

# TELPAS Writing Audit Report

Grades 2–12  
Spring/Summer 2016

# Introduction

## ***Overview***

The purpose of the Texas English Language Proficiency Assessment System (TELPAS) writing audit is to provide ongoing evidence of the validity and reliability of the holistically rated writing component of the TELPAS assessment. Under this assessment system, teachers are trained to use the TELPAS Proficiency Level Descriptors (PLDs) to rate the English language proficiency of students in specified language domains, including writing. The primary audit activities included the following:

1. evaluating the extent to which the writing English language proficiency ratings assigned by teacher raters match those assigned by state audit raters,
2. gathering feedback from the teacher raters on the quality of their training for writing, and
3. examining how well educators followed state-defined administration procedures to rate writing performance.

Six writing audits (including the 2016 audit) have occurred to date. The first audit, conducted in 2005, was relatively small, allowing the state to examine and improve audit procedures. The second audit, conducted in 2006, was a larger study in which information from a large, representative number of districts and students was collected. That audit provided results for regional education service centers (ESCs) and large districts referred to as *training entities* because of their role in directly providing TELPAS training to teacher raters. The 2006 audit results provided reliability and validity evidence supporting the accuracy of teacher ratings on the writing domain. The 2007, 2008, 2013, and 2016 TELPAS audits were smaller and served to provide ongoing evidence of validity and reliability at the level of the state rather than at the training entity level. The comparison of rater agreement across the five audits indicated that rater accuracy improved from 2005 to 2006, and was stable across 2006, 2007, 2008, and 2013. In 2016, there was a slight increase in rater accuracy from 2013. This report includes the sampling plan, the audit results, and questionnaire results for the 2016 writing audit.

## ***Sampling Strategy***

Students were sampled in grades 2–12 from across the state, including all 20 regions, for the 2016 writing audit. To minimize testing burden on Texas schools, the following types of schools were excluded from the sample:

- Campuses that had been rated as Improvement Required (IR) according to the state accountability system for three or more years
- Alternative education campuses (including JJAEP and DAEP)
- Campuses selected for the 2013 TELPAS Writing Audit sample

The target sample size for the 2016 writing audit was 2,000 students, spread equally across grade bands and proficiency levels. This target sample size allowed for a representative sample of student writing collections while minimizing the time required for completing the audit task. Given that some schools might not be able to provide writing samples for all of the students selected for the target sample, an extra 10% was added to the target sample to ensure that at least 2,000 students were included. As shown in Table 1, 138 students were sampled from each of the four proficiency levels (beginning, intermediate, advanced, and advanced high) assigned by the teacher rater and from each of the following grade bands: 2, 3–5, 6–8, and 9–12. The target sample included 2,208 students in total.

**Table 1. Total Number of Students in Target Sample**

<b>Grade Bands</b>	<b>Beginning</b>	<b>Intermediate</b>	<b>Advanced</b>	<b>Advanced High</b>	<b>Total</b>
<b>2</b>	138	138	138	138	552
<b>3–5</b>	138	138	138	138	552
<b>6–8</b>	138	138	138	138	552
<b>9–12</b>	138	138	138	138	552
<b>Total</b>	<b>552</b>	<b>552</b>	<b>552</b>	<b>552</b>	<b>2,208</b>

***Sample Characteristics***

The 2016 target sample and tested population is compared in Table 2. The 2016 target sample was representative of the population of students that took TELPAS in grades 2-12 in 2016 in terms of gender and ethnic representation, as well as representation of the percent of students in grades 2-12. The sample was somewhat less representative of the population in terms of the percent of students in English as a Second Language (ESL) and bilingual programs, due in part to the slightly lower percentage of elementary students in the sample compared to the population, and the slightly higher percentage of middle school and high school students in the sample compared to the population. In Texas, bilingual programs are more common in elementary school whereas ESL programs are more common in middle and high school.

**Table 2. Comparison of the 2016 Population and Audit Target Sample**

<b>Student Characteristics</b>	<b>2016 Population</b>	<b>2016 Audit Sample</b>
Number of 2–12 Students	667,290	2,208
% Male	53%	58%
% Hispanic	90%	85%
% Grade 2	17%	25%
% Grade 3	16%	10%
% Grade 4	14%	9%
% Grade 5	12%	6%
% Grade 6	10%	11%
% Grade 7	8%	8%
% Grade 8	7%	7%
% Grade 9	6%	11%
% Grade 10	4%	7%
% Grade 11	3%	4%
% Grade 12	2%	3%
% ESL	48%	71%
% Bilingual	42%	23%

***Rescoring of Writing Collections***

The TELPAS writing assessment for grades 2–12 required teacher raters to assemble a collection of each English language learner’s writing from classroom instruction in a variety of core content areas. After the scores were collected from the teacher raters, the Pearson Performance Scoring Center (PSC) rescored the writing collections using a two-phase process described below.

### Phase I: Initial Scoring

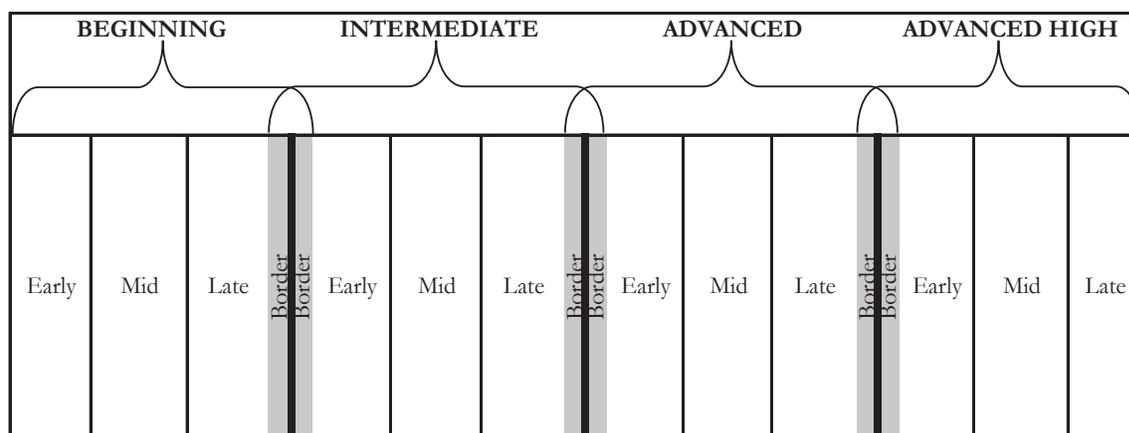
As part of the first phase, audit raters were trained to rate collections using the same rubrics and training sets used to train teacher raters. In order to qualify, the audit raters were required to rate correctly at least 80% of the writing collections from a qualifying set, a set of 15 writing collections which had been previously rated. There were two qualifying sets. If an audit rater did not qualify on the first qualifying set administered, he or she needed to qualify on the second qualifying set to participate in the audit study.

After audit raters were trained and qualified, each writing collection was scored independently by two audit raters without knowledge of the teacher rating. If the two audit ratings matched, the Phase I rating was recorded. In cases where the audit raters did not match, the collection was read by a third audit rater. Then, if two of three audit raters agreed, this rating was recorded and processed. If disagreement was still present, a highly qualified senior rater resolved the rating. Throughout this process, periodic checks were made to help ensure against rater drift among the audit raters, including rescoring of approximately 5% of Phase 1 ratings by trained scoring supervisors to identify any issues with rater drift. If rater drift was identified through these periodic checks, feedback was provided to the raters to correct the issue.

### Phase II: Analytic Scoring of Collections

Multiple purposes motivated Phase II of the scoring process. First, a second phase enabled the state to verify Phase I scores where the ratings from the audit raters and teacher were different. Second, the results from Phase II enabled the state to learn more about the reasons for adjacent ratings. Finally, the state was able to identify writing collections that could be useful for future training materials.

In Phase II, the trained team of scoring supervisors, who supervised the Phase 1 audit raters, rescored all writing collections for which the teacher and audit ratings differed by at least one proficiency level. The training conducted for this phase of scoring prepared the team of scoring supervisors to identify collections as Early, Mid, or Late within a proficiency level range or as Border (near the line between two proficiency levels) as shown in Figure 1.



**Figure 1. Depiction of Phase II Classifications**

Collections were read aloud to the team of scoring supervisors, to avoid any bias that could be created due to phonetic spelling of words. The team did not know the Phase I audit rating nor the teacher rating, only that there was a difference between the two ratings. After hearing each student's collection read aloud, the team reached consensus about the proficiency level of the student. In the Phase II

scoring, the collections were assigned not only a proficiency level score but also a designation of Border, Early, Mid, or Late. All collections receiving Border designations were subsequently rated by the lead state trainer from Pearson or by a Pearson scoring director. The ratings from Phase II (which reflected additional, more targeted training) overrode the ratings from Phase I if they differed.

Final ratings from Phase I where collections did not need a Phase II review, and final ratings from Phase II, together make up the full set of audit rating results. These ratings will be referred to throughout the remainder of the report as *final audit ratings*.

### 2016 Audit Results

Of the 2,208 student writing collections requested, sampled schools submitted 2,182 (98.8%) collections. Of the 2,182 collections received, 7 (0.32%) were classified as nonscorable by the scoring center. Collections were deemed nonscorable if they could not be used to judge the overall writing proficiency of a student because of an insufficient number, variety, or type of writing samples. Consequently, a total of 2,175 writing collections were re-rated by the state.

**Table 3. Distribution of Ratings by Teachers and Audit Raters**

Proficiency Level	Ratings by Teachers		Ratings by Audit Raters	
	N	%	N	%
Beginning	539	25	458	21
Intermediate	546	25	554	25
Advanced	550	25	659	30
Advanced High	547	25	504	23
<b>Total</b>	<b>2,182</b>	<b>100</b>	<b>2,175</b>	<b>100</b>

The distribution of ratings by teachers and the audit raters is provided in Table 3. The sample was designed to obtain equal numbers of collections across grade bands using teacher-assigned proficiency levels. Although this objective was achieved, disagreements between teacher and audit ratings resulted in more writing collections classified into the advanced proficiency level and less into the beginning and advanced high proficiency levels by audit raters.

Tables 4, 5, 6, and 7 include the perfect, modified perfect, adjacent, and modified adjacent agreement rates between the audit and teacher ratings, respectively. The teacher rating was considered to have perfect agreement with the audit rating if the teacher rating was the same as the final audit rating. The final audit ratings agreed perfectly with the teacher ratings 81% of the time overall (refer to Table 4).

**Table 4. Perfect Agreement Rates between Teacher and Audit Raters by Grade Band**

Grade Band	% Agreement
2	77
3–5	83
6–8	79
9–12	87
<b>Overall</b>	<b>81</b>

The modified perfect agreement rates were computed by considering audit ratings in agreement with teacher ratings if 1) the teacher rating was the same as the final audit rating or 2) a border designation from the Phase 2 audit rating was within a proficiency level adjacent to the rating provided by the teacher. The overall modified perfect agreement rate was 85% (refer to Table 5).

**Table 5. Modified Perfect Agreement Rates between Teacher and Audit Raters by Grade Band**

<b>Grade Band</b>	<b>% Agreement</b>
2	79
3–5	87
6–8	86
9–12	88
<b>Overall</b>	<b>85</b>

The teacher and audit ratings were considered adjacent if the teacher rating was the same as or within one proficiency level of the final audit rating. The adjacent agreement rate was 99% overall (refer to Table 6).

**Table 6. Adjacent Agreement Rates between Teacher and Auditor Raters by Grade Band**

<b>Grade Band</b>	<b>% Agreement</b>
2	99
3–5	99
6–8	99
9–12	99
<b>Overall</b>	<b>99</b>

The modified adjacent agreement rates were computed by considering audit ratings in agreement with teacher ratings if 1) the teacher rating was the same as the final audit rating, or 2) an early, late, or border designation from the Phase 2 audit rating was adjacent to the proficiency level provided by the teacher. For example, if the teacher rated the student as Intermediate, and the audit rating was either borderline Advanced or early Advanced, this would be considered a modified adjacent agreement. The modified adjacent agreement rate was 93% overall (refer to Table 7).

**Table 7. Modified Adjacent Agreement Rates between Teacher and Audit Raters by Grade Band**

<b>Grade Band</b>	<b>% Agreement</b>
2	91
3–5	94
6–8	95
9–12	95
<b>Overall</b>	<b>93</b>

The minimum standard for perfect agreement between two raters of a performance-scored item or assessment depends upon three factors:

- 1) the subject area of the assessment,
- 2) the type of rubric (holistic vs. analytic), and
- 3) the number of points on the rubric.

