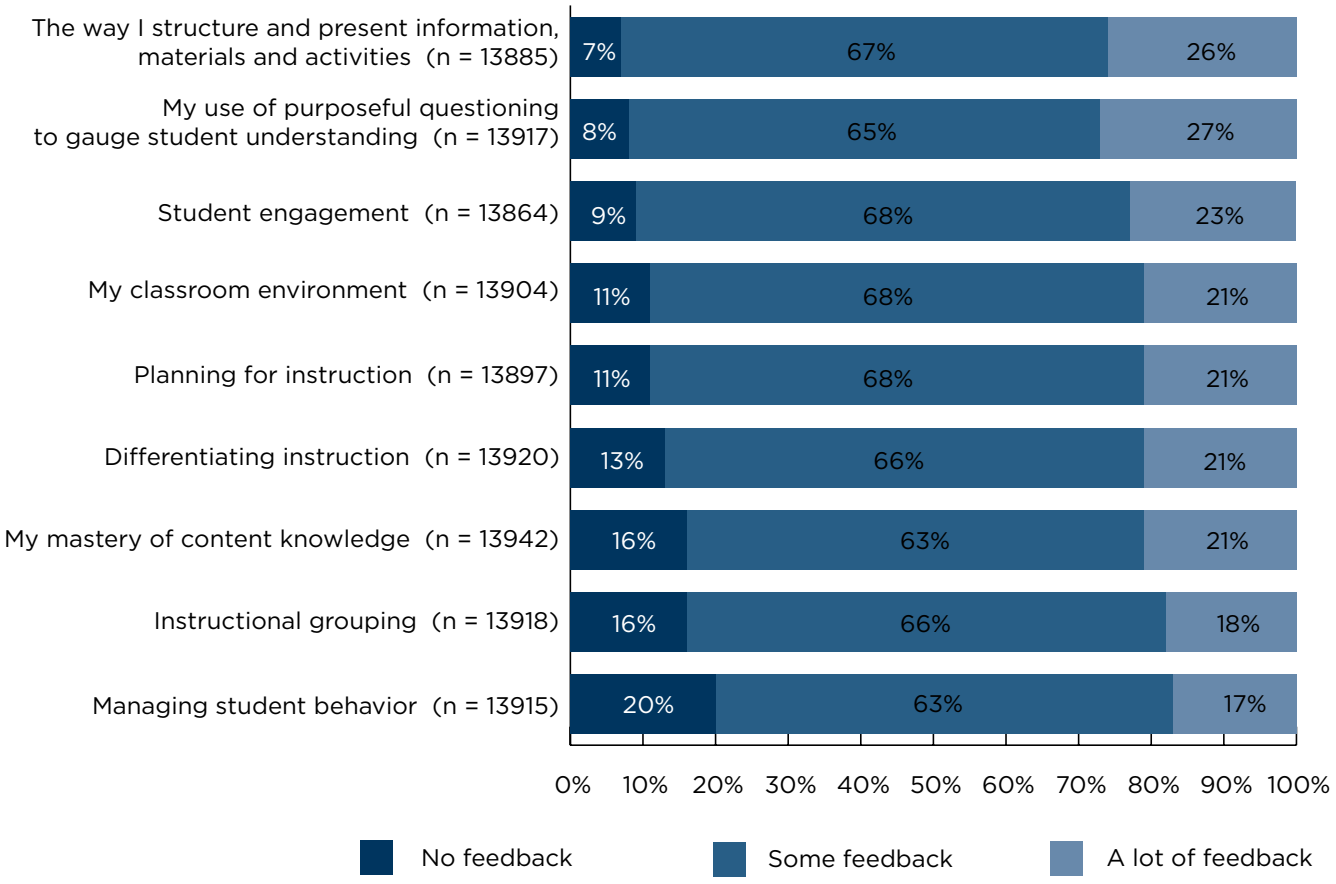
Chart 2.2.2: The utilization of the rubric to guide feedback



Research Question 2.3: What topics did feedback cover, and what was its perceived value?

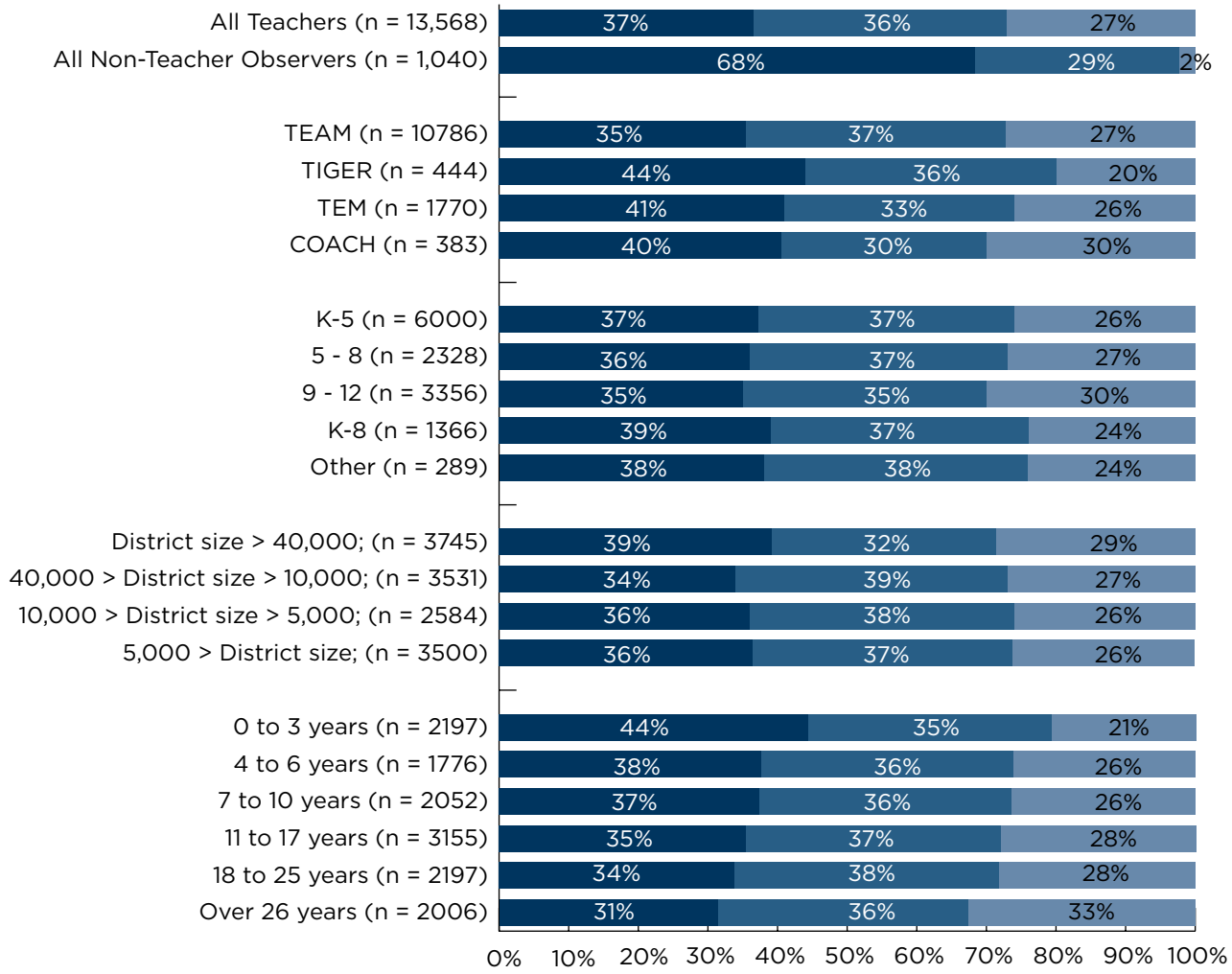
Survey respondents were asked to indicate whether they received 'No feedback', 'Some feedback', or 'A lot of feedback' concerning the educational topics below (listed from highest reported amount of feedback to lowest amount of feedback). These responses indicate that feedback was wide-ranging but not always at a level of depth.

Chart 2.3.1: Amount of feedback by topic



Respondents also indicated whether they perceived their feedback to be equally focused on improving their teaching and making a judgment, or whether it had a primary focus on one of these two goals. Results are shown in the graph below. Slightly more teachers responded that feedback was equally weighted or was more focused on improving teaching. Administrators answered a similar question "The feedback I provided was focused more/equally focused…", and more than two out of three of them responded that their feedback was focused more on improving teaching—but this proportion was not mirrored by teacher responses. Responses by model and by school and teacher demographics are also shown, and reveal little variation.

