
Making Every Observation Meaningful

Addressing Lack of Variation in Teacher Evaluation Ratings

OFFICE OF RESEARCH AND POLICY
POLICY BRIEF NOVEMBER 2014





CONTENTS

- 1 Introduction
- 2 Defining Non-Differentiating Observers
- 4 Characteristics of Non-Differentiating Observers
- 6 Identifying Non-Differentiating Observers at Different Time Points
- 8 Overlap between Non-Differentiation and Misalignment
- 10 Policy Implications

[Acknowledgments](#)

Tony Pratt, primary author

Brad Walker, designer

INTRODUCTION

Tennessee launched a new, statewide system of educator evaluation in 2011. In each of the years since, there have been hundreds of thousands of observations and conversations about teaching practice. The teacher observations that represent the primary component of teacher evaluation scores have the potential to provide valuable information about teachers' instructional strengths and areas in need of improvement—information that can be used both by district personnel and by the teachers being evaluated. But the value of observers' ratings of teaching practice depends in large part on the strength of the feedback that the ratings provide.

Currently, the most common way of assessing the quality of observers is to examine the alignment between the ratings teachers are given in observations and value-added ratings that measure teachers' impact on student test scores. This can be a highly useful metric, and it is one that we make use of in Tennessee when we think about the overall accuracy of observation ratings. The alignment metric, however, focuses more on the validity of the rating than the quality of the observer feedback.

This report describes an analysis by the Tennessee Department of Education's Office of Research and Policy that proposes a second quantitative metric for assessing the practicality and usefulness of the ratings that teacher evaluators are giving to the teachers they observe by looking at the range of ratings given. We introduce the idea of the “non-differentiating observer,” or the observer whose ratings do not usefully distinguish between teachers' relative strengths and weaknesses. This metric highlights the need for evaluations to be useful for teachers as a tool for their own improvement, and it allows us to assess evaluator performance in an ongoing way during the year.

In the following pages, we identify the number of non-differentiating observers, where they are located, and whether any particular characteristics predict whether an observer will be non-differentiating. While non-differentiating observers represent only a small proportion of the overall observer pool, we are able to identify a group that is clearly distinguishable within the data, and we propose several possible supports and interventions that might reduce the problem. It is our hope that this paper will both contribute to the process of continuous improvement to the teacher evaluation system in Tennessee and help other states, districts, and schools think critically about their own processes for assessing observer quality.

Key Findings

- Observers whose ratings do not provide teachers with a range of feedback on strengths and weaknesses are failing to provide the kind of usable information that might lead to instructional improvement.
- In Tennessee, “non-differentiating observers” make up a small but meaningful proportion of the total evaluator pool.
- Non-differentiating observers are not clustered in particular areas but are scattered throughout the state and across all types of observers.
- The real-time observation data collected by the Tennessee evaluation system can allow districts and the state to identify non-differentiating observers during the year and to take steps to ensure that teachers are receiving meaningful feedback.
- Non-differentiation is one of several indicators of low observation quality, and non-differentiating observers are not necessarily the same observers who fail to achieve reasonable alignment between observation and value-added ratings.

DEFINING NON-DIFFERENTIATING OBSERVERS

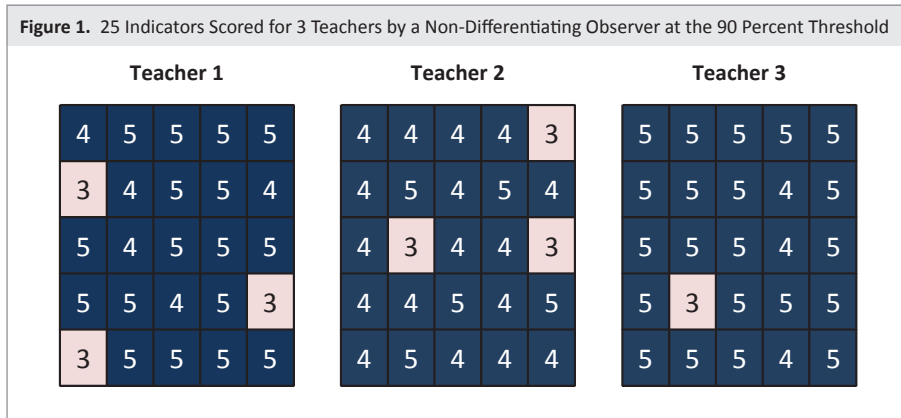
In Tennessee, “non-differentiating observers” make up a small but meaningful proportion of the overall evaluator pool.