For a writing assessment using a four-point holistic rubric, the standard is 70% perfect agreement (Pearson Performance Scoring Center International Organization for Standardization (ISO) Standards). The perfect agreement rates in Table 4 indicate that the TELPAS writing audit results exceeded the applicable standard.

Perfect agreement rates between the audit and teacher ratings are shown by proficiency level in Table 8. Perfect agreement was lowest (77%) for writing collections rated as Intermediate by the teacher, and highest (85%) for those rated as Advanced.

**Table 8. Perfect Agreement Rate between Teacher and Audit Raters by Proficiency Level**

<b>Teacher Rating</b>	<b>% Agreement with Audit Rating</b>
Beginning	82
Intermediate	77
Advanced	85
Advanced High	82
<b>Overall</b>	<b>81</b>

A cross-tabulation of teacher by audit ratings is provided in Table 9. Highlighted cells indicate perfect agreement. The sum of the highlighted cells is 81% or the overall perfect agreement rate provided in Tables 4 and 8. Table 9 also includes the frequencies of adjacent and non-adjacent ratings (non-highlighted cells). Cells above the highlighted cells indicate that the collections were rated higher by the audit raters than by the teachers. Cells below the highlighted cells indicate that the collections were rated higher by the teachers than by the audit raters. Of the collections that did not receive the same rating from teacher and audit raters, 11% of the collections were rated higher by the audit rater and 7% of the collections were rated higher by the teacher.

**Table 9. Relationship between Audit and Teacher Ratings**

Frequency Percent	Audit Rating <b>Beginning</b>	Audit Rating <b>Intermediate</b>	Audit Rating <b>Advanced</b>	Audit Rating <b>Advanced High</b>	Total Teacher Rating
Teacher Rating <b>Beginning</b>	438 20.1%	92 4.2%	3 0.1%	1 0.1%	534 24.6%
Teacher Rating <b>Intermediate</b>	19 0.9%	420 19.3%	98 4.5%	8 0.4%	545 25.1%
Teacher Rating <b>Advanced</b>	1 0.1%	37 1.7%	465 21.4%	46 2.1%	549 25.2%
Teacher Rating <b>Advanced High</b>	0 0.0%	5 0.2%	93 4.3%	449 20.6%	547 25.2%
Total Audit Ratings	458 21.1%	554 25.5%	659 30.3%	504 23.2%	2175 100%

Finally, two indices were calculated to evaluate further the degree of association between the teacher and audit ratings, inter-rater reliability and weighted Kappa (refer to Table 10). Inter-rater reliability is the correlation between the teacher and audit ratings. The teacher and audit ratings were found to be highly correlated ( $r = 0.912, p < 0.001$ ). Weighted Kappa ( $K_w$ ) also provides a measure of inter-rater agreement for categorical ratings. The  $K_w$  value can be interpreted using the criteria in Table 11 (Altman, 1991). The overall  $K_w$  from this study was 0.841 indicating a *very good* relationship between teacher and audit ratings.

**Table 10. Inter-Rater Reliability and Weighted Kappa between Teacher and Audit Ratings**

Grade Band	Inter-Rater Reliability	Weighted Kappa
2	0.892	0.800
3–5	0.923	0.856
6–8	0.906	0.825
9–12	0.931	0.883
<b>Overall</b>	<b>0.912</b>	<b>0.841</b>

**Table 11. Interpreting Kappa ( $K_w$ ) Statistics**

Value of $K_w$	Strength of Agreement
< 0.20	Poor
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Good
0.81 – 1.00	Very Good

### **Questionnaire Results**

As a part of the 2016 audit, questionnaires were given to testing coordinators from the selected districts and campuses as well as to the teachers who had rated the audited writing collections. Questionnaires were used to gather information about the training and qualification procedures in order to support the validity and reliability of the TELPAS assessment process. A brief description of the questionnaire results is provided here. The complete set of questions and responses is provided in the Appendix.

Rater training is provided to prepare teachers to rate English language proficiency in a manner consistent with the holistic rating rubrics—the PLDs from the English Language Proficiency Standards (ELPS). New teacher raters are required to complete an online grade-cluster specific basic training course. Student activities are provided for all holistically rated domains. There are two sets of online calibration activities for each grade cluster. If an audit rater does not calibrate on the first set administered, he or she will need to calibrate on the second set. For K–1, each set contains 10 students to be rated in the domains of listening, speaking, reading, and writing. Raters must rate at least 70% of the students correctly within a set to demonstrate sufficient calibration in order to qualify to rate students. For 2–12 grade clusters, each set contains 12 students to be rated in the domains of listening, speaking, and writing. Raters must rate at least 67% of the students correctly within a set to demonstrate sufficient calibration in order to qualify to rate students. Individuals not successful after taking both sets are either not used as raters (a district decision) or are provided rater support in accordance with test administration procedures.

In general, testing coordinators reported that proper procedures are in place to train raters and evaluate proper implementation of the holistic ratings. Specifically, most district testing coordinators (DTCs) reported using many of the training materials provided by TEA for teaching campus coordinators about administration procedures for the TELPAS holistic ratings. Additionally, 100% of DTCs reported implementing the required validity and reliability procedures including collaborative ratings and double rating of writing collections. Most districts (90%) selected the general procedures used to provide validity and reliability evidence for the TELPAS ratings, and many required campuses to keep documentation of the implemented procedures. The vast majority of DTCs (98%) reported that raters in their districts either passed the online training calibration by the end of the second calibration set or were provided with supplemental support. Also, 40% of DTCs reported having no uncalibrated raters in their districts after the first two attempts. Only one DTC reported submitting an incident report for a rater who did not complete online training activities. These results are based on a wide variety of districts, including districts that had between 1 and 10 raters trained (31%), and districts that had over 300 raters trained (8%).

The campus testing coordinator (CTC) responses were similar to the DTC responses. They reported using TEA's training materials to provide rater training. Over half of CTCs (58%) reported that all raters at their campuses had successfully calibrated by the end of the second calibration set. For those who had not calibrated, very few CTCs (1%) reported not providing supplemental rating support. The CTCs reported a wide range of campus sizes; 42% of campuses had 1-10 raters, while 15% had more than 26. More than half of the CTCs (61%) reported that no uncalibrated individuals served as raters in 2016; of those that did use uncalibrated raters, only 7% had three or more.

Teacher raters reported feeling sufficiently prepared to provide holistic ratings. Teachers reported that rater training sessions typically lasted between 30 minutes and 4 hours (89%). Most new raters (96%) found the online basic training course adequate to prepare them for student ratings. Nearly all raters (95%) indicated that the calibration activities provided adequate preparation for their TELPAS ratings. On calibration set 1, 76% of teachers successfully calibrated; on calibration set 2, another 20% successfully calibrated. Most raters (93%) who were unsuccessful at calibrating reported receiving rating support, as required by TEA. Raters included teachers spanning grades K–12, including special education (4%), gifted and talented (3%), bilingual (23%), and ESL (25%) teachers from all content areas. More than half of the teachers rated 10 or fewer students (62%), while only 2% of teachers reported rating 50 or more students.

In general, the questionnaire results provided additional evidence that the audit sample included a wide variety of districts, campuses, and teachers from around the state of Texas. Questionnaire responses also indicated that the vast majority of DTCs, CTCs, and teacher raters are implementing the TELPAS rater training and holistic rating procedures as described in the *2016 District and Campus Coordinator Manual* with fidelity.

### ***Conclusions/Next Steps***

The results of the 2016 audit provided evidence of rater accuracy at a slightly higher level than what was reported in the previous 2013 audit. The overall perfect agreement rate of 81% was found to be satisfactory based on the *Pearson Performance Scoring Center ISO Standards*, and the adjacent agreement rate was 99%. In addition, the high correlation ( $r = 0.912$ ) and high weighted kappa value ( $K_w = 0.841$ ) underscored the strong agreement between the audit ratings and the teacher ratings.

Furthermore, consistently high rating accuracy across four consecutive audits (perfect agreement rates were 76% in 2007, 79% in 2008, 77% in 2013, and 81% in 2016) provides evidence of stability in

inter-rater agreement over time. Since TELPAS scores across years are used in reporting student progress in language acquisition, the state's finding of rater accuracy over time supports inferences about annual student progress from TELPAS scores.

Finally, the audit questionnaire results indicated that the vast majority of testing coordinators are implementing the training and ratings procedures as intended by TEA. The district testing coordinator questionnaire results indicated that 100% of districts implemented procedures to support the validity and reliability of the TELPAS rating process. The campus testing coordinator questionnaire results indicated that campuses were appropriately implementing the supplemental support activities for raters who were unsuccessful after their second calibration attempt and that they were implementing the validity and reliability procedures for the TELPAS rating process. Questionnaire results also indicated that the vast majority of raters reported that they had received adequate training if they were a new rater (95%), were successful on the online calibration component of rater training for grades 2–12 (96%), and had adequate information to rate their students' proficiency levels in each of the language domains assessed (95%).

The audit revealed no significant problems with implementation of training procedures or rater accuracy. TEA plans to continue the writing audit process to provide ongoing monitoring of teacher rater effectiveness and to give district personnel feedback and training materials that best support the ability of raters to conduct this assessment.

## References

Altman, D. G. (1991). *Practical Statistics for Medical Research*. London, UK: Chapman and Hall/CRC Press.

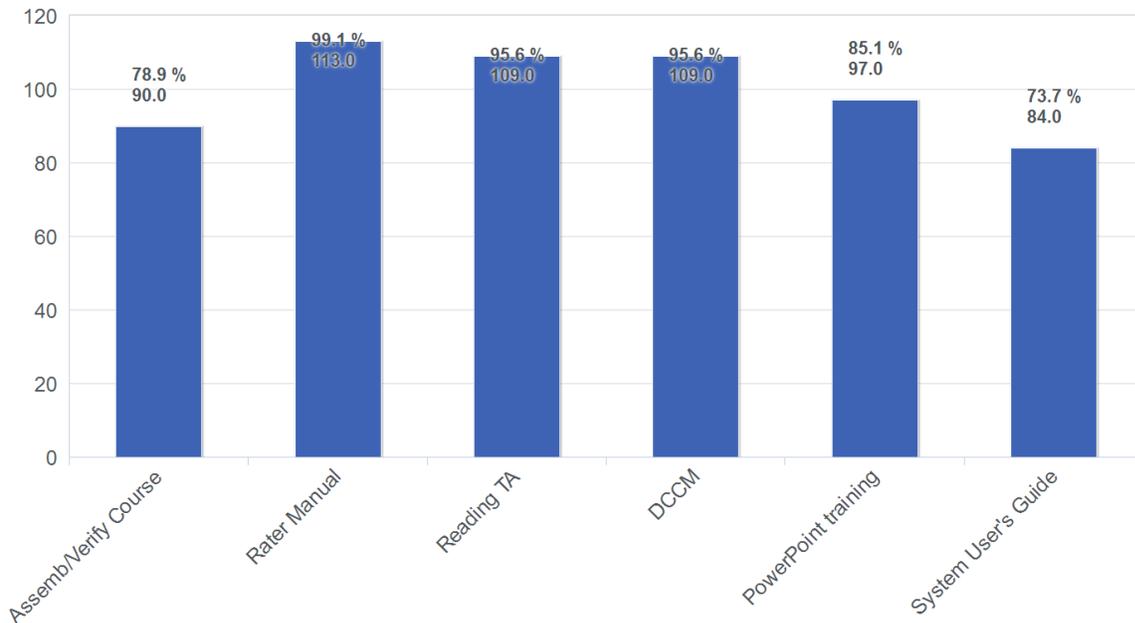
## Appendix: 2016 Writing Audit Questionnaire Results

Questionnaire results are provided for district testing coordinators, campus testing coordinators, and teacher raters below. The questions and answers provided are for the multiple-choice items only.