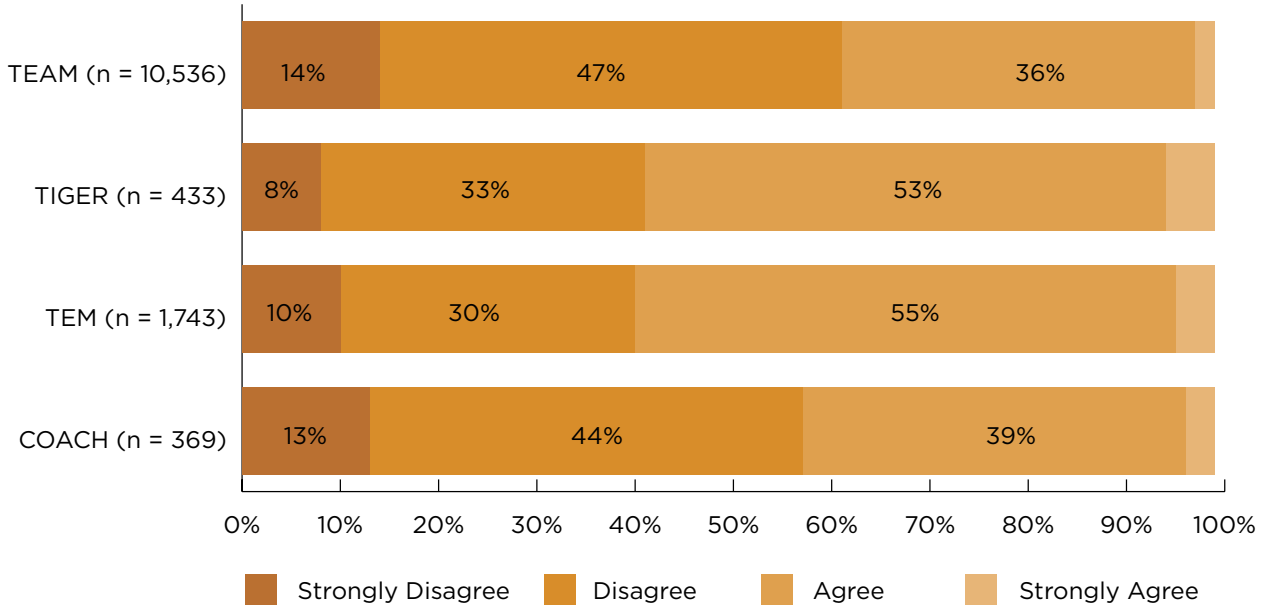
Chart 2.3.2: Extent that feedback was focused on improving teaching or making a judgment, by Teacher, Observer, Model, Tier, District Size, and Teacher Years Experience



Research Question 2.4: To what extent did feedback from evaluations inform professional development activities?

This section closes by presenting data concerning the extent that teachers report feedback informing their professional development activities, by evaluation model. Across all models, less than five percent of teachers 'Strongly Agree' with the statement that their professional development activities are informed by feedback from their evaluation. Teachers within TEAM and COACH disagree with this statement at higher rates than teachers in TIGER and TEM.

Chart 2.4.1: Extent that feedback informs professional development, by model



III. A Portrait of Observers and Their Level of Preparation

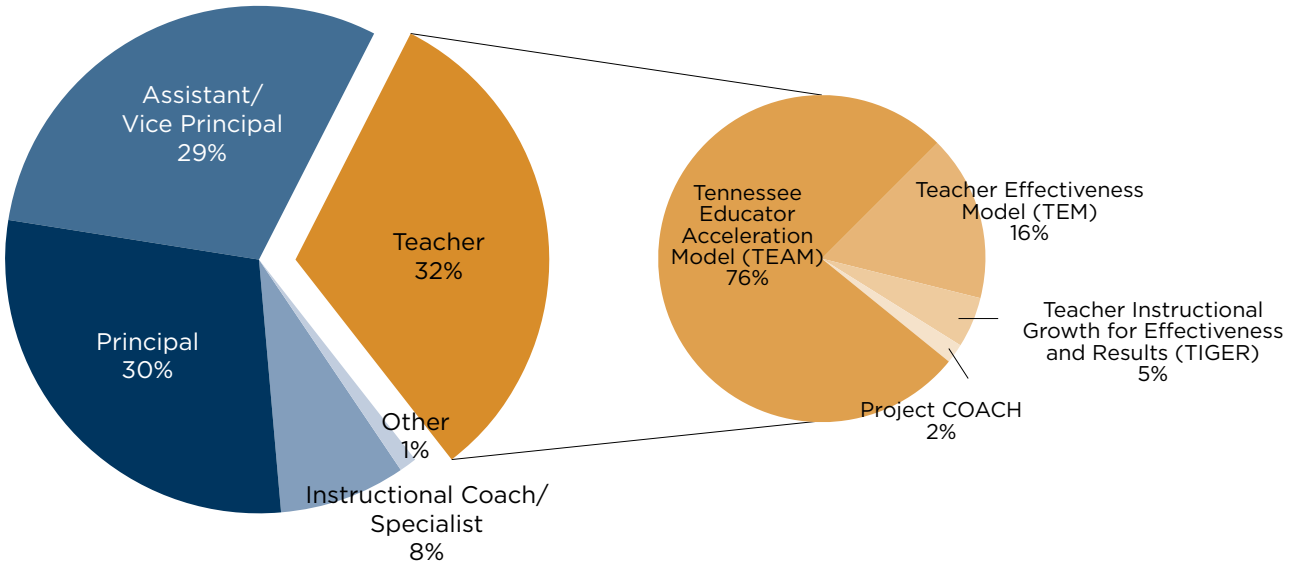
This section investigates the positions and characteristics of the individuals who self-identified as observers, then describes the extent that these persons felt adequately prepared to conduct different aspects of the observation. Question to be explored include:

- Research Question 3.1: Who served as observers?
- Research Question 3.2: How many hours of training were provided to observers?
- Research Question 3.3: How prepared did observers feel to carry out specific components of the teacher evaluation process?

Research Question 3.1: Who served as observers?

The first research question within this section investigates the educational position observers indicated on the survey as their primary role. These results are shown within the left-hand pie in the graph below. Surprisingly, 32% of observers self-identify on the survey as teachers. To further investigate this, this subset of 492 survey respondents were separated and split by evaluation model used, and these results are shown within the right-hand pie. This reveals that 76% of these teacher-observers were in schools utilizing the TEAM evaluation model. Further analyses reveal that these teacher-observers were utilized slightly more often within the 9-12 tier (39% of observers) than in the K-5 and 6-8 tiers (31% and 30%, respectively), and were not clustered by district.

Chart 3.1.1: Observers by Position, with Teacher Observers also Shown by Model (n = 1,553)



These findings might seem contradictory to charts 1.3.1 and 1.3.2, which show very few individuals who were observed reporting that a lead teacher had observed them. This contradiction is resolved, however, by investigating the number of observations conducted by observer position. One quickly notes that teachers, on average, conduct many less observations than administrators and instructional coaches. While the absolute number of teacher observers may be comparable to the absolute number of both principal observers and assistant principal observers, they primary burden of conducting observations falls on administrators.

Chart 3.1.2: Number of Short Observations Completed, by Observer Position (n = 1,535)

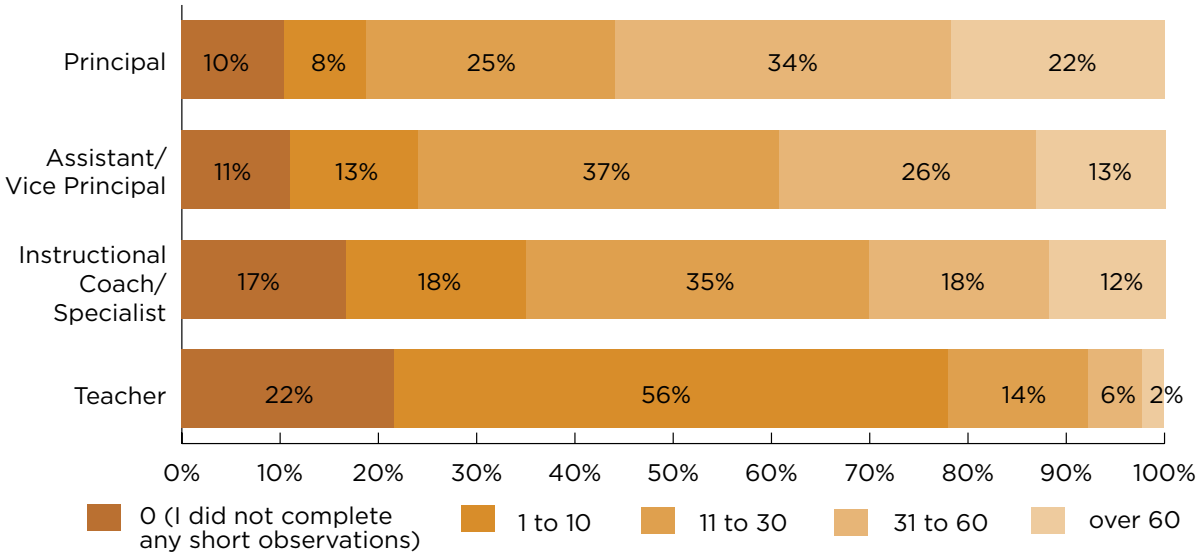
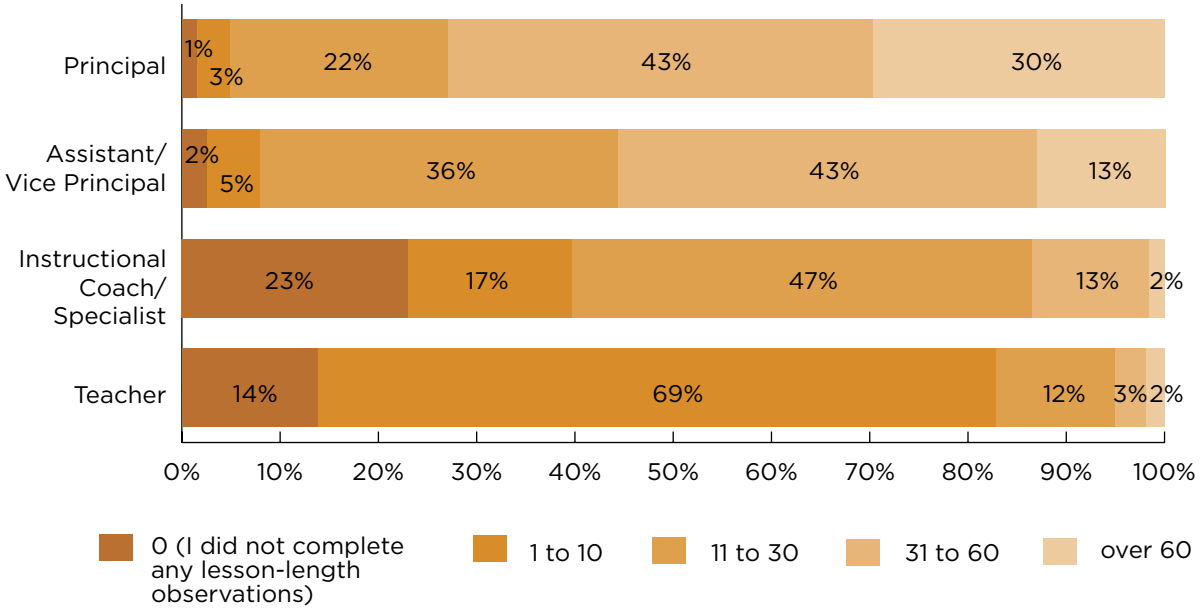