We define a non-differentiating observer as a teacher evaluator whose ratings are nearly all identical, both across teachers and across rubric indicators during a single observation.¹ By definition, these observers likely fail to provide teachers in their schools with meaningful and usable feedback about their relative instructional strengths and weaknesses. A lack of differentiation across indicators on the rubric is in direct conflict with one of the primary objectives of the teacher evaluation system—to provide educators with useful, actionable feedback that allows them to improve their professional practice.

While the concept of a non-differentiating observer is relatively straightforward, the exact measure depends upon how we set the boundary for non-differentiation, and the following analysis allows for two potential definitions of non-differentiating observers. Each of these definitions identifies a meaningful number of observers that could potentially be targeted for appropriate supports and interventions.

90 Percent Threshold

Under this model, the non-differentiating observer must score over 90 percent of the indicators within two adjacent levels on the 1–5 observation scale. In other words, the observer must have scored at least 9 of every 10 TEAM rubric indicators in just two levels. Figure 1 illustrates what 25 indicators scored by such an observer might look like for three teachers. By giving primarily scores of level 4 and 5, the observer has not offered substantial feedback along the observation scale.



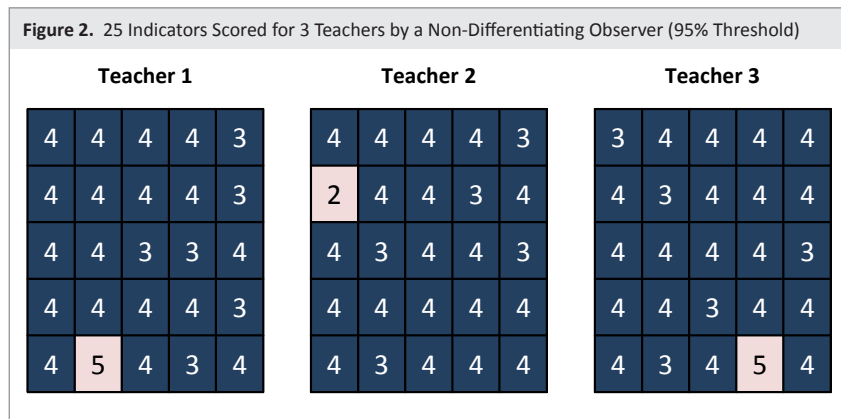
1. 118 Tennessee districts are included in this and all other analyses presented in this paper; some of the state’s districts are excluded because they use alternate observation models and/or use alternate data systems to collect observation data and did not report the observer for each observation. Only observers who scored at least 100 indicators during the 2012-13 school year were included in all analyses.

What Is an Indicator?

The majority of Tennessee districts use the [Tennessee Educator Acceleration Model \(TEAM\)](#) rubric to conduct classroom observations as one component of the teacher evaluation system. Teachers in these districts are scored on a 1-5 scale on nineteen different indicators on the rubric over a series of observations. Each indicator is intended to measure a specific skill or competency. For example, when a teacher is observed using the Instruction domain of the TEAM rubric they are scored on 12 different indicators such as “Lesson Structure and Pacing,” “Questioning,” and “Academic Feedback.”

95 Percent Threshold

Under this model, the non-differentiating observer must score over 95 percent of the indicators within two adjacent levels on the 1–5 observation scale. In other words, the observer must have scored at least 19 of every 20 TEAM rubric indicators in just two levels. Figure 2 shows what 25 indicators scored by such an observer might look like for three teachers. The figure illustrates that these observers provide an even more extreme example of a failure to differentiate within and across teachers than those at the 90 percent threshold; in this case the observer rarely gives scores outside of levels 3 and 4.²



Based on the above definitions, Table 1 illustrates the number of observers that fell into each category during the 2012–13 school year. One important takeaway is that while a number of observers fall in these categories, most observers do not; this indicates that most observers are providing a reasonable amount of differentiation in their scores across indicators. That said, non-differentiating observers evaluated 7,164 different teachers over the course of the year and targeted interventions to improve the quality of these observers’ ratings could have a large impact on the quality of feedback received by these teachers.