### *District Testing Coordinators*

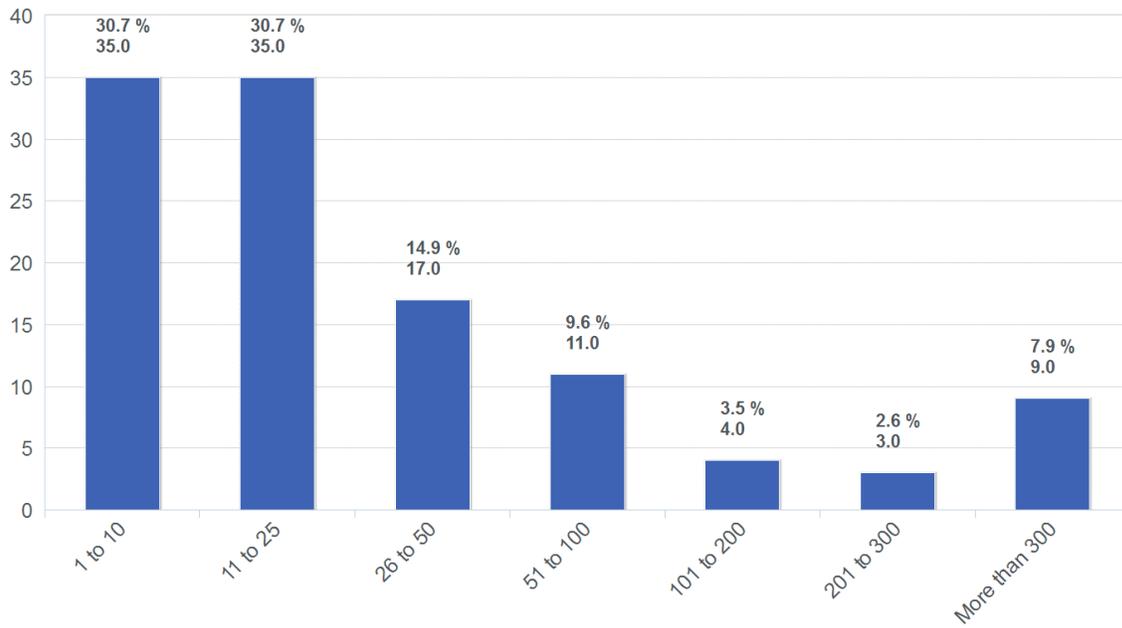
District testing coordinators (DTCs) at districts selected for participation in the 2016 TELPAS writing audit were asked to respond to a set of 12 questions. The first question asked them to select their district name. Questions 2–12 are provided below along with their responses. In general, DTCs reported that proper procedures are in place to train raters and evaluate proper implementation of the holistic ratings. A total of 115 responses were received from DTCs.

2. What materials did you use to prepare campus coordinators to conduct TELPAS administration procedures training? (Mark all that apply)

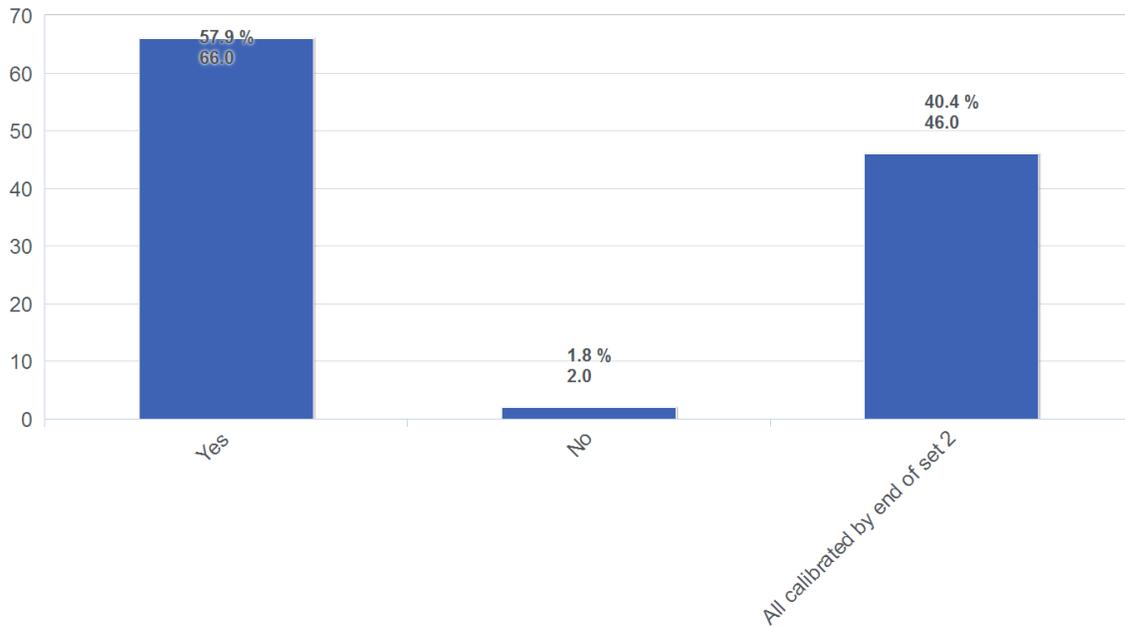


Note: Respondents were allowed to select more than one category. Therefore, totals do not necessarily equal 100%.

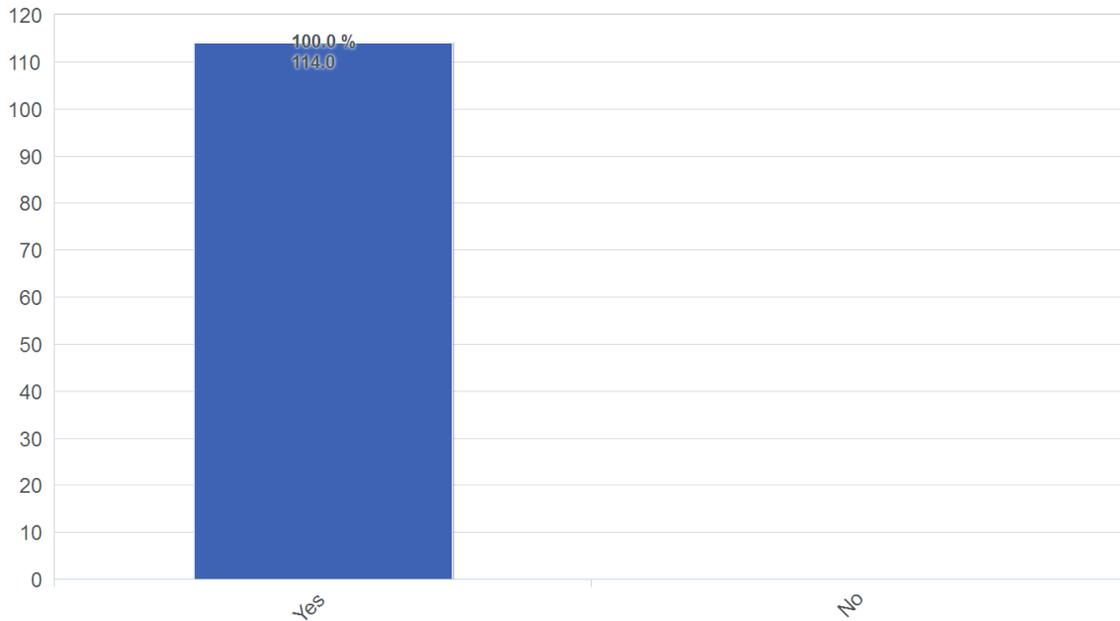
3. Approximately how many individuals served as raters in your district in spring 2016? (Include both new and returning raters.)



4. Were all raters in your district who did not successfully calibrate by the end of set 2 given supplemental support by a trained supplemental support provider?

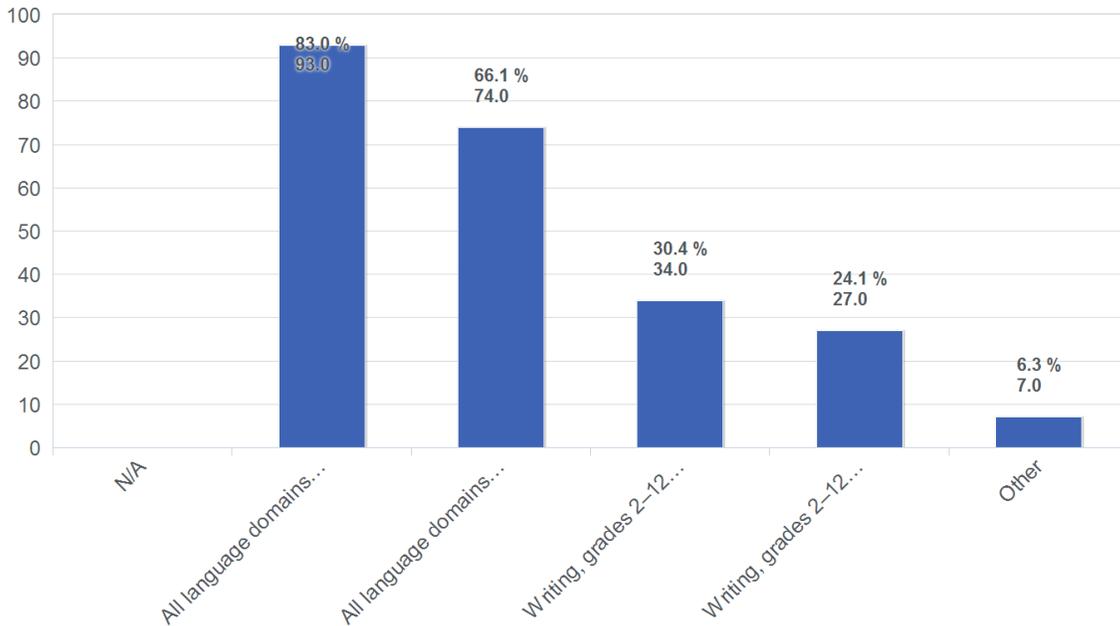


4. Did your district implement procedures to support the validity and reliability of the TELPAS rating process, as required by page T-21 and T-22 of the 2016 District and Campus Coordinator Manual?



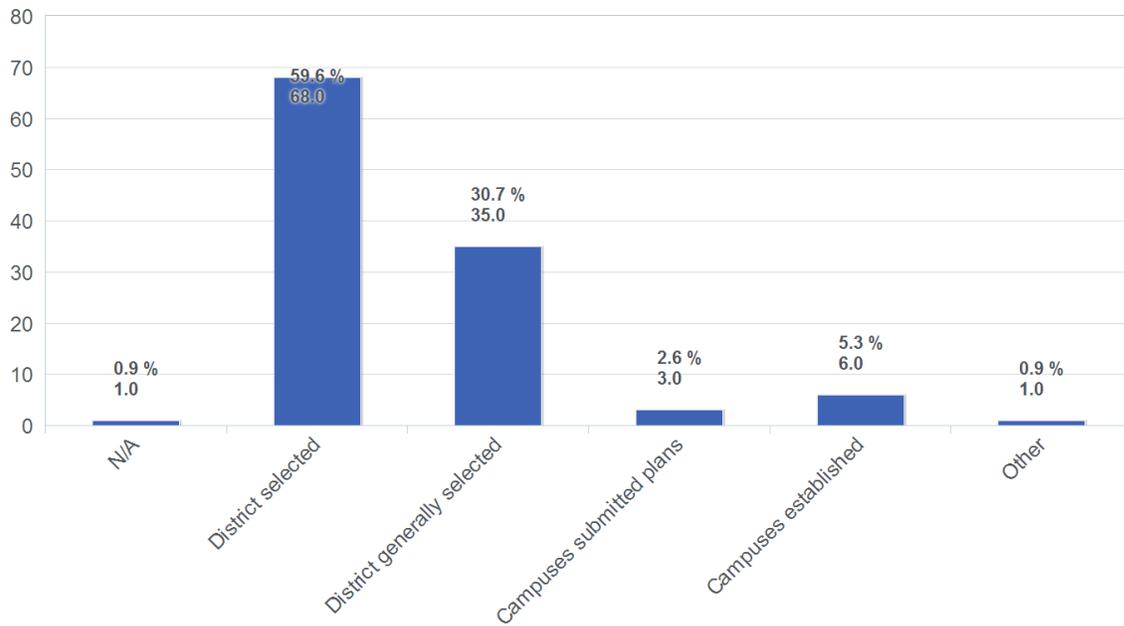
6. What procedures were implemented? (Mark all that apply.)

