Mishan Jensen, Ph.D.
Karen Looby, Ph.D

Publication 16.49
October 2017

Professional Pathways for Teachers (PPfT)

Appraisal Summary

2016— 2017 School Year

Overview and Framework of PPfT Appraisal System

Professional Pathways for Teachers (PPfT) was designed as a human capital system that blends appraisal, leadership pathways, professional development activities, and compensation. PPfT appraisal system was developed and refined during the 2014–2015 and 2015–2016 school years and implemented district wide during the 2016–2017 school year. The purpose of this summary is to provide a description of baseline appraisals to inform all subsequent years.

Teachers were appraised on three components: instructional practice, professional growth and responsibilities, and student growth. The instructional practice measure was composed of two announced observations and two classroom visits, one in fall and one in spring, each conducted by a different appraiser. These observations were scored using a rubric that covered seven strands of observable teaching behaviors. The professional growth and responsibilities measure was completed toward the end of the school year and was scored using a rubric that covered five strands related to professionalism. The student growth measure was composed of two components, a student learning objective (SLO) and school-wide value-added score. The teacher created the SLO at the beginning of the year as a goal for students' learning that school year. The school-wide value-added score, as calculated by SAS EVAAS, designed to measure the extent to which a schools' average growth exceeded, met, or fell short of the average expected growth.

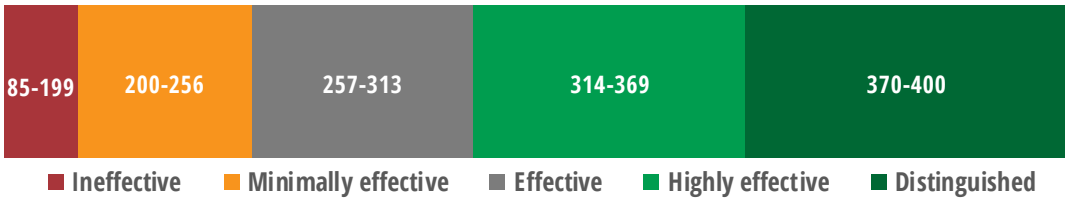
The appraisal components, based on the plan the teacher followed, were combined to produce a final summative appraisal score. Teachers were assigned to one of three appraisal plans based on their hiring status or assignment: standard, new teacher, or late contract. These plans differed by the components included, with the standard and new teacher plans including all three appraisal components, and the late contract plan excluding the student growth component. Every teacher was ranked in terms of his or her effectiveness, ranging from ineffective to distinguished, based on the teacher's final summative appraisal score.

The range of final summative appraisal scores was classified into five categories: ineffective, minimally effective, effective, highly effective, and distinguished (Figure 1). For more information on PPfT, see the *Professional Pathways for Teachers Support Guide* at www.austinisd.org/ppft/appraisal.

Department of Research and Evaluation
Austin Independent School District

Figure 1

PPfT Final Summative Appraisal Score Classification Categories



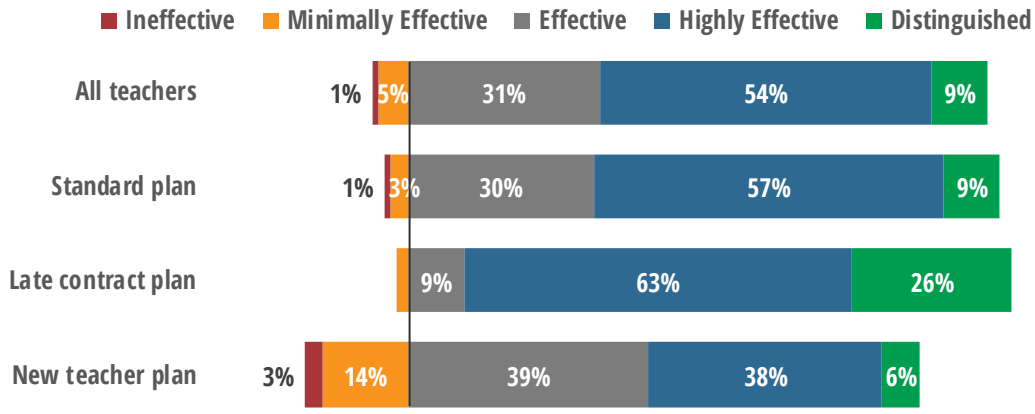
Source. Professional Pathways for Teachers Support Guide

How were teachers' appraisal score distributed?