Table 1. Number and Percent of Non-Differentiating Observers (2012-13)

	Observers with 90+% of indicators in 2 levels	Observers with 95+% of indicators in 2 levels	Observers that did not fall in either category
Number of Observers	306	96	2,483
Percent of Observers	11.0%	3.4%	89.0%
Total Number of Teachers Scored by these Observers	7,164	1,867	65,886

2. While other cut points or methodologies for identifying non-differentiating observers could certainly be defensible, we chose the 90 percent and 95 percent thresholds because they represent a clear non-differentiation problem while also identifying what seems like a manageable numbers of observers. Both categories have a sufficient number of observers to be considered problematic but also do not have so many observers that providing meaningful supports or interventions would be impractical because of scale.

CHARACTERISTICS OF NON-DIFFERENTIATING OBSERVERS

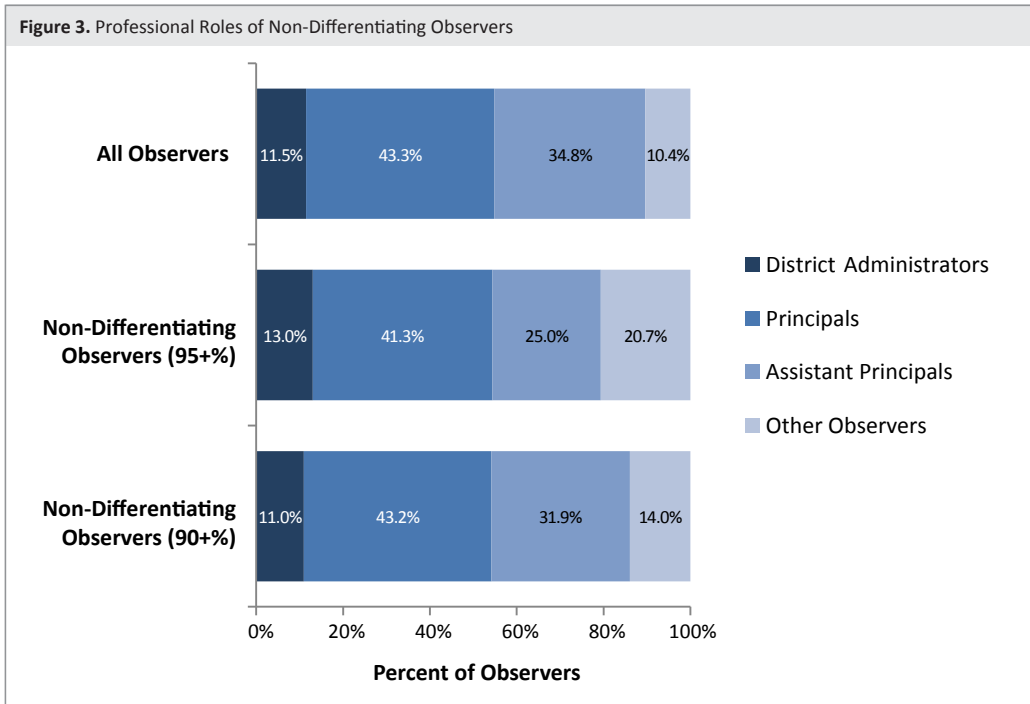
Non-differentiating observers are not clustered in particular geographic areas but are scattered throughout the state and across all types of observers.

One potential explanation for the existence of non-differentiating observers is that certain districts or groups of districts enacted policies or encouraged practices that made the issue more likely to arise; alternatively, this could be an issue that occurs across many different districts without many systematic patterns. Our analysis supports the latter explanation. While some districts were slightly more likely than others to have non-differentiating observers, this does not seem to be an issue that is driven primarily by district policy or geography. We find that few districts seem to be overpopulated by non-differentiating observers and most have a small share of such observers.

- Only two districts had more than 10 observers who scored at least 95 percent of indicators in two levels; these observers did not represent a substantial percentage of the observers in those districts.
- 48 of the 118 districts included in this analysis had at least one non-differentiating observer at the 95 percent threshold.
- Only five districts with at least ten eligible observers had at least 10 percent of their observers fall in this category.
- No district had more than 25 observers who scored at least 90 percent of their indicators in two levels while 88 of 118 districts had at least one such observer.
- Non-differentiating observers at both thresholds did not tend to be concentrated geographically in any particular areas of the state or in large or small districts.

Even if non-differentiating observers are not clustered in certain geographical areas, they could potentially share common characteristics that would allow us to understand more about why such behaviors occur. Yet we also find little evidence that non-differentiation is driven by observer characteristics.