- All language domains, K–12: Raters collaborated with other teachers of the students in determining the students' ratings.
- All language domains, K–12: Raters collaborated with each other to determine the ratings of students near the border between proficiency levels.
- Writing, grades 2–12: A district-determined selection of writing collections were rated a second time, and if the ratings differed, resolution ratings were provided and used.
- Writing, grades 2–12: All writing collections were rated a second time, and if the ratings differed, resolution ratings were provided and used.
- Other

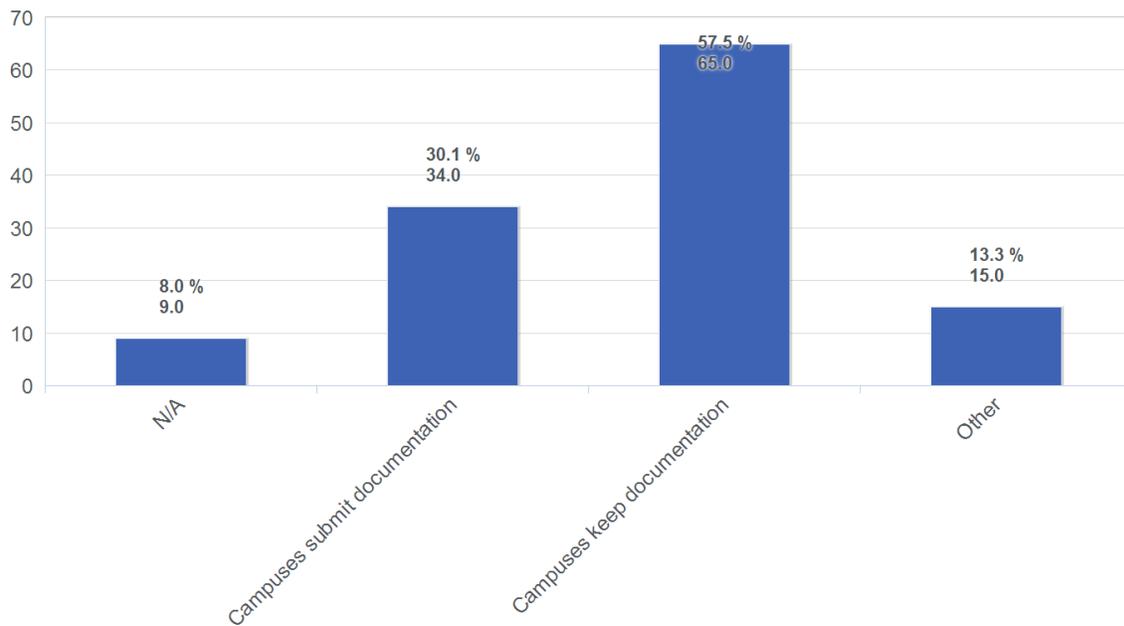


Note: Respondents were allowed to select more than one category. Therefore, totals do not necessarily equal 100%.

7. Did your district establish districtwide procedures, or did campuses select the procedures to follow?

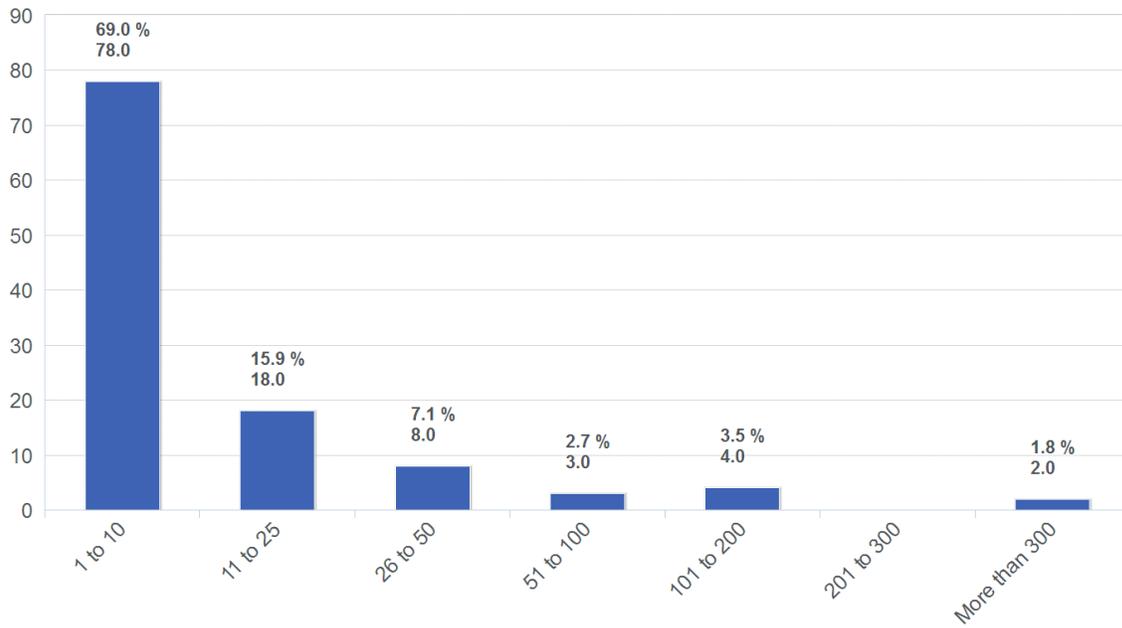


8. What steps did your district take to ensure that campuses implemented the validity and reliability procedures? (Mark all that apply.)

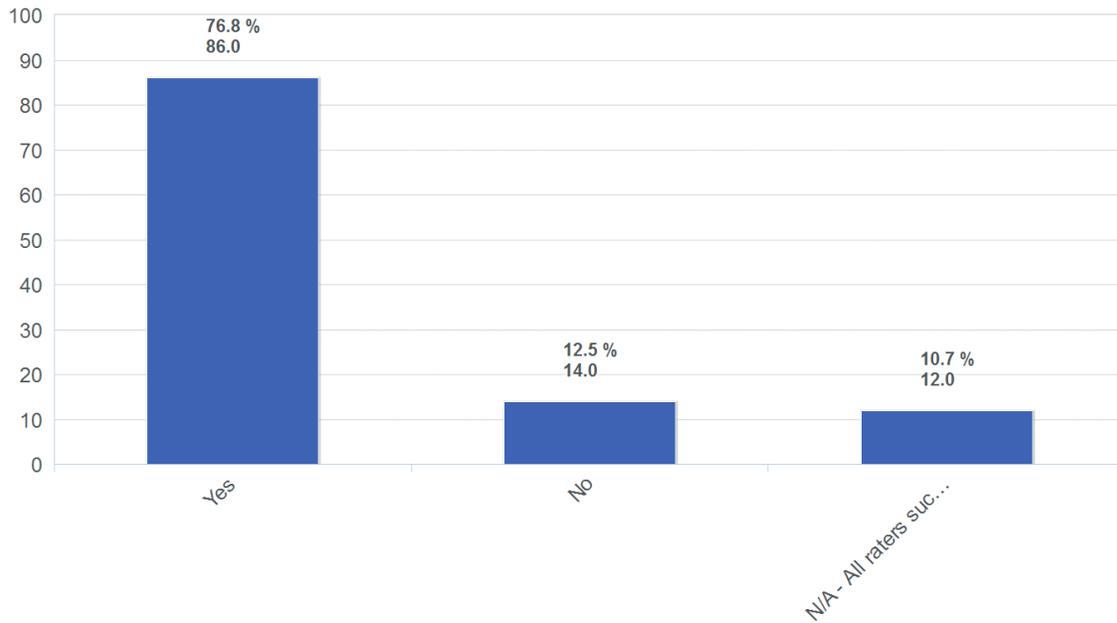


Note: Respondents were allowed to select more than one category. Therefore, totals do not necessarily equal 100%.

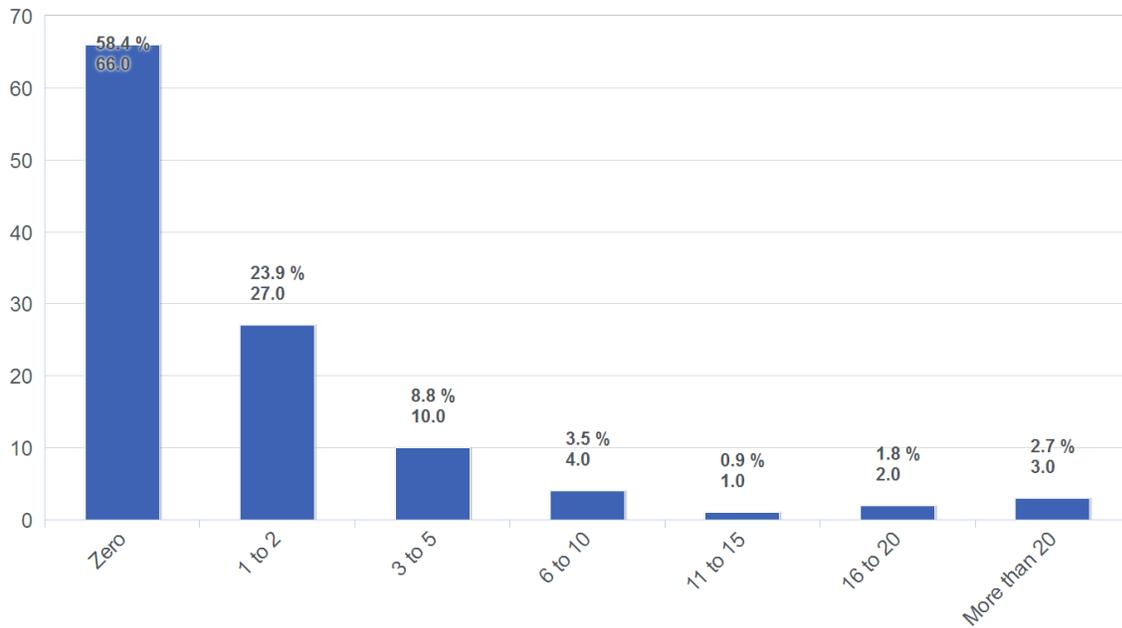
9. Approximately how many monitored calibration sessions were conducted in your district in spring 2016?



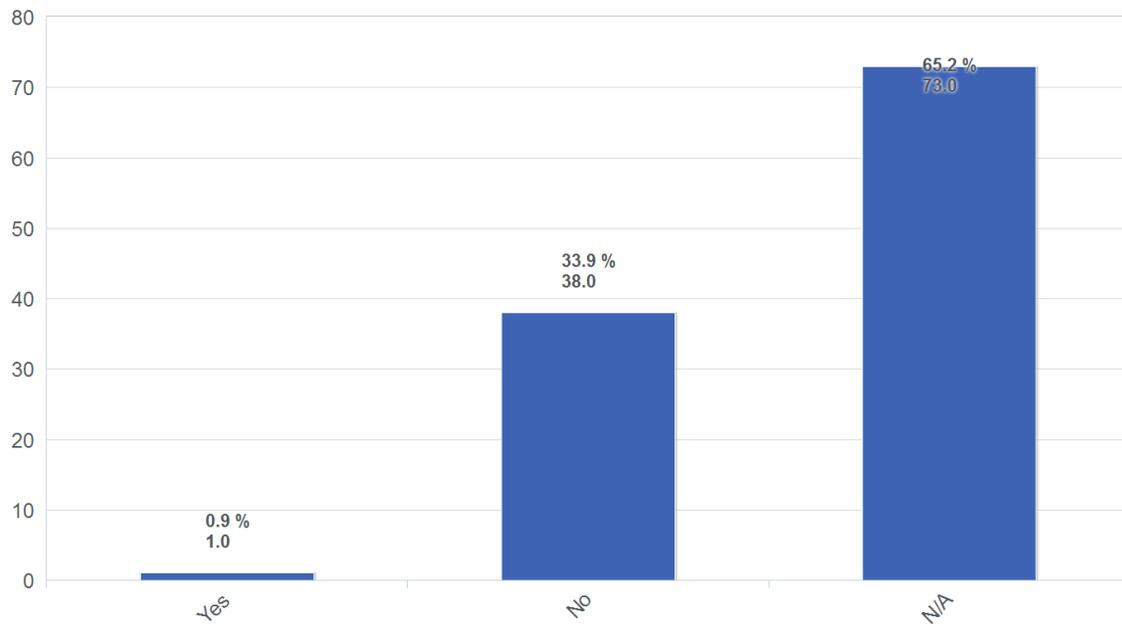
10. Were separate monitored calibration sessions conducted for raters who did not successfully calibrate on set 1?



11. Approximately how many uncalibrated individuals (individuals who completed both calibration sets unsuccessfully) served as raters in your district in spring 2016?