Overall, the majority (84.86%, n = 5,629) of teachers in the district were rated as effective or highly effective during the 2016–2017 school year. The average summative appraisal score was 322.36 (n = 5,629, SD = 39.24). The average appraisal score varied according to the plan the teacher was on, with more variability seen in the appraisal scores for new teachers (M = 302.00; n = 936, SD = 47.71). Conversely, the late contract teachers had less variability and higher average appraisal scores (M = 347.43; n = 142, SD = 32.01). Additionally, no late contract teachers were rated as ineffective, whereas 3% of new teachers were rated as Ineffective (Figure 2).

Figure 2

More late contract teachers were rated as distinguished and more new teachers were rated as minimally effective and ineffective.



Source. Human Capital Platform data collected by the Office of Educator Quality

Did teachers' appraisal scores vary by school level?

Due to the different populations of students teachers taught, some teachers' appraisal scores were isolated from the rest. The majority of teachers at all school levels were rated as effective or highly effective; however, the distribution of rating differed slightly by school level. More middle school teachers than teachers at other school

Evaluation of Inter-Rater Reliability

In the spring of 2016, AISD partnered with Regional Education Laboratory Southwest (REL SW) to conduct an analysis of the observation ratings for teachers using the PPfT appraisal system. This analysis included evaluation of the inter-rater reliability and validity of the observation scores.

REL SW was provided deidentified observation and rater data. The data were evaluated using an item response theory (IRT) model appropriate for rating data, the Rasch Rating Scale model (Andrich, 1978; Wright & Masters, 1982). The reliability was assessed in terms of reliability statistics, differential item functioning statistics, and rater severity measures.

Differential item functioning assesses the generalizability of measures across contexts, such as grade level, school type, and students served.

Reliability assesses the degree to which a person's scores are free from measurement error.

Rater severity assesses the extent to which the teachers' observation scores were differentially influenced by different raters.

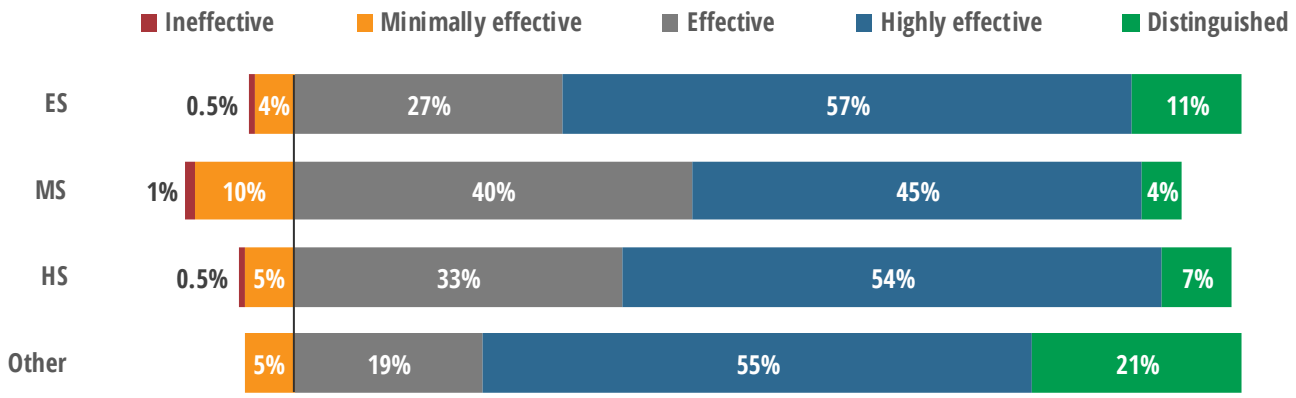
After REl SW analyzed the data provided, the DRE team was apprised of the results of the analysis. The results indicated that the inter-rater reliability was high, differential item functioning was nonexistent, and rater severity was low. The teachers' observation scores did not depend on the rater assigned, nor did items used to rate the teacher function differentially for subpopulations.

levels were rated as effective, minimally effective, and ineffective. Elementary teachers had the highest average appraisal scores, with a mean of 327.26 (n = 3,063, SD = 37.57). High school teachers had similar average appraisal scores (M = 321.13, n = 1,318, SD = 37.54). Middle school teachers' average appraisal scores were lower, with more variability (M = 307.95, n = 1,083, SD = 41.79).