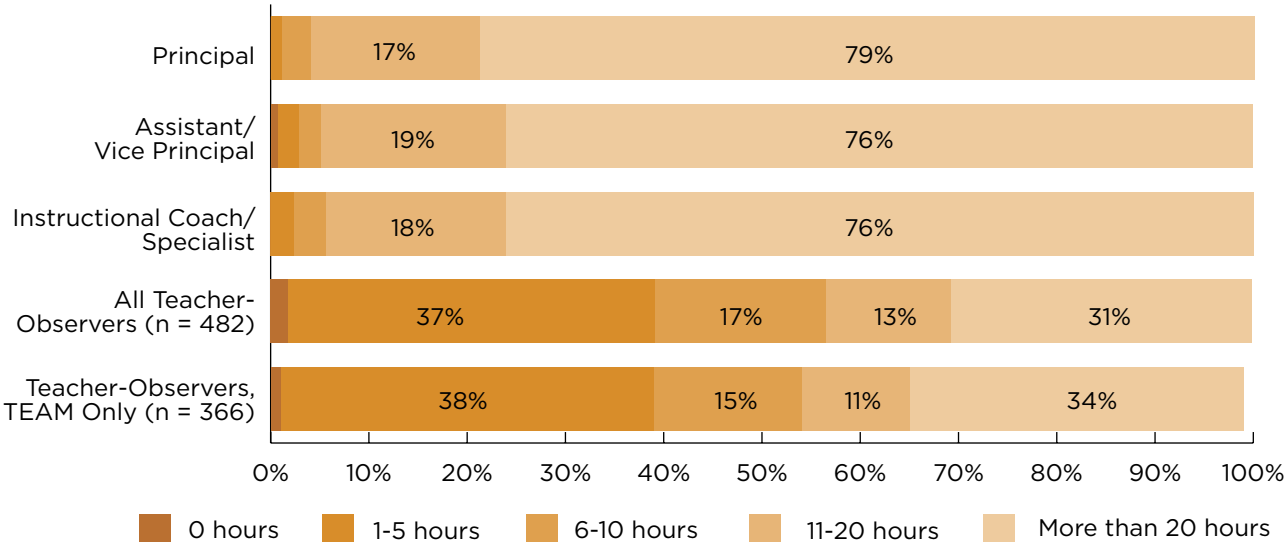
Chart 3.1.3: Number of Lesson-Length Observations Completed, by Position (n = 1,525)



Research Question 3.2: How many hours of training were provided to observers?

Survey results indicate that an overwhelming percentage of principals, assistant principals, and instructional coaches/specialists received greater than ten hours of training on their evaluation model, with three out of four receiving more than twenty hours of training. Both teacher-observers generally and teacher-observers within the TEAM evaluation model, on average, received many fewer hours of training that administrators and specialists. There was no substantial difference in the amount of training received by educators when separated by tier or by the size of the district (not shown).

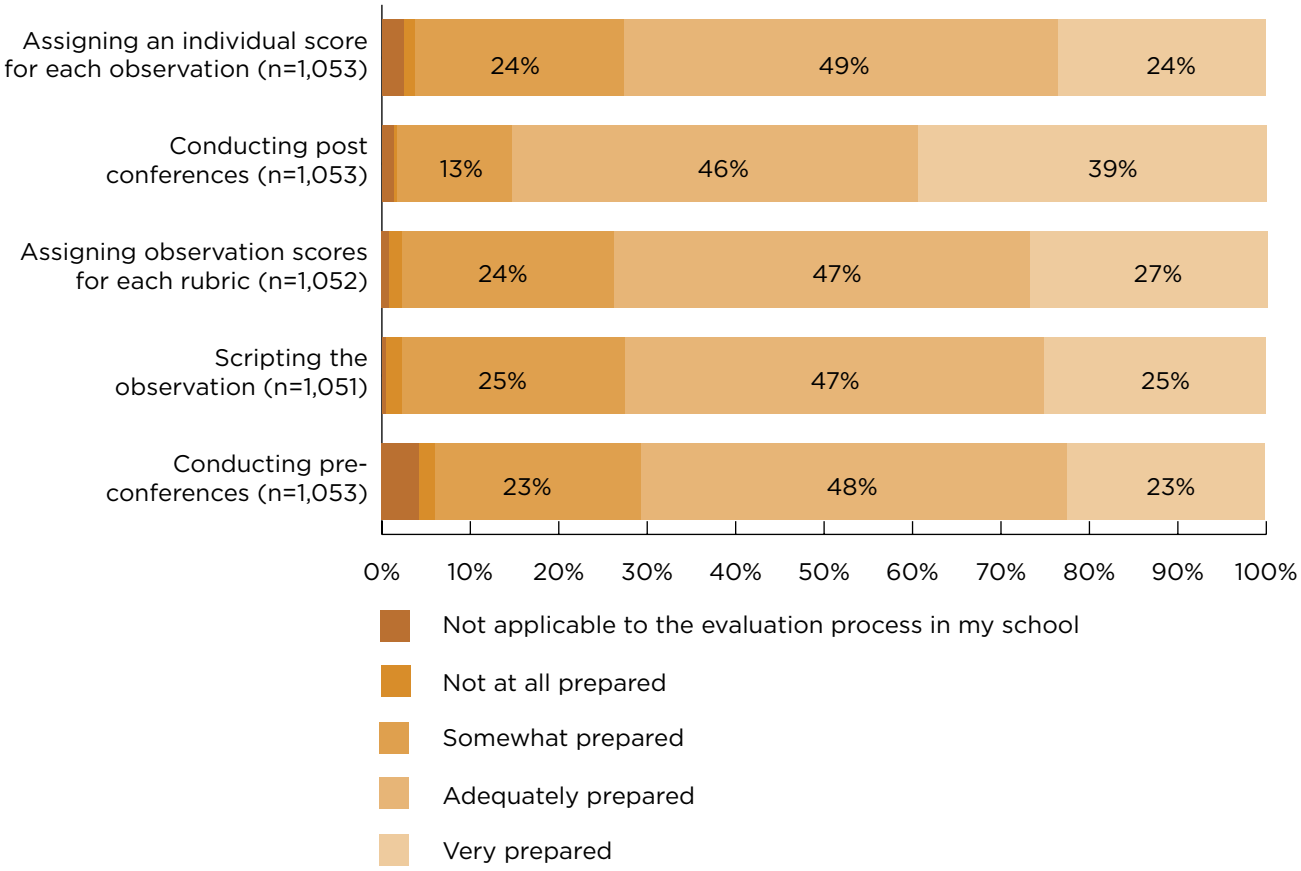
Chart 3.2.1: Number of Hours of Training Received by Position (n = 1,545)



Research Question 3.3: How prepared did observers feel to carry out specific components of the teacher observation and evaluation process?

The final graph in this section portrays responses from principal, vice-principal, and instructional coach observers on the extent that they felt prepared to conduct specific components of the observation evaluation process.[6] Paired with the graph above, these results portray survey respondents who received a significant amount of training and who felt adequately prepared by it to conduct evaluations within their schools. Results were found to be similar by Tier.