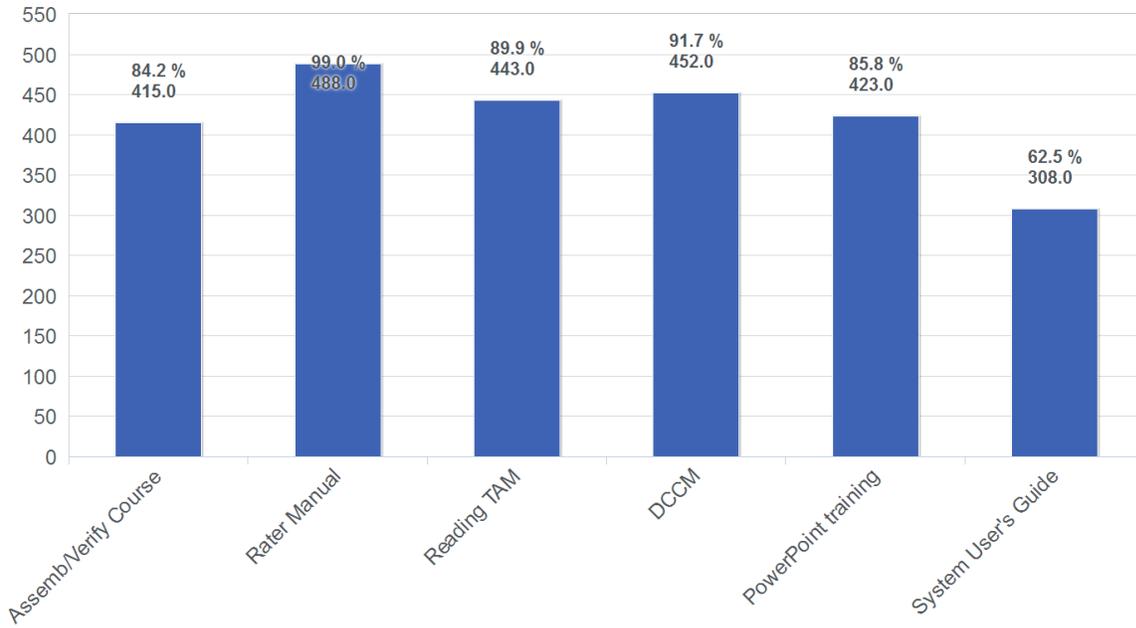
12. Were incident reports submitted for any raters that did not complete required online training activities?



### ***Campus Testing Coordinators***

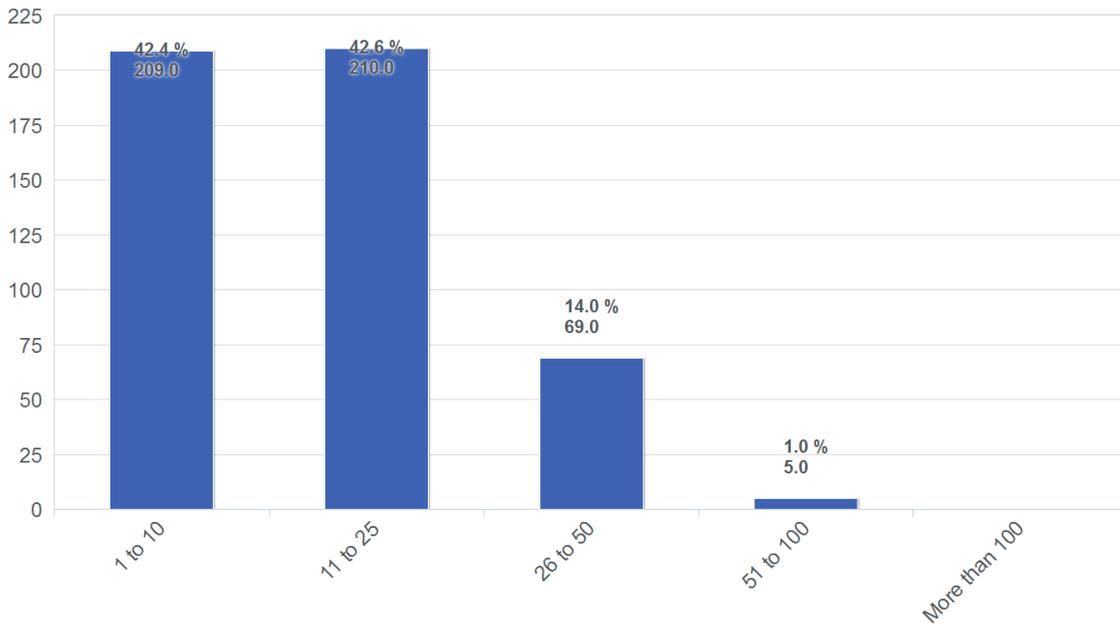
Campus testing coordinators (CTCs) at campuses selected for participation in the 2016 TELPAS writing audit were asked to respond to a set of 12 questions. The first question asked them to select their district and campus names. Questions 2–12 are provided below along with their responses. In general, CTCs reported that proper procedures are in place to train raters and evaluate proper implementation of the holistic ratings. A total of 494 responses were provided by CTCs.

2. What materials did you use to conduct TELPAS administration procedures training? (Mark all that apply)

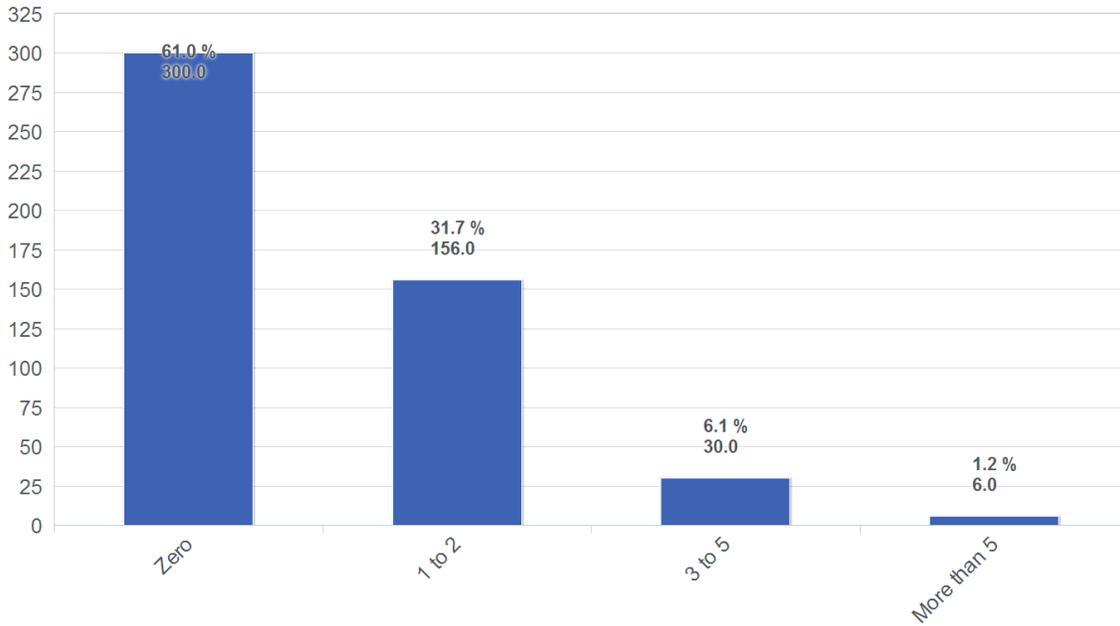


Note: Respondents were allowed to select more than one category. Therefore, totals do not necessarily equal 100%.

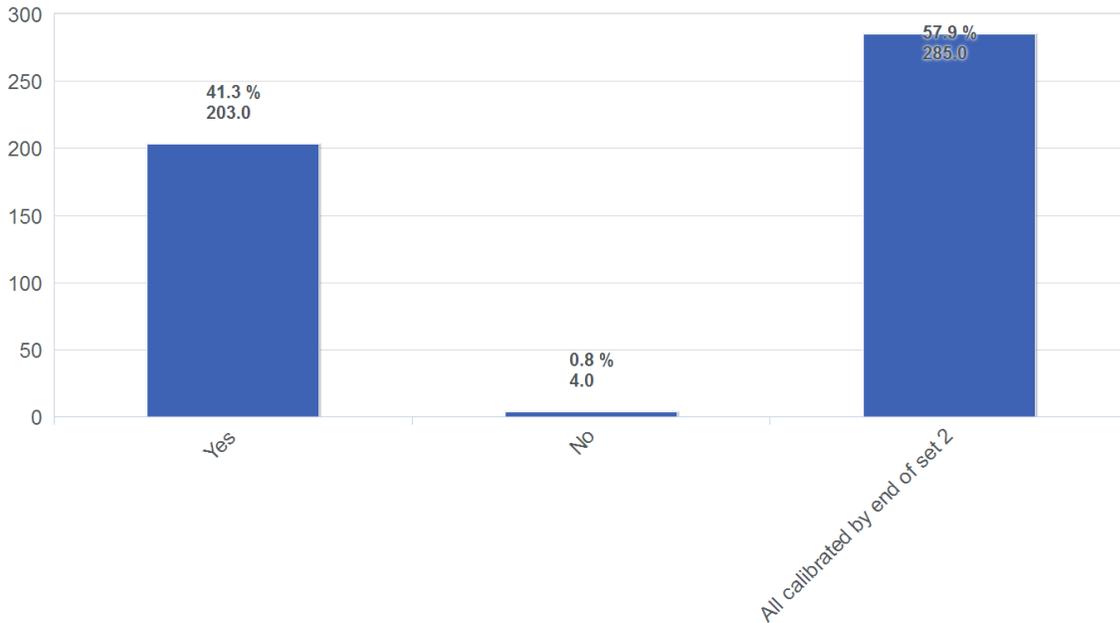
3. Approximately how many individuals served as raters on your campus in spring 2016? (Include both new and returning raters.)



4. Approximately how many individuals who served as raters on your campus did not calibrate by the end of calibration set 2?



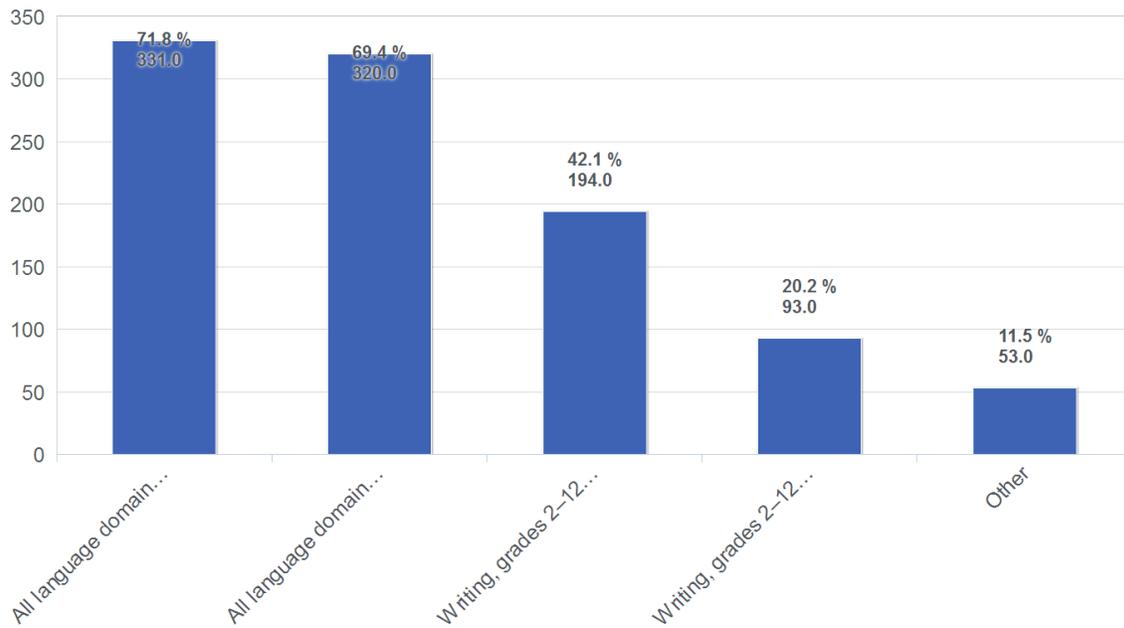
5. Were all raters on your campus who did not successfully calibrate by the end of set 2 given rating support?



6. What procedures were implemented on your campus to support the validity and reliability of the TELPAS rating process? See pages T-21 and T-22 of the 2016 District and Campus Coordinator Manual for more information. (Mark all that apply).