The 2016–2017 school year included all teachers, including teachers at special high schools (e.g., International High School, Lanier Graduation Path, and Travis Graduation Path). Teachers at other schools (e.g., Rosedale, Alternative Learning Center, and Leadership Academy) and non-school sites (e.g., Austin State Hospital and Dell Children's Medical Center [DCMC] Education Center) were included. Additionally, teachers assigned to homebound students and special courses were included. The average appraisal score for these teachers was 335.56 (n = 165, SD = 40.55). Teachers at these special campuses had appraisal ratings that stood out from the rest, with a larger portion rated as distinguished, and none rated as ineffective (Figure 3).

Figure 3

More middle school teachers were rated as minimally effective, while more teachers at other locations (e.g., special high schools and non-school sites) were rated as distinguished.



Source. Human Capital Platform data collected by the Office of Educator Quality

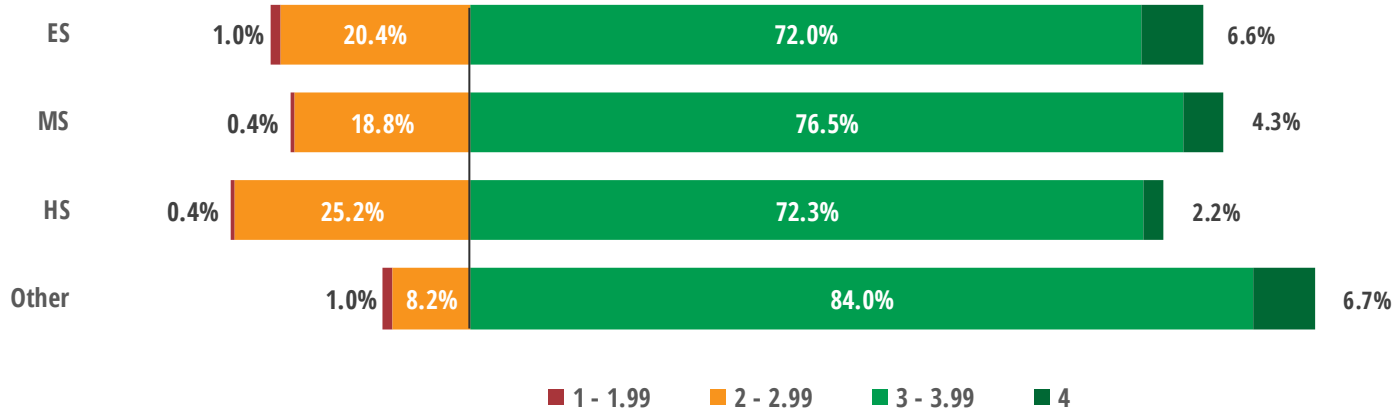
Note. Other includes teachers assigned to special high schools (e.g., Lanier Graduation Path, International High School, Leadership Academy), other schools (e.g., Clifton Center, Phoenix House Alternative Learning Center), non-school sites (e.g., Austin State Hospital and DCMC Education Center), and other (e.g., Homebound Program, Learning Support Services, Performing Arts).

Did teachers' instructional practice scores vary by school level?

Teachers were observed once in the fall and spring, each conducted by a different appraiser. The observations were scored based on an appraisal rubric that covered seven strands of observable teaching behaviors. These observation scores ranged from one to four, with a score of three being the expected standard. The seven strands were averaged for both the fall and spring observations. The two instructional practice ratings were averaged and then multiplied by 0.50 in calculating the final summative appraisal score. For more information on PPfT instructional practice appraisal ratings, see the *Professional Pathways for Teachers Support Guide* at www.austinisd.org. Teachers' instructional practice scores varied by school level, with more teachers at special high schools, other schools, and non-school sites scored in the 3— 3.99 and 4 categories in the fall observation (Figure 4).

