Texas Education Agency

House Bill 1164

Texas Writing Pilot Program

Report to the Governor and the Texas Legislature

October 23, 2018

TABLE OF CONTENTS

Texas Education Agency

EXECUTIVE SUMMARY

Overview

The Texas Writing Pilot was structured to study a more robust, portfolio-style writing assessment, to meaningfully integrate summative assessment into daily instruction. The study included the collection and scoring of a range of student writing samples produced throughout the school year. This included two timed writing samples, two instructional writing process samples from different genres, and an instructional portfolio containing these writing samples. This portfolio assessment pilot is a formative series with the intent of supporting instruction throughout the school year. The aggregated results of the formative assessment series could then be combined to demonstrate student growth over a school year, as a measure of summative achievement.

Results

The data from the Texas Writing Pilot suggests the following conclusions:

- Scoring correlations and rater-agreement never reached the same level as STAAR, at scale.
- There was inconsistency between Year 1 and Year 2 due to adjustments in the pilot, as would be expected in the development of any new assessment.
- Appropriations to the project derived from STAAR savings supported the initial development of materials and implementation. Limited appropriations to the project reduced the ability for true piloting of a standardized assessment prototype, including possible variables related to training, scoring, and tools used.
- Teachers reported more intentional and focused writing instruction because of the Texas Writing Pilot and generally felt that the prompts were an authentic assessment tool.
- Teachers reported stronger student engagement in their writing instruction.

Recommendations

The Texas Writing Pilot provided the opportunity to begin an investigation into alternative forms of writing assessment in the state. This work contributed to the following recommendations:

- Materials should be freely available to local education agencies (LEAs) and teachers.
- The Texas Education Agency (TEA) should continue to explore options for what authentic writing assessment could look like, pending appropriations and statute. This would include continuing to investigate the inclusion of automated scoring of writing samples to ensure minimum validity and reliability in scoring. Research suggests that computers can adequately evaluate four of the six recognized traits of writing. Preliminary conversations addressed the possibility of combined automated and human scores, which could be explored in later iterations.

While the Texas Writing Pilot was not able to validate the creation of an alternative writing assessment as outlined, the pilot reflected improved writing instruction. Educators indicated they experienced a more intentional instruction methodology and a more thorough integration of the writing standards throughout the year. The professional development offered through the pilot enhanced teachers' understanding of the Texas Essential Knowledge and Skills (TEKS) and promoted writing throughout the year. It further demonstrates that when adequate resources, time, and training are provided, assessment can be meaningfully incorporated into routine instructional practices so that student learning is reflected accurately and in alignment with classroom instruction.

PROJECT OVERVIEW

As required by House Bill (HB) 1164, 84th Texas Legislature, 2015, TEA has conducted a pilot study during the 2016–2017 and 2017–2018 school years to examine alternative methods of assessing writing.

The pilot study included the collection and scoring of a range of student writing samples produced throughout the school year. The writing products completed, submitted, and scored were:

- two timed writing samples completed at the beginning and end of the school year based on a specific writing prompt chosen by each student from a selection of three prompts;
- two instructional writing process samples from different genres—personal narrative, expository, persuasive, or analytic—that include evidence of a writing process from start to finish (e.g., planning, drafting, revising, editing, and publishing); and
- an instructional portfolio containing the writing samples listed above.

Scoring of the student writing samples consisted of several components. Each student's teacher of record initially scored the student samples. Additionally, the samples received a second blind score. The blind scoring included local teachers who were certified to teach reading language arts (RLA). This second round of scoring was coordinated at the local level by participating Education Service Centers (ESCs). Finally, TEA and its contractor, Educational Testing Service (ETS), pulled a sample of the student's writing and provided a third score.

The Texas Writing Pilot assessed writing in grade 4, grade 7, English I, and English II. Similar to other writing portfolio assessment models, the pilot demonstrated that authentic student assessment creates more engagement from students. It also provided educators with the ability to adjust and improve writing instruction with fluidity based upon consistent student evaluations. This report details the pilot's design, educator feedback on varied experiences, and data evaluation for reliability of scoring. Notable work for year two of the pilot are listed below.

- Pilot participation significantly increased from about 1,700 in year one to over 30,000 students in year two.
- Pilot participation included grade 4 students who were assessed in Spanish.
- There were 596 educators recruited in spring 2018 for blind scoring of student samples.
- TEA created and piloted a calibration model for raters to supply standardized training and rigor for accuracy of scoring.
- An interactive online platform and innovative communication avenues using technology promoted efficiency of assignment completion, refinement of performance, and collaboration of leadership.