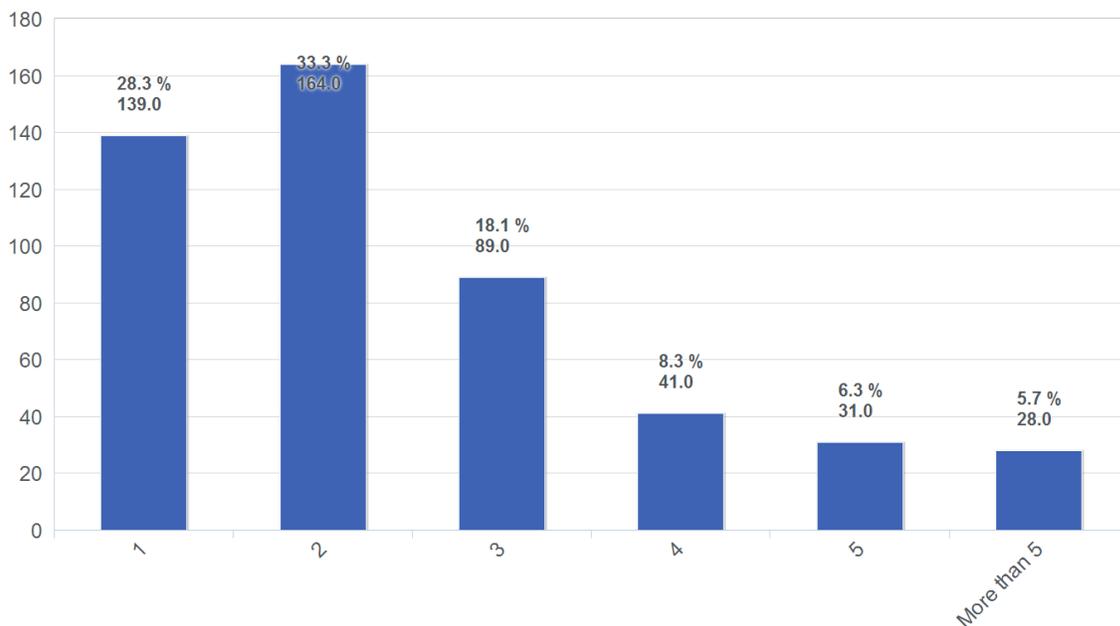
- All language domains, K–12: Raters collaborated with other teachers of the students in determining the students' ratings.
- All language domains, K–12: Raters collaborated with each other to determine the ratings of students near the border between proficiency levels.
- Writing, grades 2–12: A district-determined selection of writing collections were rated a second time, and if the ratings differed, resolution ratings were provided and used.

- Writing, grades 2–12: All writing collections were rated a second time, and if the ratings differed, resolution ratings were provided and used.
- Other

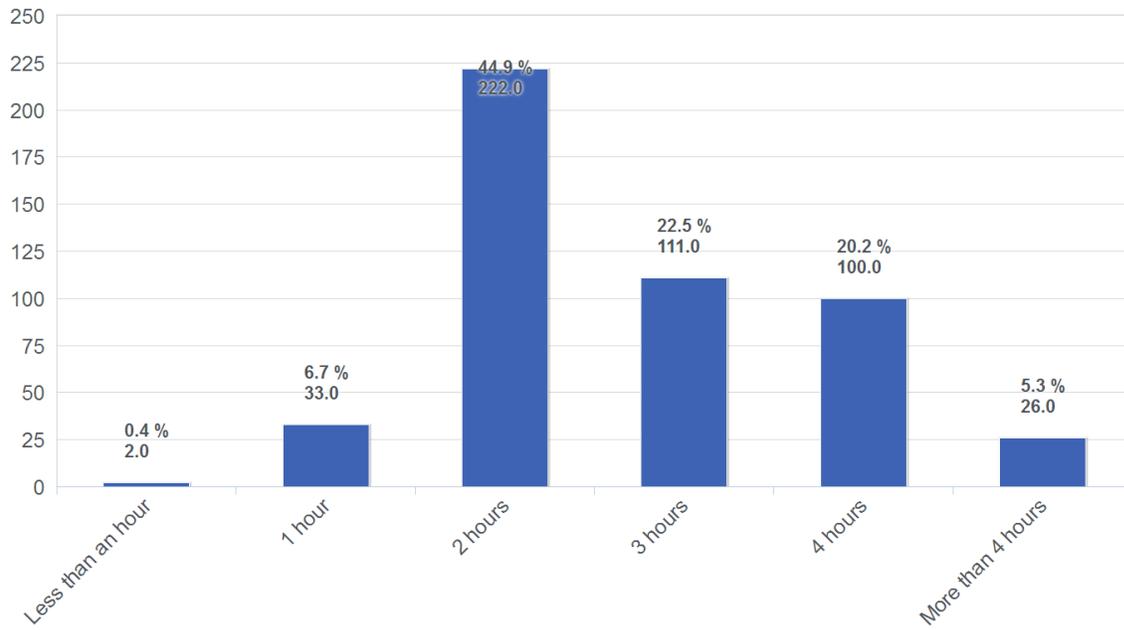


Note: Respondents were allowed to select more than one category. Therefore, totals do not necessarily equal 100%.

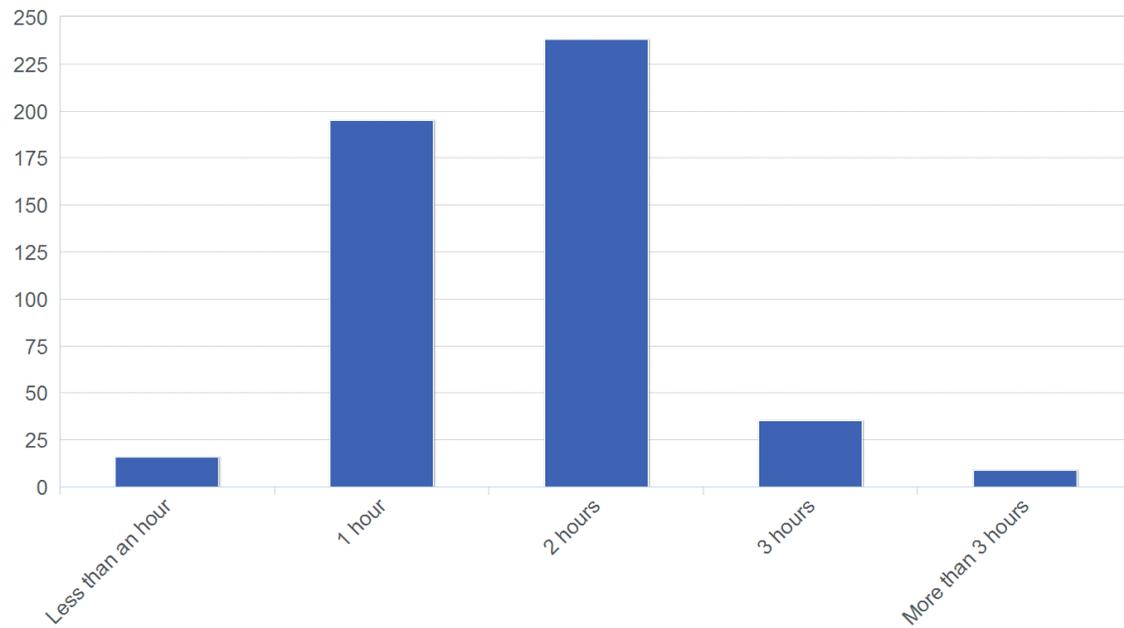
7. Approximately how many monitored calibration sessions were conducted on your campus in spring 2016?



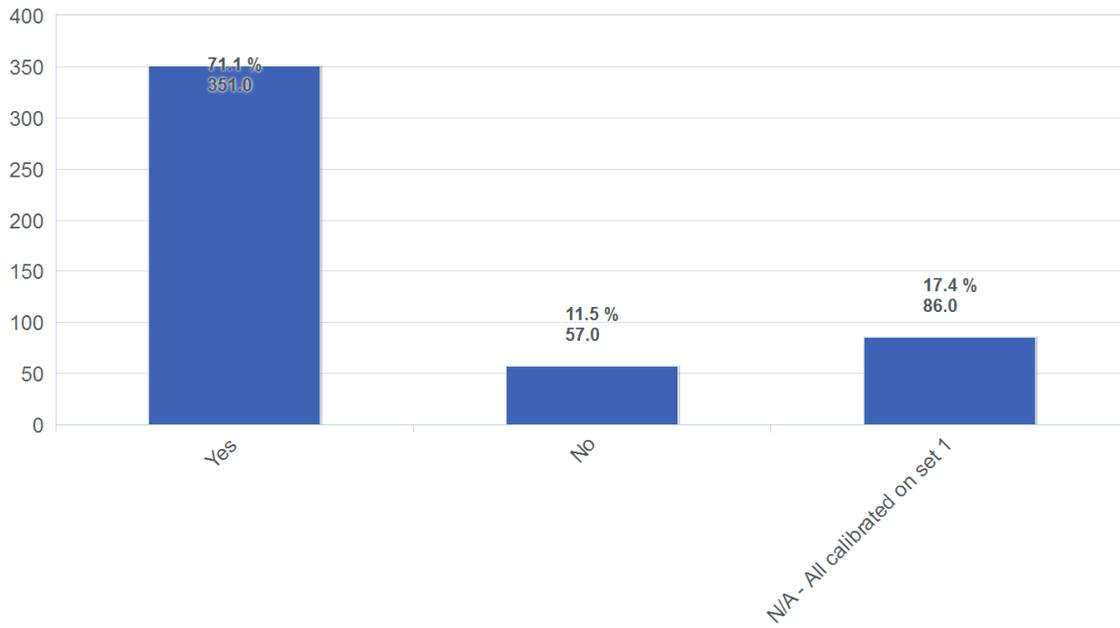
8. On average how long was each monitored calibration session scheduled for?



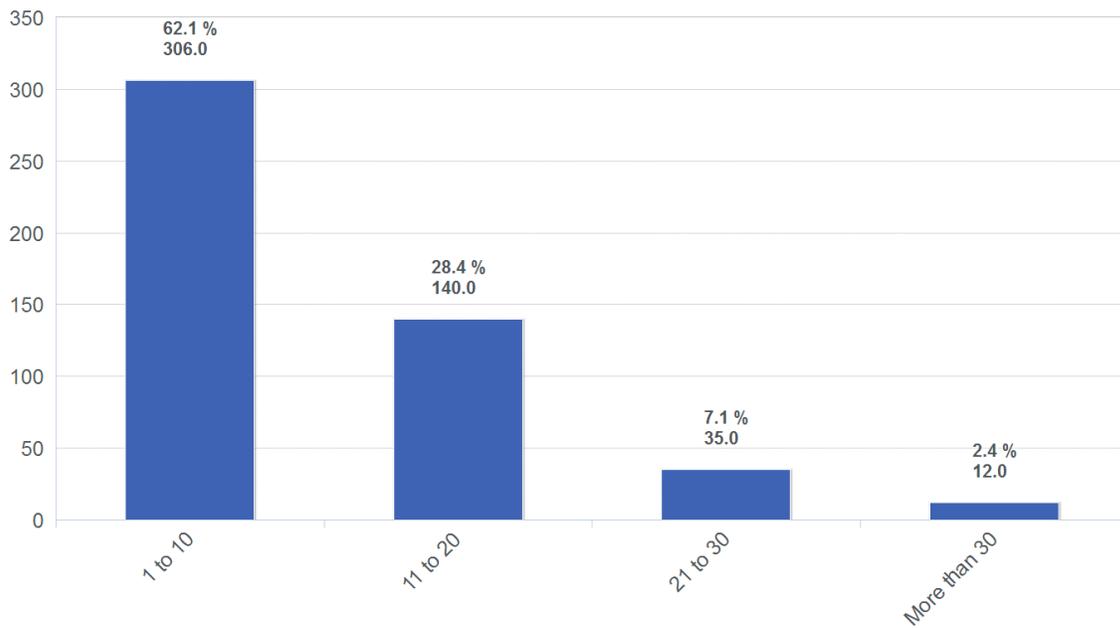
9. On average how long did it take a rater to complete a calibration set during a monitored calibration session?