Figure 4

During the fall observations, teachers assigned to other types of campuses were rated higher, on average, for instructional practice scores than teachers at elementary, middle, and high schools.



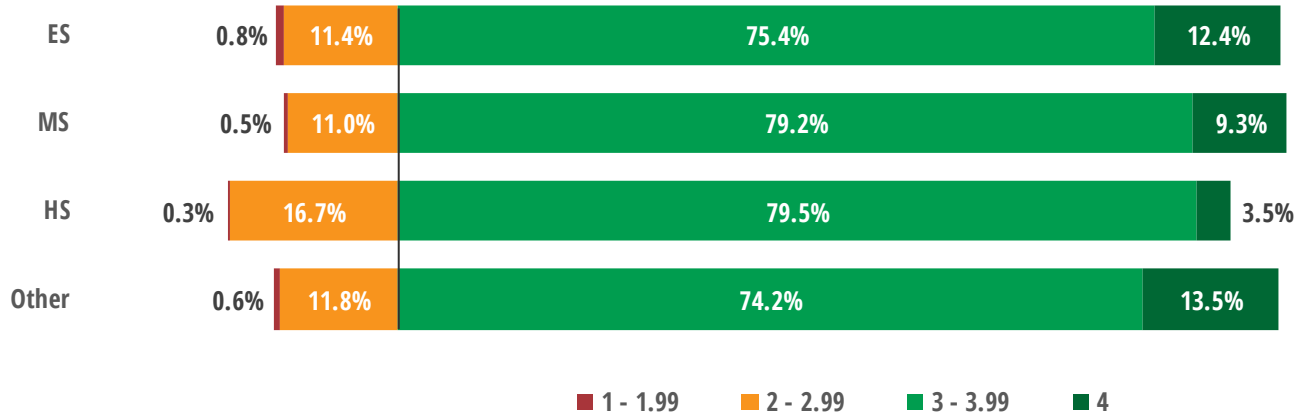
Source. Human Capital Platform data collected by the Office of Educator Quality

Note. Other includes teachers assigned to special high schools (e.g., Lanier Graduation Path, International High School, Leadership Academy), other schools (e.g., Clifton Center, Phoenix House Alternative Learning Center), non-school sites (e.g., Austin State Hospital and DCMC Education Center), and other (e.g., Homebound Program, Learning Support Services, Performing Arts).

The spring observation rating distributions were slightly different, with more teachers across all school levels improving their average instructional practice ratings. The proportion of middle school teachers' whose average score that fell into the 3— 3.99 range increased, whereas the teachers assigned to other types of campuses average rating score in that same range decreased. The proportion of elementary and middle school teachers' whose average rating was a 4, more than doubled compared to the fall observation ratings (Figure 5).

Figure 5

During the spring observations, more teachers were rated above 3 for all school levels.



Source. Human Capital Platform data collected by the Office of Educator Quality

Note. Other includes teachers assigned to special high schools (e.g., Lanier Graduation Path, International High School, Leadership Academy), other schools (e.g., Clifton Center, Phoenix House Alternative Learning Center), non-school sites (e.g., Austin State Hospital and DCMC Education Center), and other (e.g., Homebound Program, Learning Support Services, Performing Arts).

What were teachers' perceptions of the PPfT appraisal system?

The majority of the teachers surveyed in the 2016–2017 Employee Coordinated Survey did not perceive the implementation of the appraisal system to be challenging (Figure 6). Teachers indicated that writing an SLO as more challenging than doing the other components of PPfT, with scheduling post-observation conferences the least challenging.