Chart 3.3.1: Reported Readiness by Evaluation Component - excludes Teachers



6 Due to the relatively small number of observations conducted by teachers—and the divergent nature of their training patterns—teacher observers are excluded from this analysis.

IV. Evaluation Components and Issues of Time

Relative to what they replaced, the new evaluation models used within Tennessee, along with the legislative requirement of evaluating every educator annually, necessitate a greater overall time commitment from both observers and those being observed. In light of this, the extent of the time burden, by component and for both groups, is explored within Section IV. Due to the lack of consistency in the processes of short observations, data were collected on components of lesson-length observations only. The COACH model does not conduct lesson-length observations, and hence are excluded from the graphs within this section.

- Research Question 4.1: How much time did observers and those being observed report they spend on evaluation components?
- Research Question 4.2: To what extent did observers report that the evaluation model used in their school is a burden?

Research Question 4.1: How much time do observers and the observed report they spend on evaluation components?

The three graphs below show the reported time that both observers and those being observed spend on preparing for an observation; conducting a pre-conference; and providing (observer) or receiving and reviewing (the observed) feedback. Note first that roughly two out of three individuals who are observed in the TEAM model report spending more than 90 minutes preparing for an observation, and slightly less than one out of two report spending more than three hours preparing for an observation. Individuals who are being observed in other models do not report spending as much time preparing for an observation, and observers generally spend much less time preparing for observations than those who are observed.

Observers and those being observed spend relatively similar amounts of time related to the pre-conference, although those who are observed report spending slightly more time than observers. This finding is mirrored within feedback findings, with those who are observed reporting more time spent on this component than observers.

Chart 4.1.1: Amount of time spent preparing for a typical, announced, lesson-length observation, by model, by observed and by observer

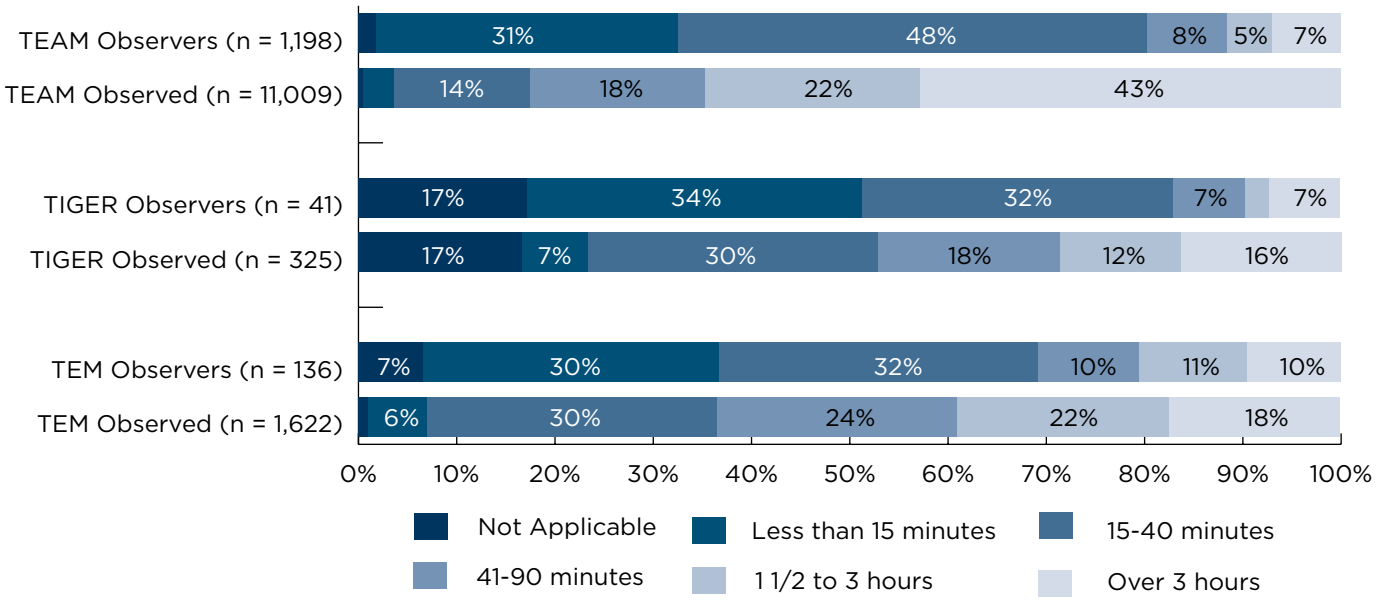


Chart 4.1.2: Amount of time spent on a typical pre-conference, by model, by observed and by observer

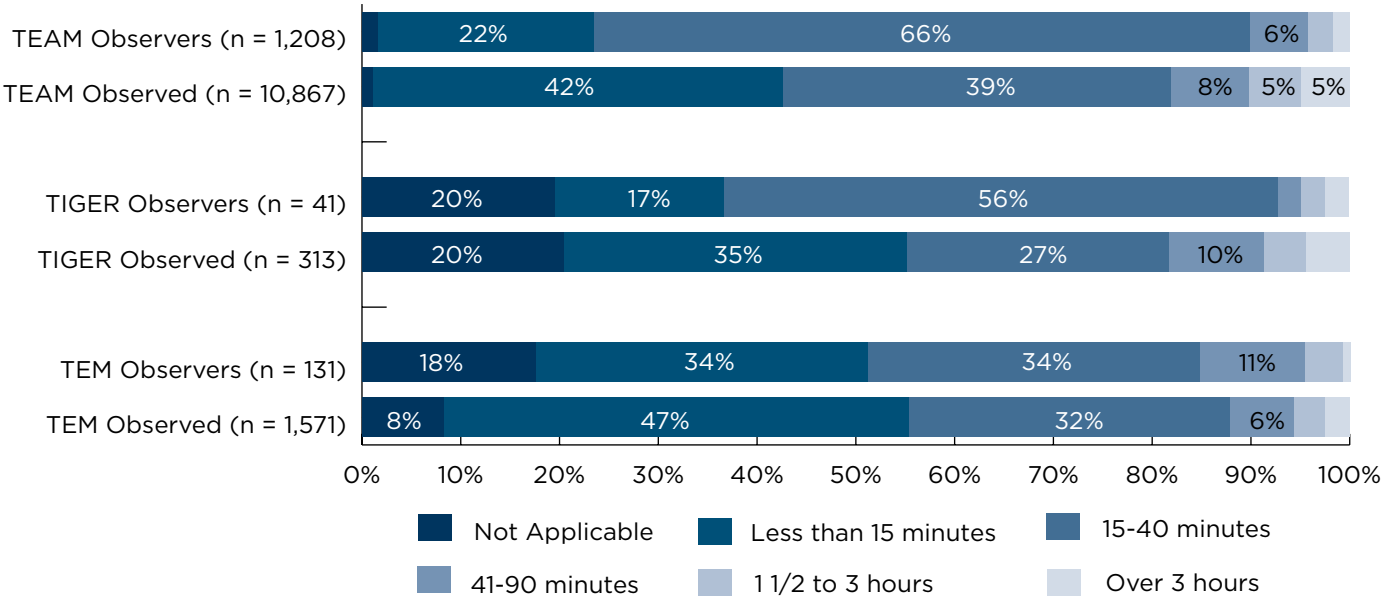
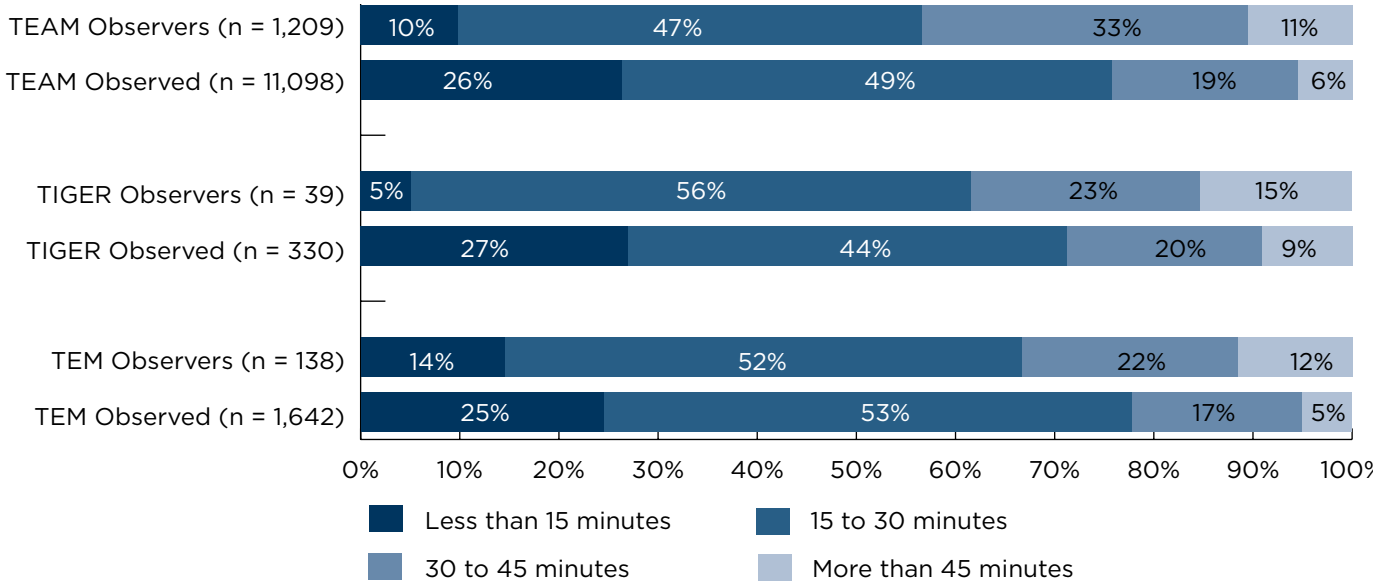