Observers in the state are grouped into one of the following four professional roles: district-level administrator, principal, assistant principal, or other (typically instructional coaches or lead teachers). Figure 3 shows how the distribution of these roles differs across the full group of observers and across those that we have labeled non-differentiating. Interestingly, we see some significant differences. The “other” observers—usually instructional coaches or lead teachers—were slightly more likely to be non-differentiating at the 90 percent threshold. These observers were even more overrepresented at the 95 percent threshold, accounting for 20.7 percent of non-differentiating observers at this level while only accounting for 10.4 percent of all observers. This information offers some insight into the make-up of the non-differentiating observer group. On average, these observers are less likely to be school or district leaders, suggesting that we might want to offer some extra support for this category of observer. At the same time, the differences are not large enough to indicate that the state’s issue with non-differentiation lies solely with one observer group. Non-differentiating observers are scattered across each observer category.



The same finding holds true when we look at other observer characteristics. Despite some minor differences, observers’ years of experience and highest degree level attained are also not strong predictors of whether or not they will be non-differentiating observers. Table 2 below summarizes these characteristics for observers that fell in each of non-differentiating observer categories and those that did not. While there are some minor differences across groups, the differences are small overall.

Table 2. Characteristics of Non-Differentiating Observers (2012-13)

	Observers with 90+% of indicators in 2 levels	Observers with 95+% of indicators in 2 levels	Observers that did not fall in either category
Bachelor’s Degree only (%)	3.7%	3.2%	2.3%
Master’s+ (%)	96.3%	96.8%	97.7%
Doctorate (%)	9.9%	11.8%	10.8%
Years of Experience in Education in Tennessee (Average)	22.0	23.4	20.0

IDENTIFYING NON-DIFFERENTIATING OBSERVERS AT DIFFERENT TIME POINTS

The real-time observation data collected by the Tennessee evaluation system can allow districts and the state to identify non-differentiating observers during the year and to take steps to ensure that teachers are receiving meaningful, differentiated feedback.

In order to meaningfully intervene with non-differentiating observers, we need to know when these observers can be reliably identified.

In Figure 4, we show the overlap between the group identified as non-differentiating if we use all the data across 2012–13 and the group identified if we use only data collected through December 2012. This figure only includes observers who had scored at least 100 indicators through December 2012. The results show that most non-differentiating observers can be reliably identified near the midpoint of the school year. The overwhelming majority of observers who could have been identified as non-differentiating at the end of December were non-differentiating observers over the course of the full school year. Over 80 percent of the non-differentiating observers identified at the 95 percent threshold in December were non-differentiating at either the 90 or 95 percent threshold over the course of the year. This suggests that midyear results could potentially be a viable option for identifying non-differentiating observers for possible intervention.