Figure 6

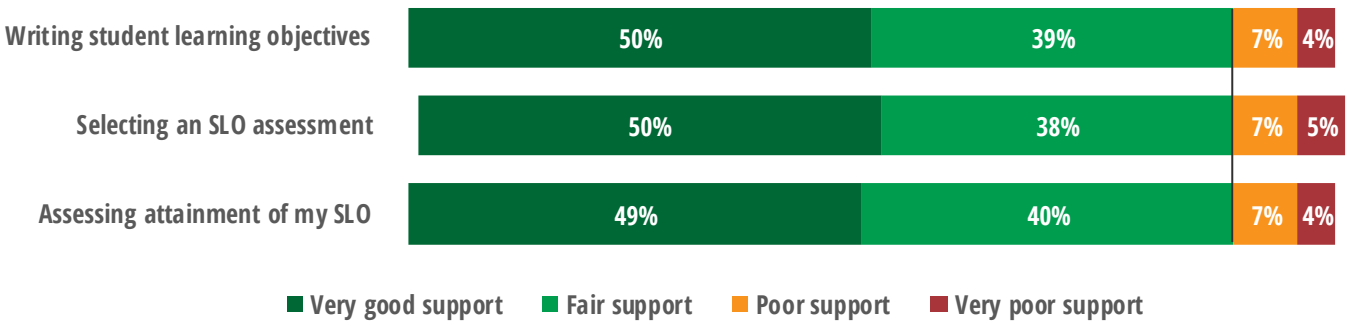
How challenging did you find the following features of the Professional Pathways for Teachers (PPfT) appraisal system this year?

Source. AISD Employee Coordinated Survey, 2016— 2017

The majority of teachers (almost 90%) surveyed indicated they received very good to fair support writing SLOs, selecting an SLO assessment, and assessing attainment of their SLO (Figure 7). The perception of the PPfT appraisal system in terms of distinguishing effective teachers from ineffective teachers was less decisive (Figure 8). Slightly more teachers indicated they did not find the system good at making this distinction than indicated that the system did do a good job. Although the teachers' perceptions about distinguishing effective teachers from ineffective teachers was more divided than were their perceptions about other aspects of the PPfT system, the number of teachers responding to this is item was much lower than that for other items.

Figure 7

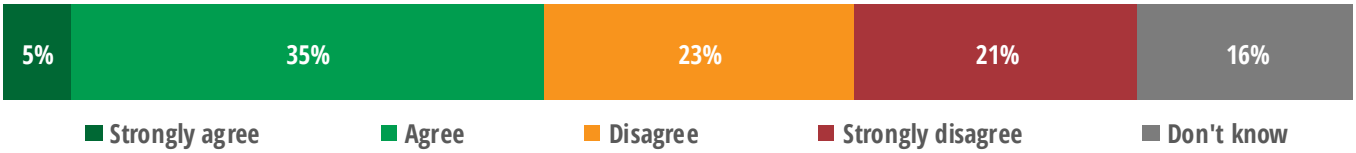
Indicate the quality of support you received with the following features of the PPfT appraisal system.



Source. AISD Employee Coordinated Survey, 2016— 2017

Figure 8

The PPfT appraisal system does a good job distinguishing effective from ineffective teachers.



Source. AISD Employee Coordinated Survey, 2016— 2017

Conclusions and recommendations

The PPfT appraisal system was developed in the 2014–2015 school year and refined during the 2015–2016 school year. The 2016–2017 school year was the district-wide roll out of the PPfT appraisal system. The majority of teachers were rated as effective (31%), highly effective (54%), and distinguished (9%). Significant differences were observed in the distributions of appraisal scores by appraisal plans, e.g., standard, new teacher, and late contract. Additionally, a partnership between the Regional Education Laboratory Southwest and AISD studied the inter-rater reliability of the appraisal scores and ratings were found to be reliable. However, the study utilized data from the pilot year and contained a subset of teachers in the district. Because there were significant differences found in the distribution of appraisal ratings by appraisal plan, it is recommended that re-evaluation of the inter-rater reliability be completed with the district-wide data.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561–573.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement.* Chicago, IL: MESA Press.

Austin Independent School District

Mishan Jensen, Ph.D.
Karen Looby, Ph.D.

Department of Research and Evaluation

1111 West 6th Street, Suite D-350 | Austin, TX 78703-5338
512.414.1724 | fax: 512.414.1707
www.austinisd.org/dre | Twitter: @AISD_DRE

October 2017

Publication 16.49