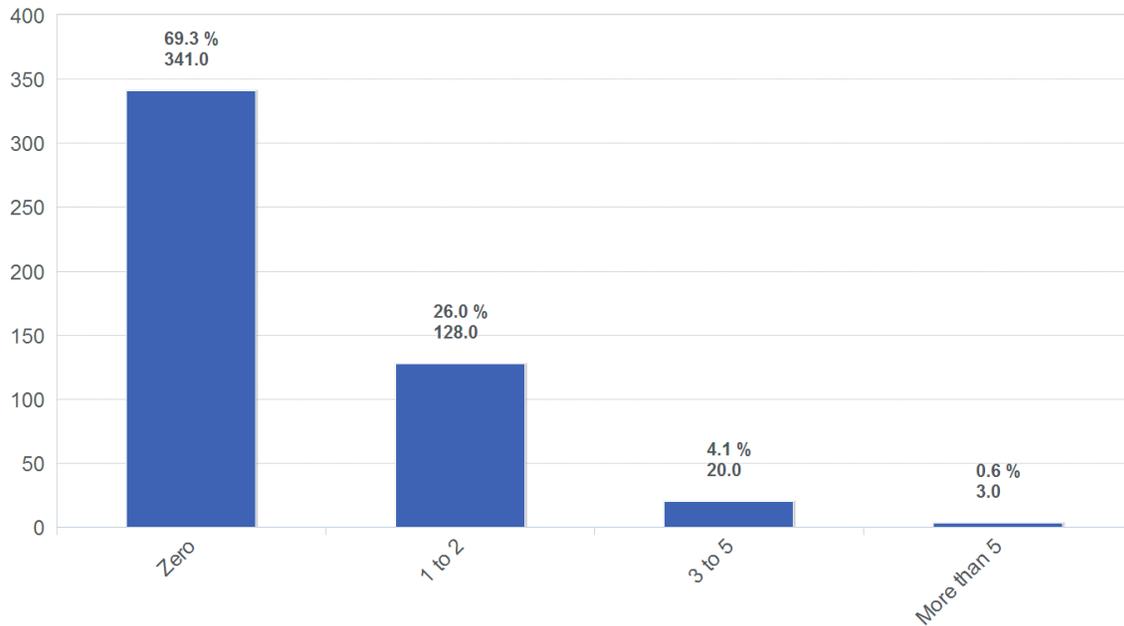
10. Were separate monitored calibration sessions conducted for raters who did not successfully calibrate on set 1?



11. Approximately how many raters attended each monitored calibration session?



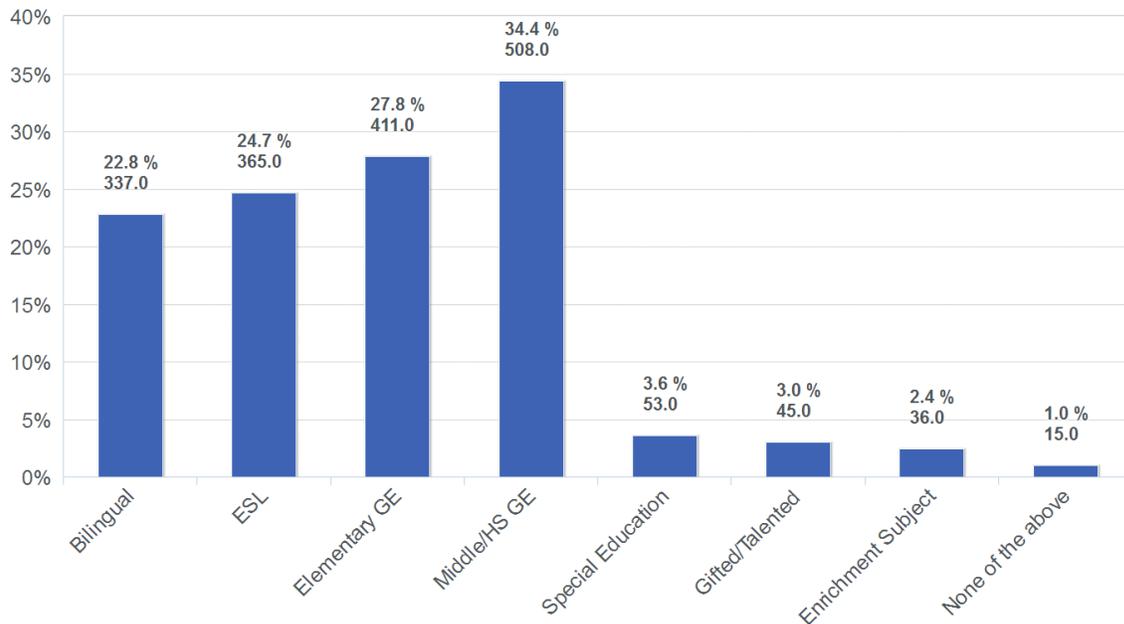
12. Approximately how many uncalibrated individuals (individuals who completed both calibration sets unsuccessfully) served as raters on your campus in spring 2016?



**Teacher Raters**

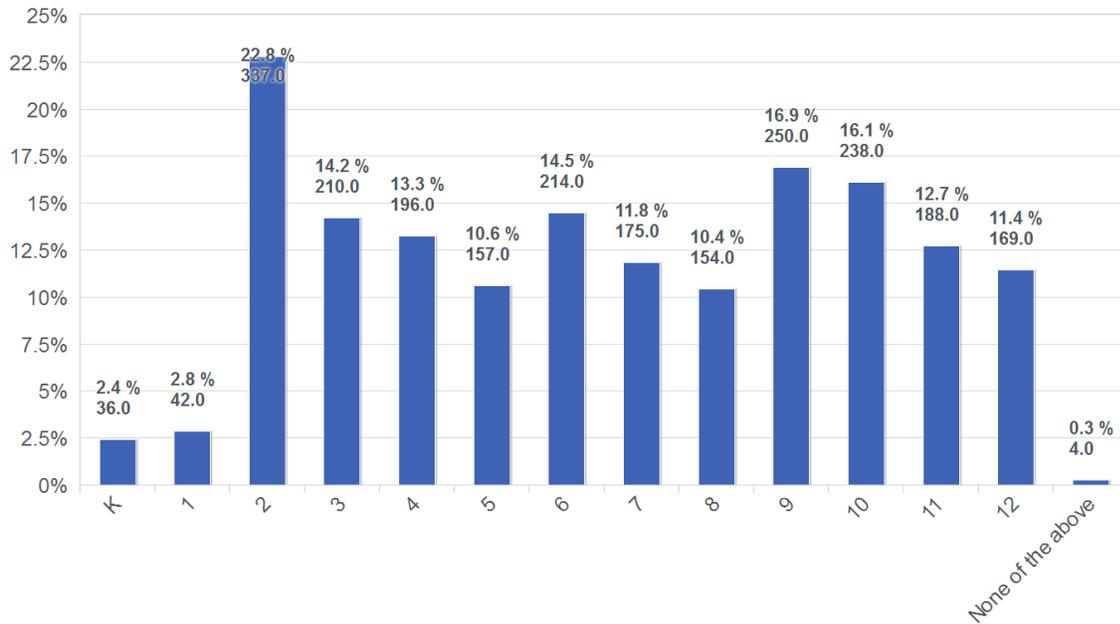
Teacher raters selected for participation in the 2016 TELPAS writing audit were asked to respond to a set of 11 questions. The first question asked them to select their district and campus names. Questions 2-11 are provided below along with their responses. In general, teachers reported receiving adequate training to provide holistic ratings and were confident in the scores they provided. A total of 1,479 responses were provided by teacher raters.

2. Choose the role that best describes you. (Mark all that apply.)



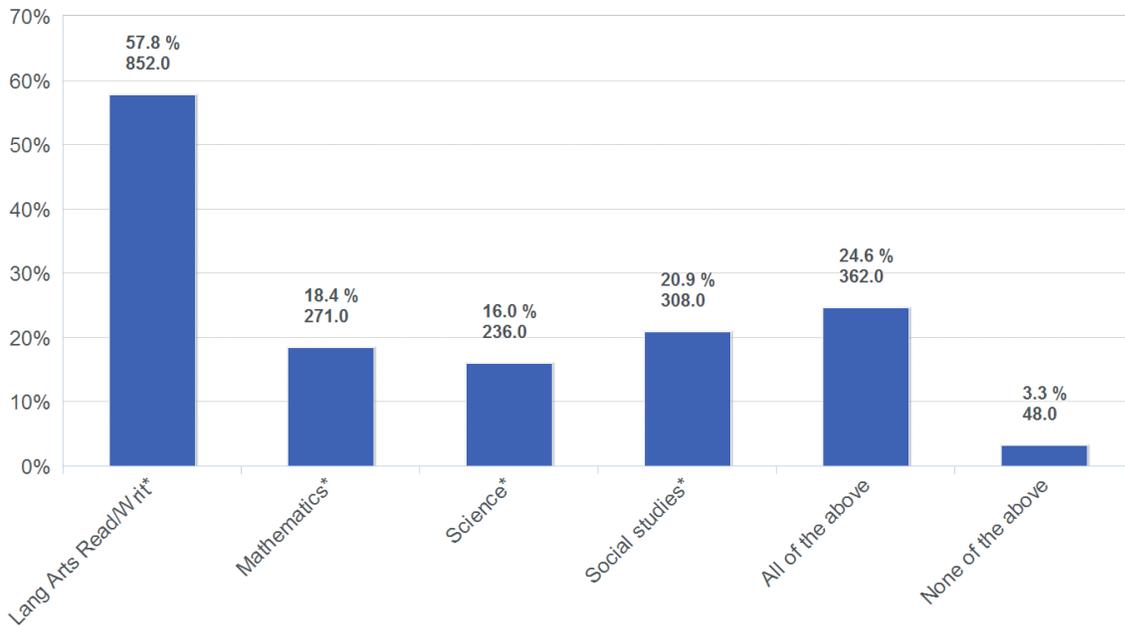
Note: Respondents were allowed to select more than one category. Therefore, totals do not necessarily equal 100%.

3. What grade(s) do you teach? (Mark all that apply.)



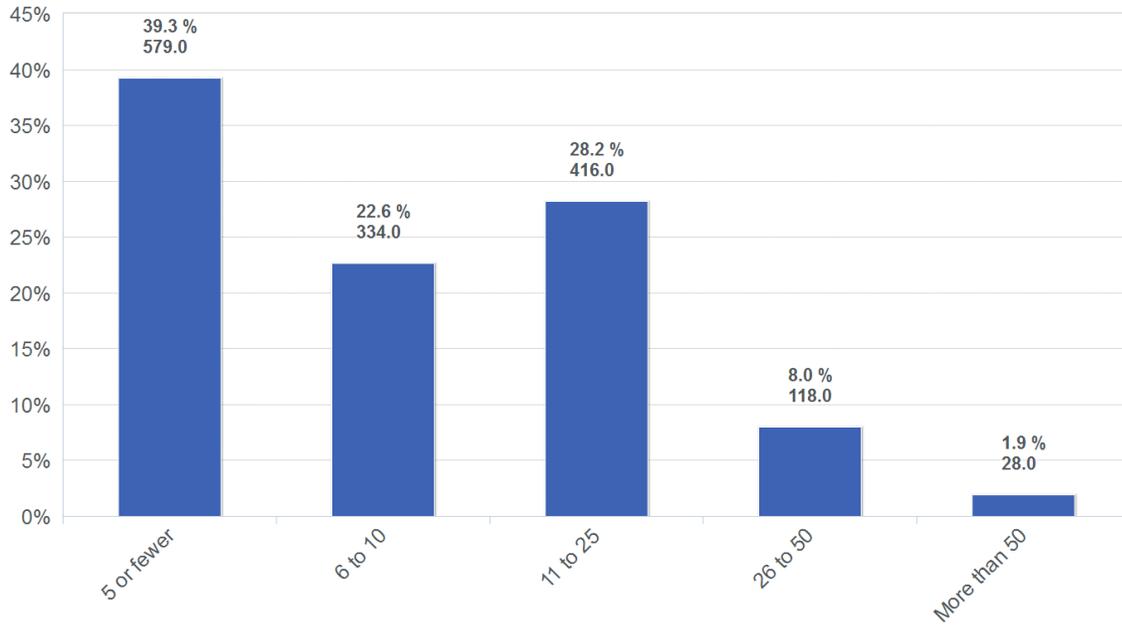
Note: Respondents were allowed to select more than one category. Therefore, totals do not necessarily equal 100%.

4. In which foundation subjects do you teach the student(s) you rated? (Mark all that apply.) \*Any course within this discipline.