Chart 4.1.3: Amount of time spent providing (observer)/ receiving and reviewing (observed) feedback

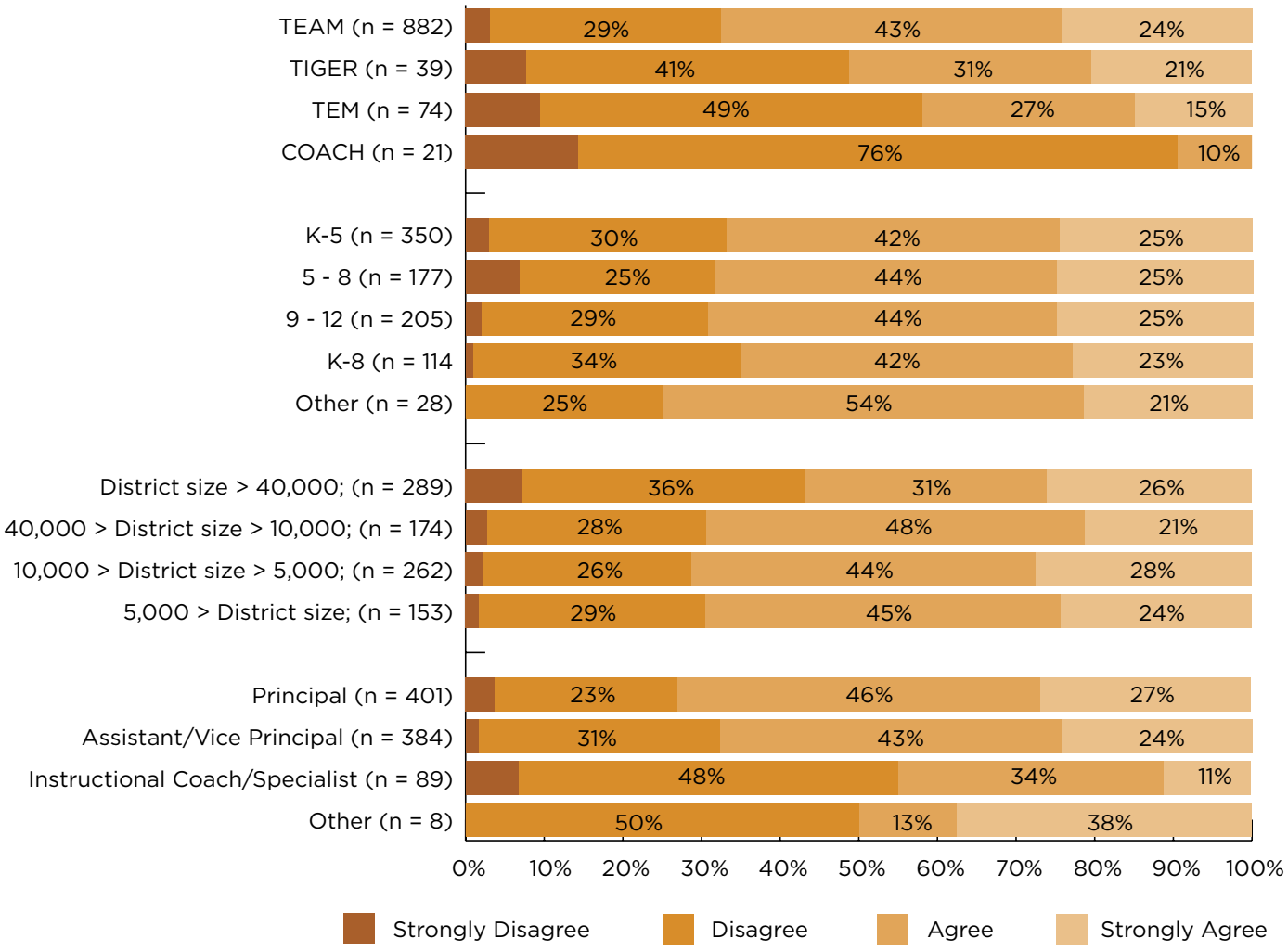


Research Question 4.2: To what extent do observers report that the evaluation model used in their school is a burden?

The survey also asked observers to rate the extent that the teacher evaluation process is burdensome for them, and results are shown by model and by school category in the graph below. Observers in the COACH model appear the least burdened by the evaluation system.

Excluding the by-model results, all data below are limited to administrators and instructional specialists in TEAM only, and find consistent results by tier and, to a lesser extent, district size. Instructional specialists report feeling less burdened by TEAM than administrators.

Chart 4.2.1: Extent of the agreement of administrator and instructional specialist observers that the evaluation process is burdensome



V. Understanding of and Support for the TEAM Evaluation Model

As the annual evaluation of all teachers and administrators moves forward in Tennessee, it is critical to examine general understanding of and support for reforms. All of the data presented within this section comes from survey questions that have answer choices along a four point Likert scale {Strongly Disagree, Disagree, Agree, and Strongly Agree}. Individuals who self-identified as teachers within the survey are grouped together, as are "administrators" – indicating principals, vice principals, and instructional specialists who serve as an observer.[7] Results are presented for respondents in schools utilizing the TEAM model only.

Key research questions include:

- Research Question 5.1: To what extent did teachers and administrators feel that the TEAM evaluation rubric promotes attainable goals, is comprehensive, and is adequately descriptive?
- Research Question 5.2: To what extent did teachers and administrators understand and support the components of the TEAM evaluation model?
- Research Question 5.3: How did teachers and administrators believe results from the TEAM evaluation model should be utilized?
- Research Question 5.4: To what extent did teachers and administrators understand and support the TEAM evaluation processes and model?

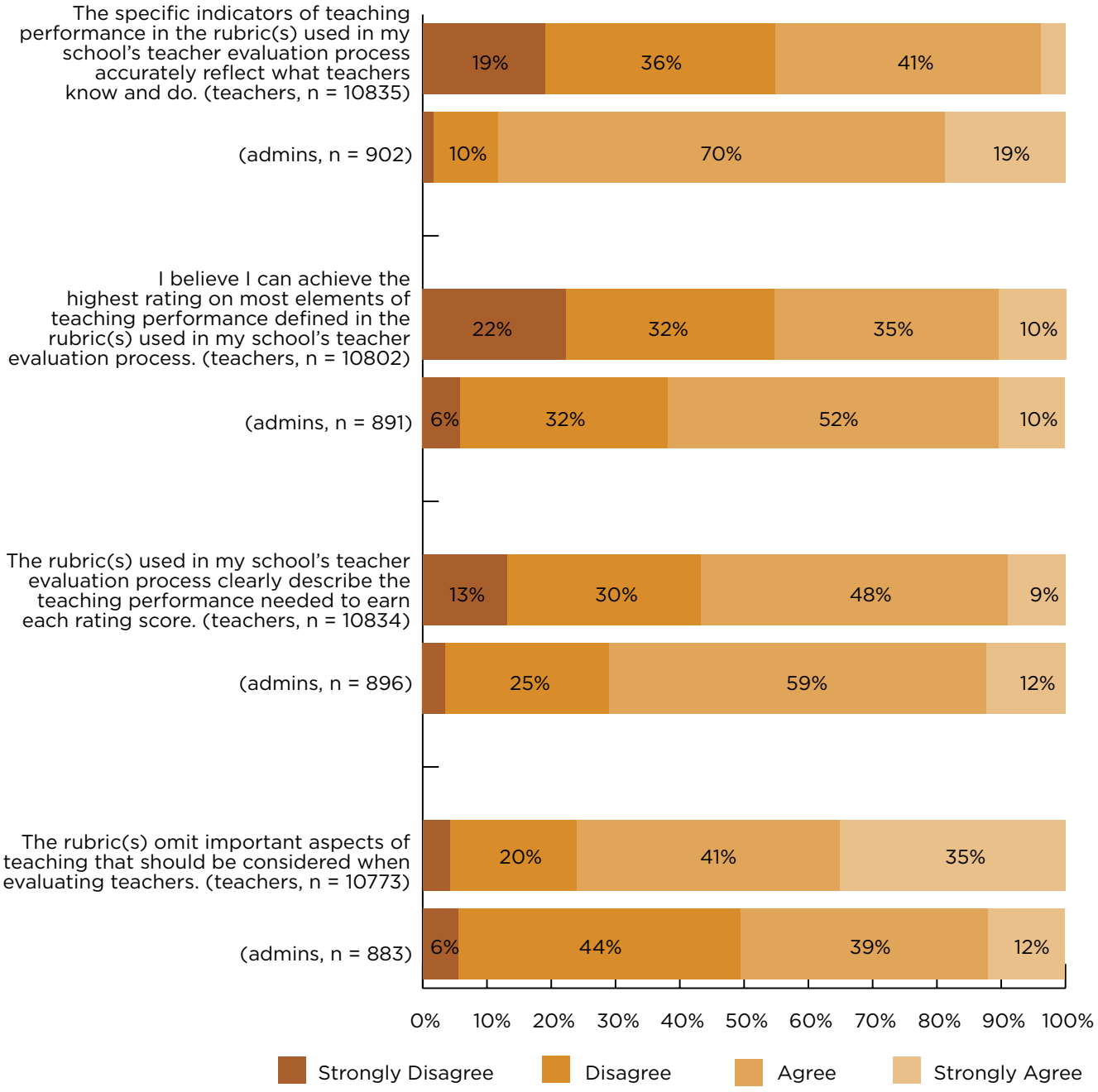
Research Question 5.1: To what extent do teachers and administrators feel that the TEAM evaluation rubric promotes attainable goals, is comprehensive, and is adequately descriptive?