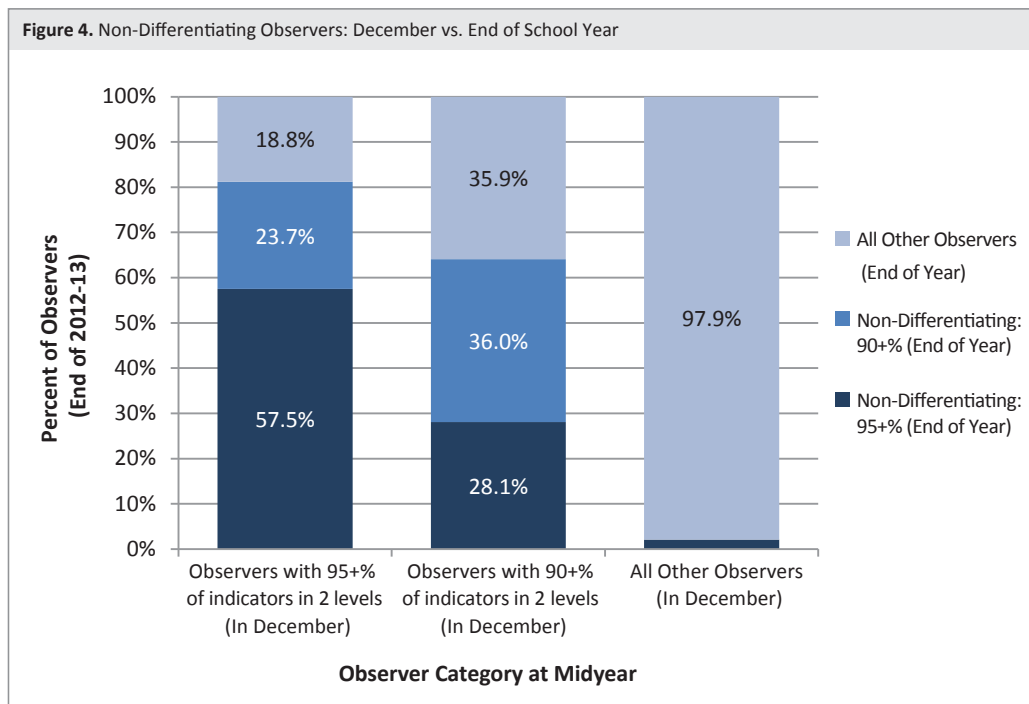
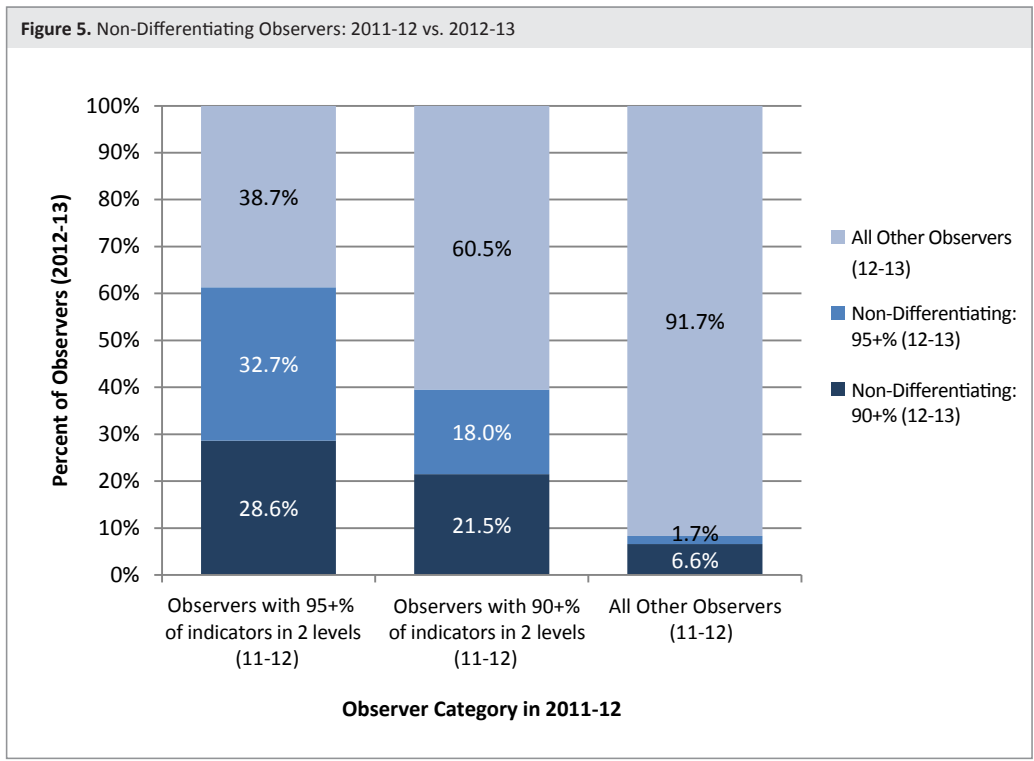


Figure 5 reaches back even farther in time and uses data from the previous year’s observation ratings. We find that about 40 percent of observers exceeding the 90 percent threshold in 2011–12 exceeded the 90 percent threshold in 2012–13; over 60 percent of those exceeding the 95 percent threshold in 2011–12 exceeded the 90 percent threshold in 2012–13. At the same time, just 8.3 percent of observers who were not non-differentiating in 2011–12 were identified as non-differentiating in 2012–13 while just 1.7 percent of these observers exceeded the 95 percent threshold. One potential explanation for why an observer may have been identified in 2011–12 but not 2012–13 is that observers could have become more comfortable with giving a wider variety of indicator scores after receiving additional training or after becoming more comfortable with the rubric through experience. Overall, this analysis suggests that, although some observers identified as non-differentiating in 2011–12 were not identified in 2012–13, using the 95 percent threshold to identify observers from the previous year for potential intervention is a viable option.