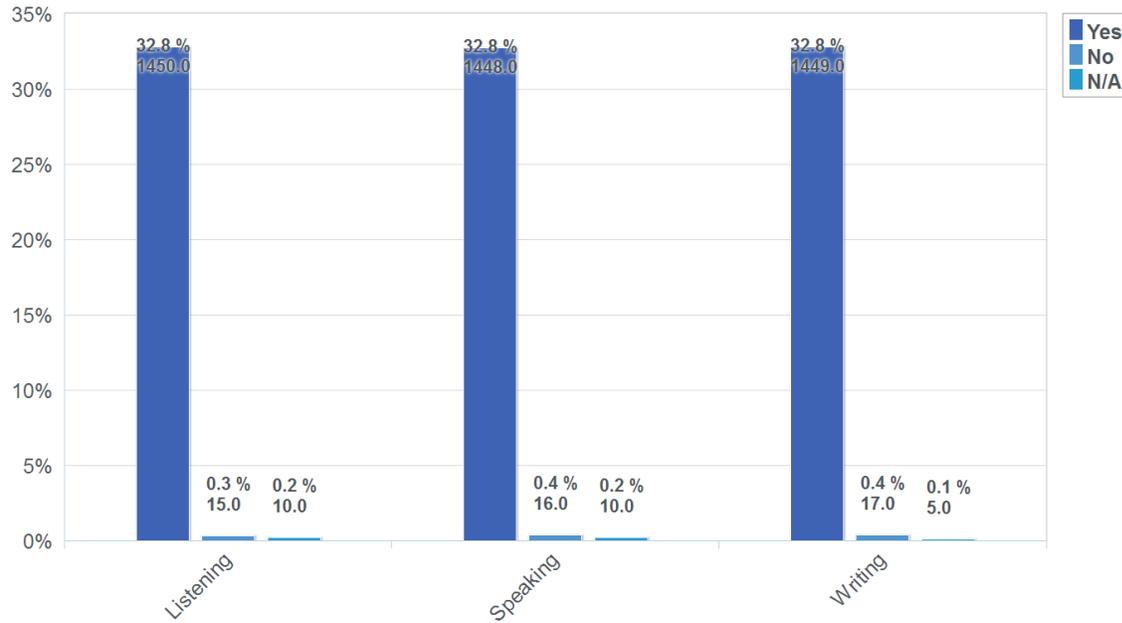


Note: Respondents were allowed to select more than one category. Therefore, totals do not necessarily equal 100%.

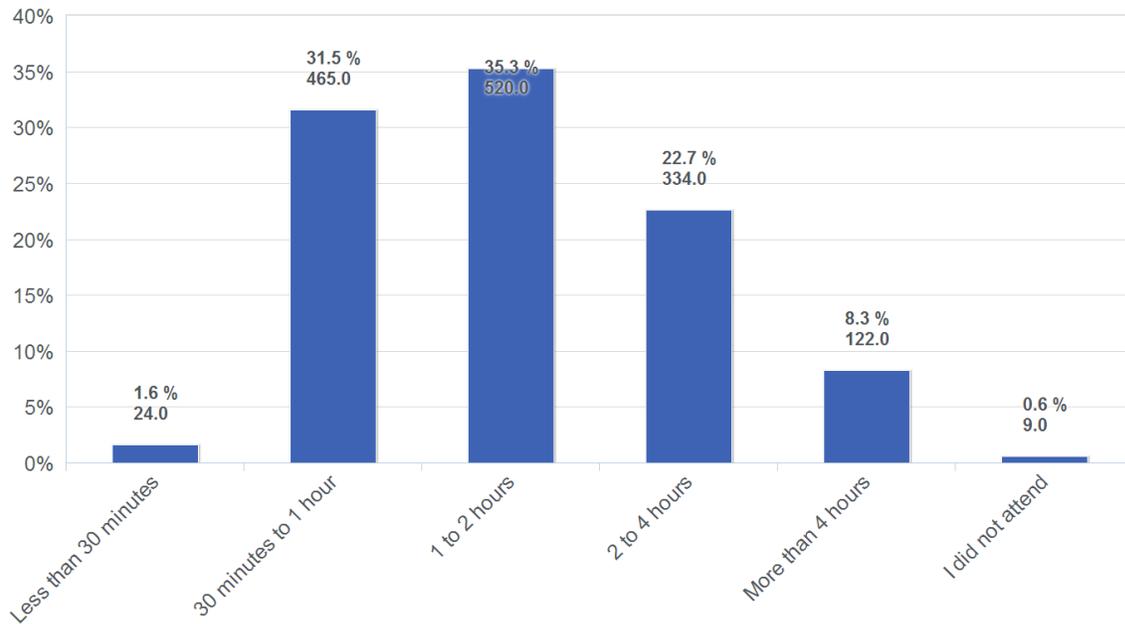
5. Approximately how many students did you rate?



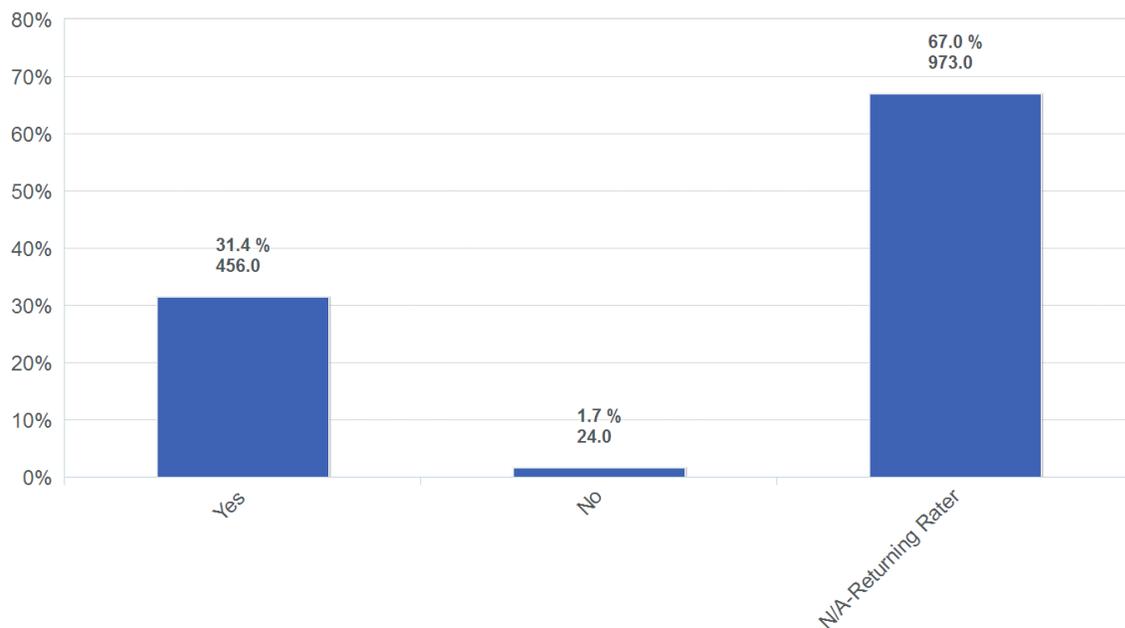
6. Did you feel that you had enough information about the English language abilities of your students to make judgments about their proficiency levels in each domain?



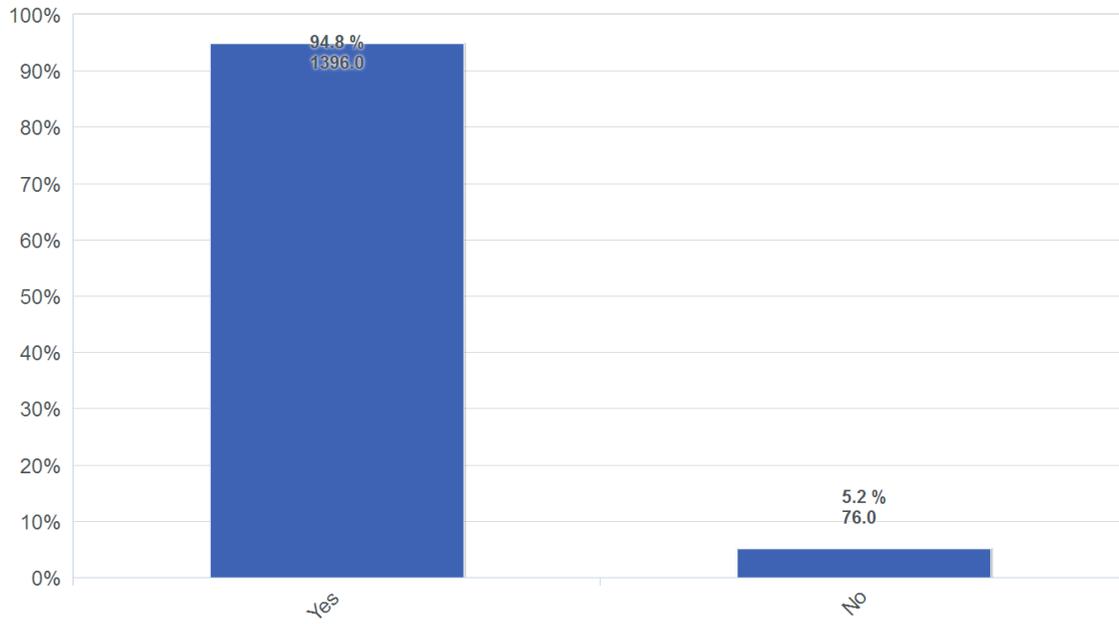
7. How long was the TELPAS administration procedures training session for raters and verifiers you attended this spring? This training covers security procedures, how to assemble writing collections, fundamentals of the holistic rating process for new raters, holistic rating training requirements for raters, etc. If this training was conducted in conjunction with other training, please indicate just the approximate amount of time spent on TELPAS administration procedures. If you completed the Assembling and Verifying Grades 2-12 Writing Collections online course, do not include the completion time for that course in your response.



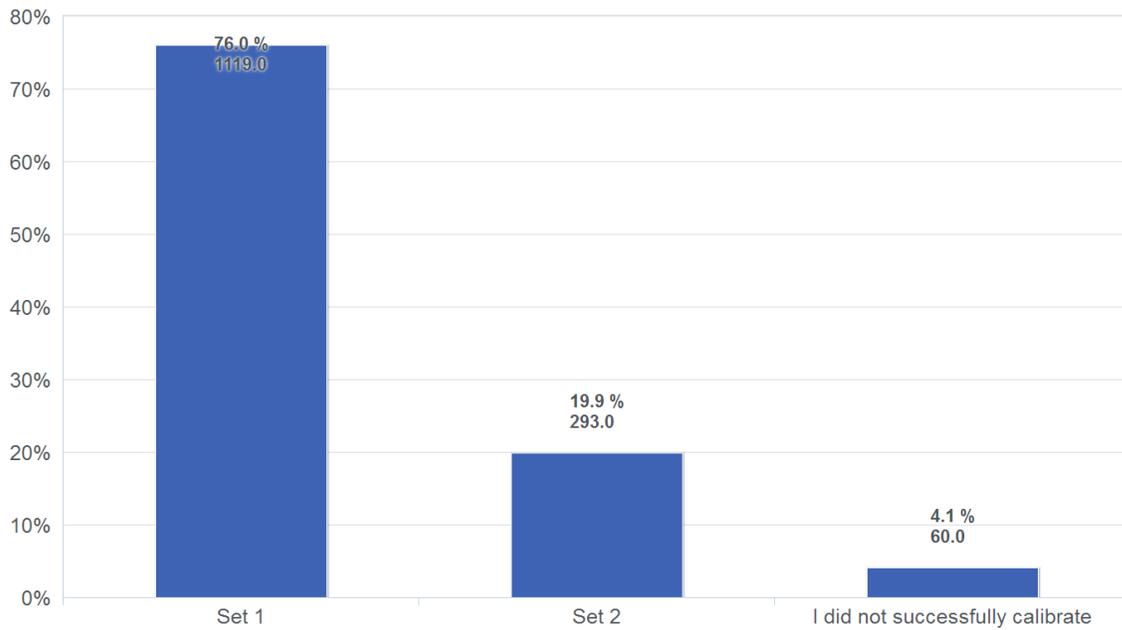
8. If you are a new rater and took the online basic training course this year, did the training content adequately prepare you to rate students?



9. Did the calibration activities adequately prepare you to rate students?



10. On which set did you successfully calibrate?



11. If you did not successfully calibrate, did you receive rating support?