The results within the graph below reveal that administrators are generally positive about the TEAM rubric, and are relatively more positive than teachers. More than three out of four teachers perceive that the rubric "omits certain aspects of teaching", although this perception is shared only by one out of every two administrators. This disagreement between teachers and administrators is also mirrored within the question pertaining to whether the rubric accurately reflects what teachers know in do, with administrators being relatively more positive.[8]

7 As with earlier analyses, the small percentage of teachers who conducted observations are excluded from this analysis.

8 Administrator questions from the survey were modified appropriately—e.g., the first one began "My teachers believe that they can achieve the highest rating…"

Chart 5.1.1: Teacher and administrator perceptions of the TEAM rubric



Research Question 5.2: To what extent do teachers and administrators understand and support the components of the TEAM evaluation model?

Chart 5.2.1 shows teacher and administrator level of support for different scoring components within the TEAM model. Teachers indicate the lowest level of support for the 15% achievement measure and highest level of support for the growth measure, while administrators have the most confidence in the qualitative measures of teaching performance.

Chart 5.2.1: The level of agreement of those observed/observers with the statement: "I believe that the {individual scoring component of TEAM} included in my/my teacher's overall effectiveness rating accurately reflects my teaching/ a teacher's performance"

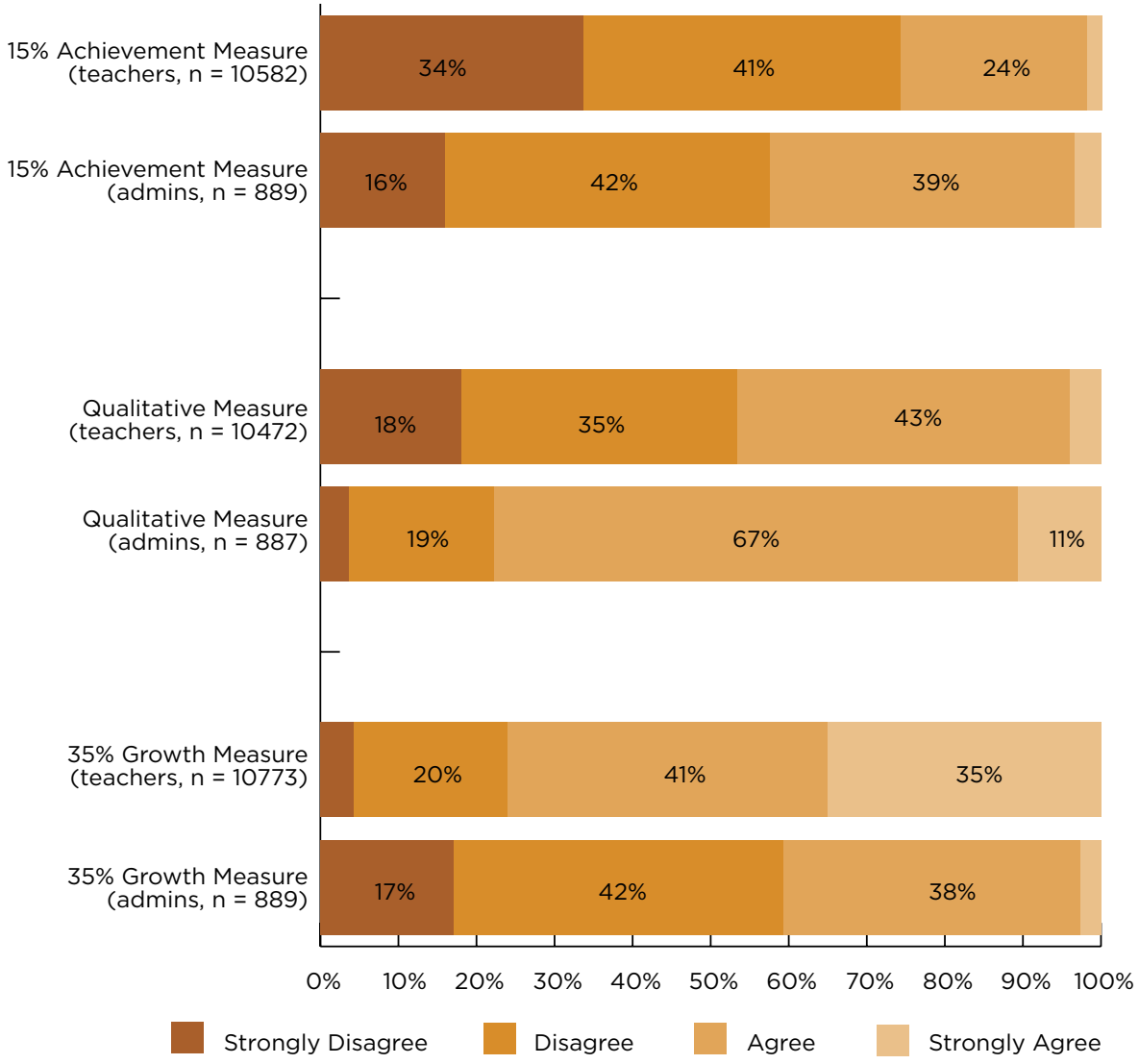
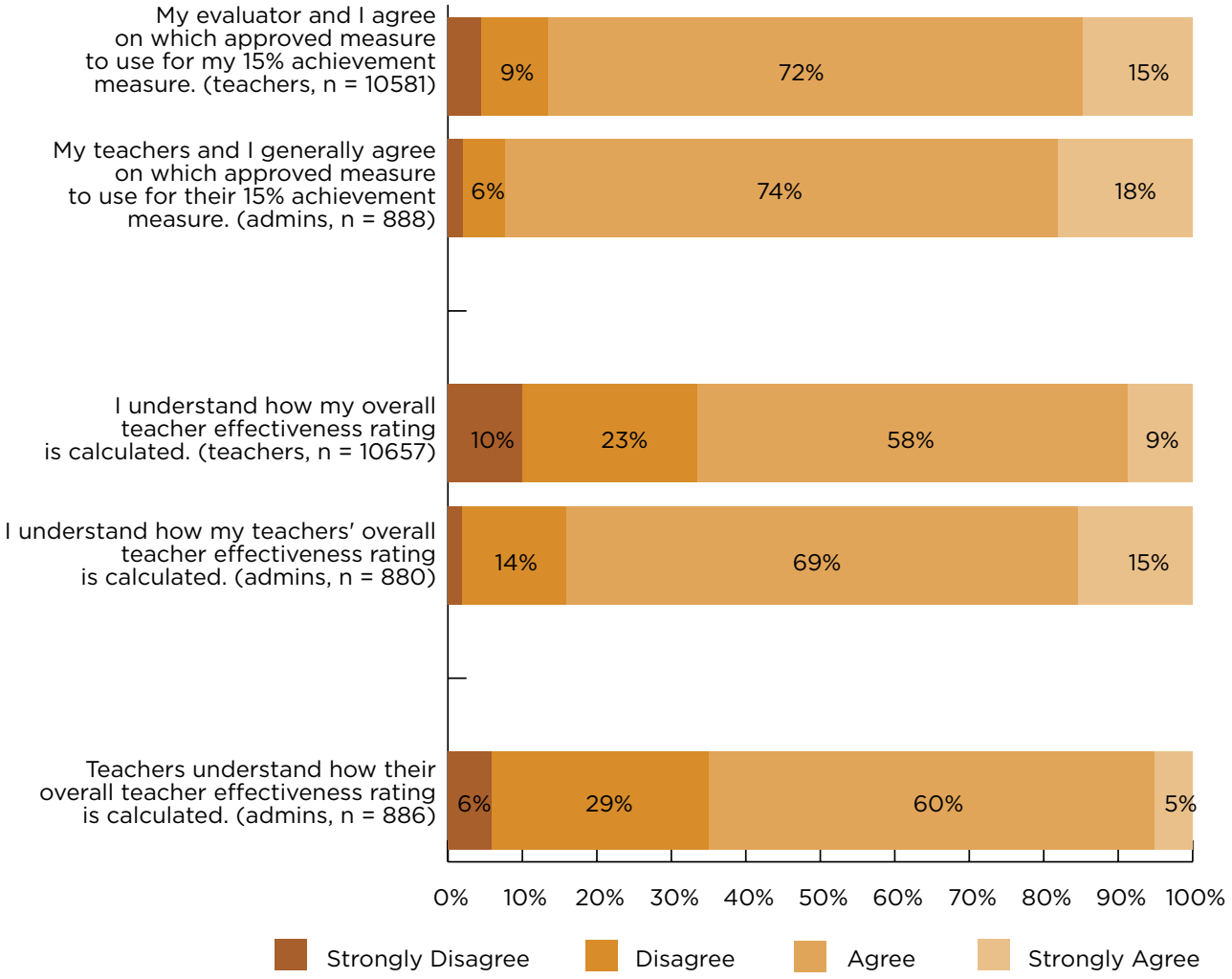


Chart 5.2.2 shows teacher and administrator level of understanding and agreement on components of the TEAM model. Approximately two out of three teachers claim an understanding of how their teacher effectiveness score is calculated. There also appears to be little disagreement between teachers and administrators concerning what to utilize as a 15% approved additional measure.