The correlations and rater-agreement of scoring never reached the same level as STAAR, at scale. While there were some sporadic highlights across the population in both Year 1 and Year 2, the overwhelming

variance in data suggests that training enough educators to be standardized scorers would not be possible. This is generally consistent with the broader literature base on inter-rater reliability and mass scoring. In particular:

- **mean rater scores** varied across tests and categories.
- the **percentage of exact agreement** between raters ranged from a low of 28% to a high of 65%.
- the **percentage of adjacent agreement** between raters ranged from a low of 72% to a high of 99%.
- the **correlations between pilot scores and STAAR scores** were low to medium.
- the **percentage of exact agreement** between raters was greatest between Trained Rater 1 and 2 in most cases.

Analysis of the available data, policies, and operational narratives has been synthesized to incorporate the following supportive recommendations for the Texas Legislature.

- Materials from the Texas Writing Pilot should be produced for LEAs and teachers to use for free through the interim assessment portal and through Texas Gateway. These resources should include the rubric, online training materials (modules, documents, and videos), calibration activities, sample annotated student writing, and an implementation guide. This positive outcome will support teachers in transitioning to the use of a meaningful assessment.
- TEA should continue to explore options for what authentic writing assessment could look like, and the impact of strong reading and writing instruction when paired with authentic writing assessments.
- Pending the availability of resources appropriated for the purpose, TEA should begin investigating the inclusion of automated scoring of writing as a way to ensure minimum validity and reliability in scoring, and also control for the costs of implementing a statewide, authentic writing assessment.

Texas Education Agency

YEAR ONE OVERVIEW

YEAR-ONE PARTICIPATION REQUIREMENTS

For year one, three ESCs were selected to participate with a total of seven partnering independent school districts (ISDs). Region 6 (Huntsville) partnered with Calvert ISD and Huntsville ISD. Region 10 (Richardson) partnered with Athens ISD, Garland ISD, and Sunnyvale ISD. Region 16 (Amarillo) partnered with Amarillo ISD and Dumas ISD. In total, 37 teachers and 1,707 students in grade 4, grade 7, English I, and English II participated in year one of the Texas Writing Pilot.

The 2016–2017 school year began with RLA representatives from the partnering ESCs attending a kick-off planning session with TEA and ETS in Austin. Once the writing pilot rubric was established, a companion scoring training was developed to introduce participating teachers to using the rubric to assess student writing. TEA and ETS then facilitated a virtual train-the-trainer session for the three regional ESC representatives who, in turn, held in-person scoring trainings for the participating teachers in their region.

YEAR-ONE COMMUNICATIONS

Communication and collaboration were high priorities during year one. Representatives from TEA, ETS, and ESCs met weekly to plan and monitor pilot program activities. In addition to the weekly meetings, both TEA and ETS were available for one-on-one support to any ESC, district, or teacher who needed assistance. In this collaborative method, a series of ongoing resources were developed.

YEAR-ONE WRITING SAMPLES

To establish a baseline of student writing, Timed Writing Sample 1 (TS1) was assigned. Students were given an in-class timed writing assignment and had the opportunity to choose from three prompts. While there was a time restriction (see table below), there was no length restriction. Students were free to write as much as they wanted within the given time limit. TS1 was collected at the end of September 2016.

GRADE/COURSE	TIME LIMIT
Grade 4	35 minutes
Grade 7	45 minutes
English I and English II	60 minutes

During the fall and spring semesters, teachers worked on the instructionally based writing process samples with their students. The three process samples—Process Sample 1 (PS1), Process Sample 2 (PS2), and Process Sample 3 (PS3)—were assigned and collected according to the appropriate grade-level genres outlined in the TEKS.

Teachers were provided with designated timeframes and submission windows for assigning and collecting each of the three writing-process samples. Participating districts and teachers could choose the writing genre to collect during each submission window. Submission windows and choice of genre gave teachers the flexibility to fully align the assessment with local instruction and scope and sequence of curriculum. In addition, to better support districts in their writing instruction scope and sequence, a decision was made mid-year by the pilot leadership team to collect two rather than three writing-process samples. These untimed samples were evidence of the student's writing process—planning, drafting, revising, editing, and publishing.

Timed Writing Sample 2 (TS2) was assigned during the last two weeks of April 2017. Students were given a choice of three prompts and the same time allotment and genre as TS1. Both timed samples (TS1 and TS2), as well as the writing-process samples, were compiled into a student's writing portfolio.