OVERLAP BETWEEN NON-DIFFERENTIATION AND MISALIGNMENT

Non-differentiating observers are not necessarily the same observers who fail to achieve reasonable alignment between observation and value-added ratings.

As noted earlier, one method of assessing the quality of an observer's observation ratings is by measuring the alignment between the observation and the value-added scores that rated teachers receive. In this section, we investigate whether this measure captures the same group of evaluators who are flagged as non-differentiating observers.

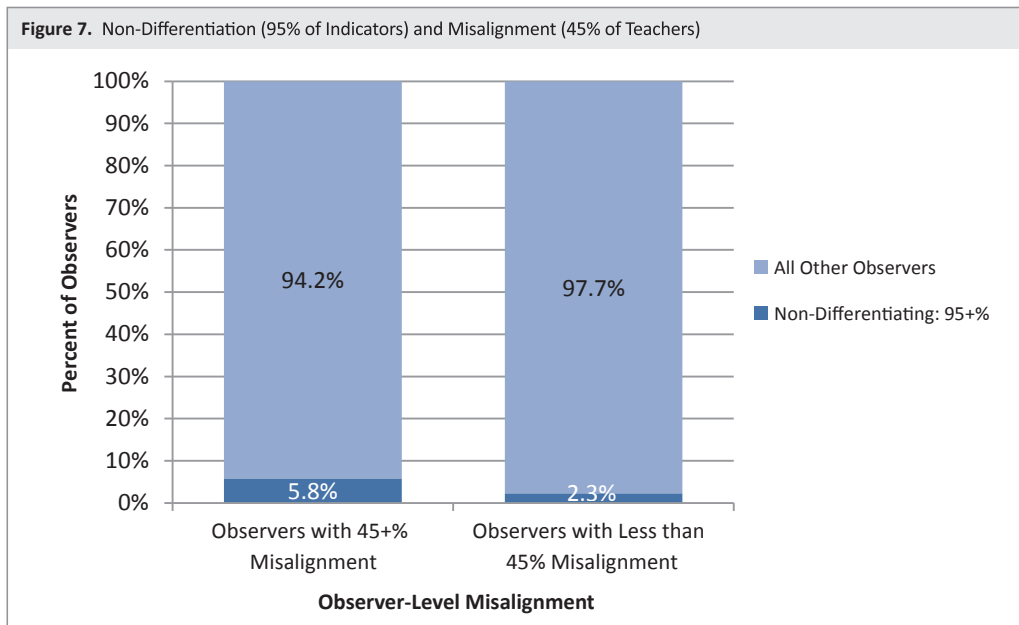
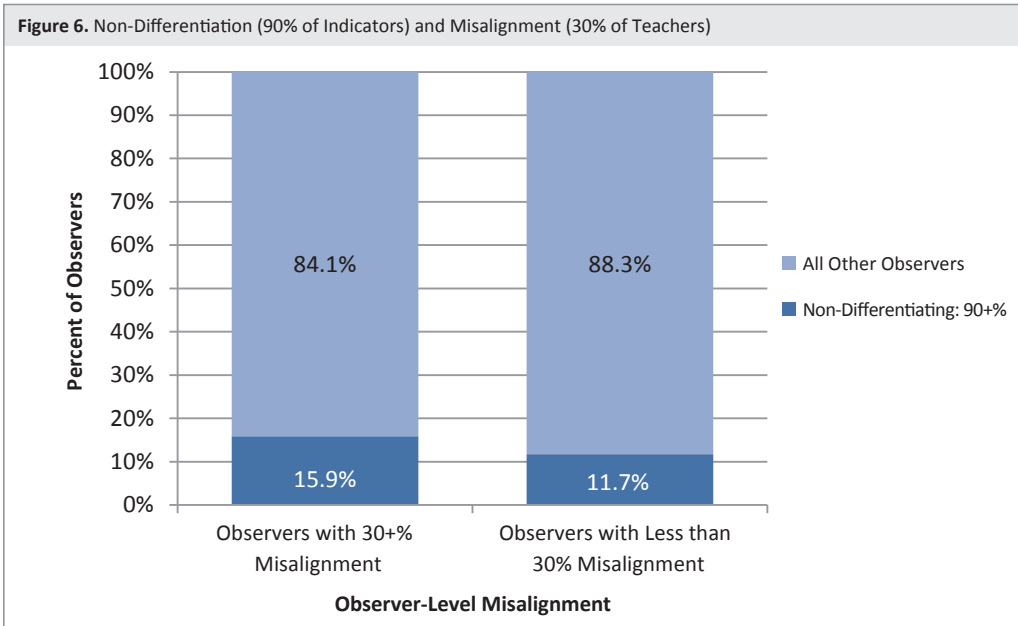
In order to assess whether non-differentiating observers also tend to produce results misaligned with growth scores, we first calculated teacher-level observation averages for each observer. We scaled these averages into the five levels used in the TEAM model and compared them to individual Tennessee Value Added Assessment System (TVAAS) composite score levels, which are also reported as one of five levels. While we would not expect perfect alignment between the two measures, scores misaligned by three or more levels indicate a fundamental disagreement in whether a teacher's performance is above or below expectations; this is the threshold that is currently used in [State Board policy](#) to define misalignment.