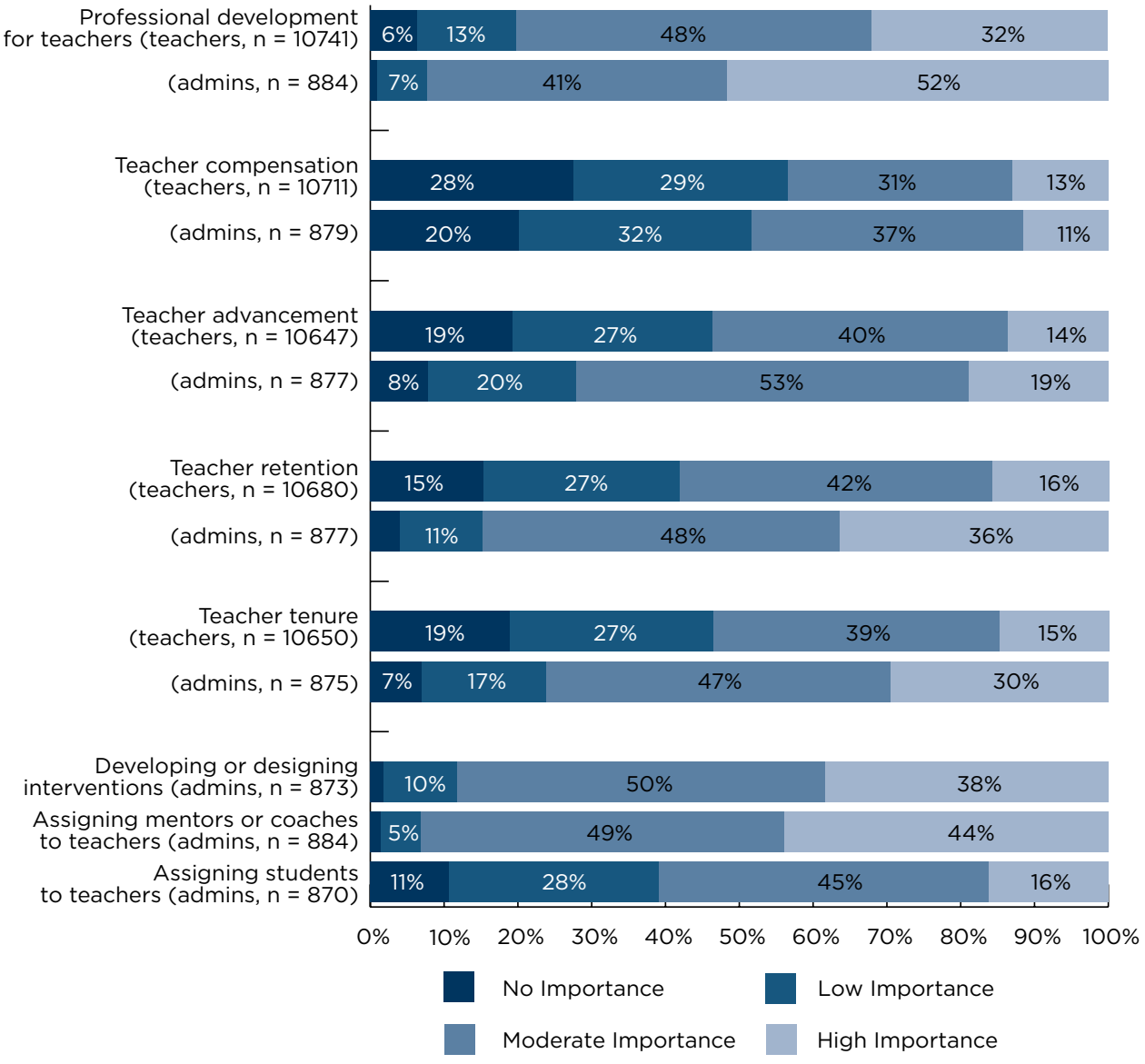
Chart 5.2.2: Teacher and administrator level of understanding and agreement on the individual scoring components



Research Question 5.3: How do teachers and administrators believe results from the TEAM evaluation model should be utilized?

Chart 5.3.1 shows teacher and administrator level of support for the utilization of teacher effectiveness measures in certain policy decisions. Administrators are generally positive about utilizing results to inform teacher tenure, retention and advancement. They are, perhaps surprisingly, less supportive of utilizing these results within teacher compensation decisions. Teacher responses are similar to administrator perceptions, although teachers are generally less supportive of utilizing evaluation results to inform policy decisions.

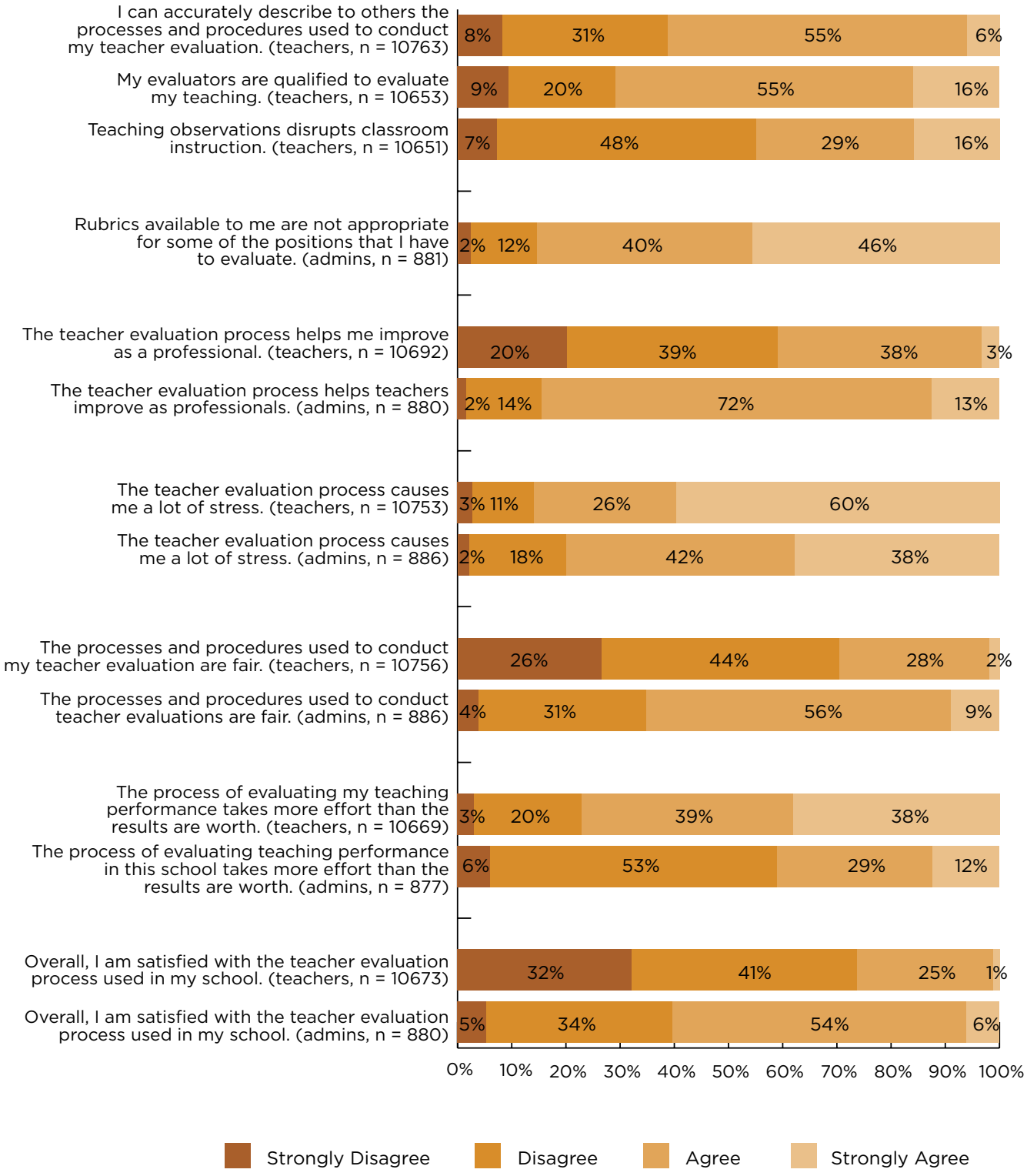
Chart 5.3.1: To what extent do TEAM teachers and administrators think results from the teacher evaluation process should inform selected teacher policy decisions?