YEAR-ONE MATERIALS AND COLLECTION

Classroom teachers scored the writing pilot samples at varying times throughout the school year using the holistic writing pilot rubric (see Appendix A). With the writing pilot rubric, classroom teachers scored the students' TS1 assignments, the final copy of the writing-process samples, and TS2 assignments upon completion in accordance with the scoring deadlines. All teacher-of-record scores, along with student samples, were submitted throughout the year and stored in the secure writing pilot database.

Year-one student samples were collected and housed according to the decision of each local district. Some teachers asked their students to work on a computer for their assignments while others asked their students to complete the assignments on paper. All samples to be scored for year one were periodically uploaded throughout the year to a secure online database where they could be accessed for blind scoring and TEA scoring.

YEAR-ONE SCORING

Blind scoring is a type of scoring in which no rater had access to any score from other raters. Blind scoring sessions for writing samples were held in June 2017. During the blind scoring sessions, all students' writing samples and portfolios were scored at the local regional level by teachers certified to teach RLA. Each of the three participating ESCs recruited teachers within their respective regions for the blind scoring. Each regional blind scoring session consisted of three full days. Over the course of the three days, teachers at each regional session scored a random sample of the statewide writing pilot samples and portfolios. All teacher raters completed end-of-scoring-session evaluation surveys providing input on their scoring experience.

A sampling of the writing samples was scored by ETS on behalf of TEA during the last week of June 2017. ETS recruited Texas-based experienced raters who were certified for scoring the State of Texas Assessments of Academic Assessments (STAAR®). An ETS RLA assessment specialist involved with the writing pilot trained the raters using the same materials and training time used by the ESCs.

YEAR-ONE DATA ANALYSIS

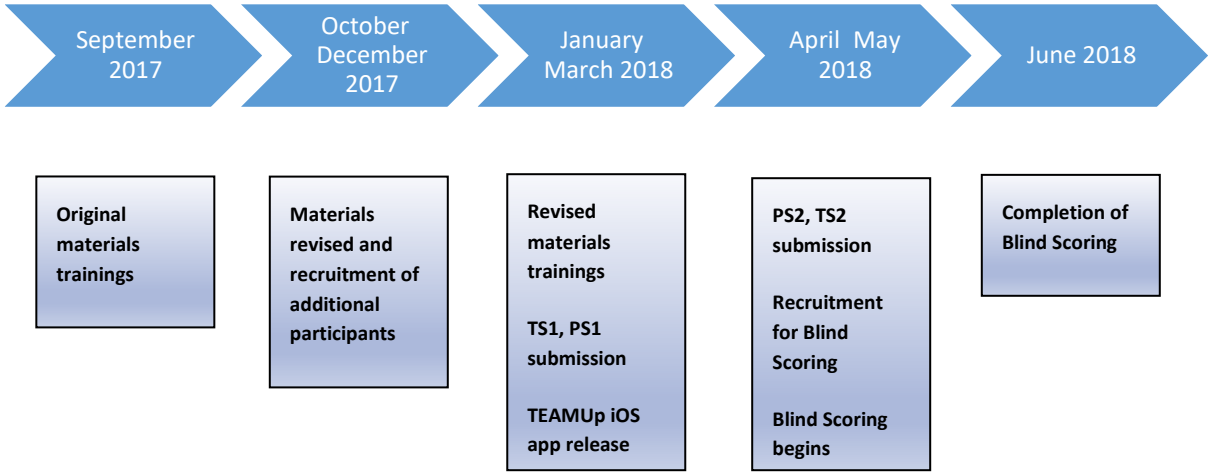
The year-one analysis showed that across all four writing samples and rater pairs:

- the mean correlations over the rating scores were between 0.37 and 0.58;
- the mean percentages of exact agreement over the rating scores ranged from 39% to 47% (compared with 57% to 60% for STAAR);
- the mean percentages of exact or adjacent agreement over the rating scores ranged from 87% to 94%; and
- the maximum correlation, exact agreement rate, and exact or adjacent agreement rate across the rating scores were 0.69, 61%, and 100%, respectively.

The maximum correlation and exact agreement rate for a class across all subjects, rater pairs, and rating scores were 0.88 and 68%, respectively.

YEAR TWO OVERVIEW

Based on feedback from year one of the Texas Writing Pilot, the scope and processes for year two were increased and improved. The progression of activities for year two of the writing pilot are shown below.



YEAR-TWO PARTICIPATION REQUIREMENTS

The enrollment goal for year two was to double student participation counts to 3,500 by August 2017. Based on the recommendation of participating ESCs, positive public response, and the language of the statute, TEA opened enrollment in December 2017 to a wider audience. The requirement for district participation included submission of pertinent campus information for communications and participant counts. A letter from each current and proposed district signed by both district- and campus-level administration that assured the following was also required.