We considered observers who had at least five teachers who had individual TVAAS scores in this analysis. Next, we examined whether each teacher's observation average from a given observer was three or more levels removed from their individual TVAAS score on the 1–5 scale. We then calculated the percent of misaligned teachers for each observer and identified thresholds for misalignment that roughly corresponded with the percentage of observers identified by each of the non-differentiating observer thresholds in order to foster meaningful comparisons. These thresholds were 30 percent misaligned by three or more levels³ and 45 percent misaligned by three or more levels.⁴

Figures 6 and 7 display the proportion of observers that fell in either or both of the misalignment and non-differentiation categories. Although observers with high levels of misalignment were slightly more likely to be non-differentiating observers, the two methods tend to identify different sets of observers. As a result, identifying observers in need of support using the non-differentiation approach presented here has the potential to be complementary to the current misalignment-only approach. Only four of the 69 observers who had over 45 percent misalignment also scored over 95 percent of their indicators in two levels. Similarly, just 30 of 263 observers with over 30 percent misalignment also scored at least 90 percent of their indicators in two levels.

3. 12.1 percent of observers fell in this category, roughly equivalent with the percentage of observers above the 90 percent threshold for non-differentiation.

4. 3.2 percent of observers fell in this category, roughly equivalent with the percentage of observers above the 95 percent threshold for non-differentiation.



POLICY IMPLICATIONS

Non-differentiating observers are not concentrated geographically or in particular districts, they are not linked by a unified set of characteristics, they tend to be a different group of observers than those with the greatest misalignment between teachers' observation scores and growth scores, and they can be identified reliably using either prior year or midyear data. As a result, this issue appears to warrant a targeted intervention aimed at reducing the chance that a teacher will have a non-differentiating observer. The Tennessee Department of Education has taken initial steps to address and further understand the issue by deploying TEAM coaches to provide one-on-one support to non-differentiating observers. We also intend to monitor non-differentiating observers over time in order to determine any emerging trends in the number of observers identified and their characteristics. While there are undoubtedly many additional ways to address non-differentiation, the following general strategies could be used by schools, districts, and other states:

Reporting

Reports outlining non-differentiation could be provided directly to non-differentiating observers, those tasked with supporting observers, and/or to district leaders. These reports could identify non-differentiating observers and additional information such as how frequently they scored indicators in each level from 1–5. This low-cost strategy has the potential to raise overall awareness of the need to address non-differentiation and could directly or indirectly cause observers to better differentiate their observation scores.

Coaching

Coaches—such as TDOE's TEAM coaches or district-level staff tasked with implementing teacher evaluation—could provide direct, one-on-one coaching to non-differentiating observers. This coaching could include activities such as co-observations, scoring of anchored video lessons, and discussions of how to appropriately score particular indicators on the rubric.

Additional Training

Observers identified as non-differentiating could be provided with additional training. This training could give non-differentiating observers practice in appropriately using the observation rubric and could help them to differentiate their scores in order to give more useful feedback to educators. The training could be designed specifically for groups of non-differentiating observers or, alternatively, non-differentiating observers could be invited or required to attend existing training for new observers.

Changes in Roles and Responsibilities

Non-differentiating observers who continually fail to differentiate their scores across indicators over time could be pulled from their role as observers. This should only be considered after additional strategies have been pursued unsuccessfully and could involve a change in the non-differentiating observer's role within their school or district and/or a loss of certification to conduct observations.