Research Question 5.4: To what extent do teachers and administrators understand and support the TEAM evaluation processes and model?

Research question 5.4 investigates a variety of questions pertaining to teacher and administrator experiences with and support for the TEAM evaluation processes, including a summative question that asks for participants' overall satisfaction with the model. Overall, approximately three out of four teachers are dissatisfied with the TEAM evaluation model, while almost two out of three administrators are satisfied with the model. A similar teacher dissatisfaction is found when asking them whether the costs of the process outweigh the benefits of the model. Both teachers and administrators claim that the evaluation process causes them a lot of stress. Finally, almost three out of four teacher respondents strongly disagree or disagree with the statement "Overall, I am satisfied with the teacher evaluation process used in my school", while administrators are more positive, with sixty percent strongly agreeing or agreeing with this statement.

Chart 5.4.1: Teacher and administrator level of understanding and support for the TEAM evaluation process



Methodology

Data Sources

The process of creating a group of educators from which to survey began by extracting a list of all certified staff appearing within the Tennessee Department of Education (TDOE) Educator Information System (EIS), which is the primary state-level current staffing database. These data were extracted in mid-April, 2012, then linked to school-specific grade-levels served from the TDOE Tennessee School Directory. A final linking added variables from a TDOE TEAM database extract, which collects data on all evaluation models utilized within the state.

This pre-randomization file included data on all certified staff within Tennessee, including their gender, ethnicity, years educational experience, highest educational level, email address, evaluation role, and regional service center. It also included a created variable named Tier that recoded all possible grade levels within the state into one of six categories.

A variety of minor data problems were encountered, such as a handful of license numbers that were linked to two different email addresses. These were cleaned utilizing available information. The only significant data issue was the finding that 3,812 of the 74,650 (5.15) certified staff that appear within EIS do not appear within the TEAM database. Because the TEAM database is the source of educator email addresses, these 3,812 individuals could not be surveyed. Due to inaccuracies within the variable Assignment Code, it is not clear, however, that this group of educators should appear in TEAM. A memo describing these findings was presented to the TDOE on May 2nd, 2012.

Sampling Frame

In order to maximize the likelihood that educators responded to the survey invitation, Consortium researchers felt that it was imperative that the survey could be reasonably completed within twenty to twenty-five minutes. In order to reach this goal, educators were invited to take one of six survey versions. Each version of the survey contained the same core questions that investigated aspects of the evaluation model used within the educator's school. Each version then also contained a distinct module that covered one of the following topics: Great Teachers and Leaders, Professional Development, Data Systems & Resources to Support Instruction, Standards and Assessment and Knowledge of and Attitudes Towards Reform, Instructional Practices and Testing, and Teacher Compensation. Through this method, a wide range of survey topics could be measured while also achieving the goal of survey of reasonable length.

Educators were randomly assigned to these six modules through a process of stratified random assignment, utilizing the following strata:

- Size of district, four categories
- Teacher years experience, three categories
- Tier, two categories {ES/MS/K-8; HS}
- Tennessee grand division, three categories

This process was followed with one exception. A small subset of Tennessee schools that are participating within a TIF grant had already been surveyed within 2011-12 on the questions included within the Compensation Module. Educators within these schools, who had already had an opportunity to respond to questions concerning teacher compensation, were randomly assigned to one of the five non-compensation modules.[9]

Survey invitations were sent to all school-level certified staff, including instructional coaches, librarians, and counselors. All educators took the same survey because branching structures within the survey sent educators into distinct question pathways depending on their responses. The first key branch in the survey was based on position, and collected demographic data of administrators separate from the demographic data of non-administrative, certified staff. The second key branch in the survey was based on whether or not the respondent had served as an observer. Those who answered no were sent to questions concerning their experiences as an individual being observed, while those who answered in the affirmative were asked questions concerning their preparation and perception of conducting observations. While most observers are administrators, a handful of observers serve double duty—conducting observations and serving within a position that requires them to be observed. Note that the branching design does not collect data from teacher-observers when they are observed; this was an intentional decision due to length considerations if teacher-observers were expected to provide data on both their observer and observe experiences.

Representativeness

Consortium researchers conducted a check of representativeness by comparing the distribution of observable characteristics of survey-takers to the entire EIS/TEAM universe. These results are shown in Table 3.1, and reveal a response rate of 27.3% for administrators and 24.8% for non-administrators. The table reveals that administrator respondents are more likely to be female, more likely to be white, and more likely to be in medium and very small districts. They are also more likely to be in an elementary or K-8 tier. These differences, although statistically significant, are not substantial for most differences, with an exception of the disproportionality in ethnicity and the size of district.

Non-administrators respondent are more likely to have a masters degree or higher, more likely to be female, more likely to be from small and very small districts, and more likely to have more

9 A verification of this randomization was completed by comparing the distribution of teacher characteristics within each module to the overall sample. No substantial differences were found except for within the compensation module, which was overrepresented by large districts (enrollment > 40,000). This finding was expected, however, due to the pattern of TIF participation by districts.

years of experience. The most important of these differences may be in the variables educational level and years experience, with more experienced, educated teachers more likely to participate. The impact of this disproportionately will be further investigated in later work.

2012 First to the Top Survey
A Check of Representativeness of Administrator and Non-Administrator Responses
Tennessee Certified Educators Versus Survey Respondents

Variable	Administrators* (3,320 in EIS; 905 Respondents)			Non-Administrators** (67,247 in EIS; 16,705 Respondents)		
	% of EIS, Admin	% of Respondents, Admin	Chi-square EIS to Respondents	Chi-square EIS to Respondents	% of Respondents Non-Admin	% of EIS, Non-Admin
Highest Educational Level			2.9 df=4	174.8 df=4		
Bachelor's	1.0%	0.6%	p > .05	p < .001	36.2%	40.4%
Master's	42.6%	42.9%			44.2%	42.1%
Master's Plus	17.2%	16.6%			10.3%	9.1%
Education Specialist	29.8%	30.1%			8.1%	7.5%
Doctorate	9.4%	9.8%			1.2%	0.9%
Sex			13.0 df=1	188.4 df=1		
Female	55.4%	60.5%	p < .001	p < .001	83.4%	79.7%
Male	44.6%	39.5%			16.5%	20.3%
Ethnic Origin			47.7 df=2	82.1 df=2		
White	80.3%	87.8%	p < .001	p < .001	89.4%	87.4%
Black or African-American	19.4%	11.8%			10.1%	12.1%
Other	0.3%	0.3%			0.4%	0.6%
Urbanicity of District			44.6 df=3	189.8 df=3		
Enrollment > 40,000	33.2%	24.5%	p < .001	p < .001	28.6%	32.5%
40,000 > E > 10,000	25.3%	27.1%			26.8%	26.9%
10,000 > E > 5,000	15.1%	16.7%			19.0%	16.9%
5,000 > Enrollment	26.4%	31.7%			25.5%	23.8%
Tier			16.0 df=4	23.02 df=4		
K-5	38.0%	40.9%	p < .01	p< .001	45.1%	44.0%
5 - 8	20.6%	19.1%			17.7%	18.2%
9 - 12	28.7%	24.9%			25.1%	25.8%
K-8	9.9%	11.9%			10.1%	9.7%
Other	2.8%	3.3%			2.0%	2.3%
Years Experience			13.4 df=5	326.4 df=5		
0 to 3 years	0.7%	1.0%	p < .05	p < .001	15.4%	18.8%
4 to 6 years	2.5%	2.8%			12.7%	14.5%
7 to 10 years	10.0%	7.7%			15.1%	15.5%
11 to 17 years	31.3%	29.4%			23.8%	22.1%
18 to 25 years	25.1%	26.6%			16.8%	14.6%
Over 26 years	30.5%	32.4%			16.3%	14.5%

*Administrators were flagged using the *School Administrator* flag maintained within the TDOE TEAM Database. A very small percentage of survey respondents who were flagged as administrators within the TEAM Database but who also self-identified as non-administrators on the survey were dropped from this category.

** Non-administrators includes teachers and all other certified support staff, including instructional coaches, librarians/media specialists, counselors, speech pathologist, and other miscellaneous certified staff.

Limitations

Several limitations should be considered when reviewing these survey findings. First, despite email invitations and multiple reminders, these results represent the perceptions of only one in four Tennessee educators. Surveying efforts were undoubtedly hurt by soliciting participation during the last month of school. One also cannot discount the possibility that negative-response bias is present.

Although differences in the representativeness analyses are statistically significant, these differences are relatively small (all differences less than 5%), and the reported results represent the perceptions and experiences of over 16,000 Tennessee educators.