- A portfolio method of assessment embedded in classroom writing instruction will be compatible with the school's current writing instructional practices.
- There will be participation and support from district- and campus-level administration, including testing and curriculum coordinators, for all aspects of the writing pilot program.
- The district and participating campuses have the technological capacity to commit to an online platform for the submission of student samples.

ESCs were required to sign a letter of continued support for year two that assured the following.

- The ESC will support all pilot activities, including hosting pilot events and supporting any required professional development.
- There will be an institution of higher education (IHE) partner that will work with the ESC to support writing pilot activities.

Enrollment closed on January 12, 2018 with a significant increase in the number of students—from about 3,500 to over 50,000. However, pilot information was clarified, requiring Public Education Grant (PEG), Improvement Required (IR), and Focus campuses to administer STAAR writing so those test scores could be used as an accountability measure. As a result, participation counts leveled to about 30,000 in February 2018. Participation numbers for year two of the writing pilot are listed below. Specific campuses involved in the writing pilot are listed in Appendix B.

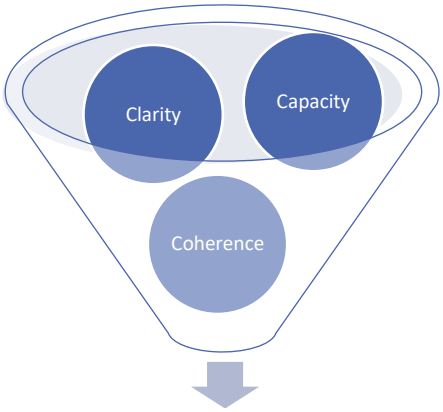
PARTICIPATION CATEGORY	PARTICIPATION NUMBERS
Region	16
District	67
Campus	233
Grade 4	*15,193
Grade 7	11,559
English I	1,985
English II	1,673
Total Number of Students	30,410

*This number includes 724 Spanish writing students

In Year 2, the increase in the number of participants led to delays in implementation, so the full integration into instruction could not occur. For the purposes of data analysis, the study focuses on those who participated over both years. To provide inclusion of grade 4 students who take Spanish writing assessments, TEA partnered with Grand Prairie ISD in January 2018. TEA and ETS conducted an on-site training session at Grand Prairie High School on May 30 and 31, 2018. TEA English Learner (EL) specialists translated the analytic rubric and other pilot materials into Spanish for use in year two of the pilot. EL specialists and educators from Grand Prairie ISD were trained on the new analytic rubric. They were also trained on scoring in the TEA Measuring Upward Progress (TEAMUp) online platform. Raters adjusted the samples to view only grade 4 Spanish responses and provided numeric scores to the responses, as well as commentary on the analytic rubric translation.

YEAR-TWO COMMUNICATIONS

Because of the influx of participating ESCs and districts in January 2018, an immediate need arose to streamline communications from TEA to participating districts. ESCs played a key role in effective communication. Communication best practices—Clarity of Role, Capacity to Provide Support, and Coherence of Responsibility—were implemented to maintain successful communication among TEA, ESCs, and participating districts.



Communication Best Practices

- Clarity of Role—key stakeholders know what role they play.
  The writing pilot achieved greater levels of success when key stakeholders played the following roles.
  - TEA managed the pilot study to ensure the completion of legislative requirements and worked with ETS to develop materials and trainings.
  - ESCs designated an RLA specialist who served as a liaison between the TEA and districts.
  - District testing coordinators (DTCs) interacted with ESC points of contact to disseminate information to educators.
  - Educators provided instruction and scoring of students' writing samples.

- Capacity to Provide Support—key stakeholders have the resources necessary to be successful.
  TEA requested ESCs serve a key role in an innovative assessment that grew at a rapid pace. In addition to regular work responsibilities, ESCs provided support to numerous districts during year two of the writing pilot. DTCs and educators were also asked to exceed typical work expectations through stringent timelines and trainings. Purposeful time-management and the understanding of one's ability to provide useful support and guidance was a crucial lesson learned from the state to the local level.

- Coherence of Responsibility—key stakeholders know what duties they must perform.
  The writing pilot achieved greater levels of success when key stakeholders completed the following actions.
  - TEA and ETS created materials and systems, provided training of materials to ESCs, provided training of systems to DTCs and educators, provided technical and instructional support, communicated regularly with ESCs via email, phone, and video-conferencing.
  - ESCs were responsible for transmitting pertinent documents and messaging, supplying materials directly from TEA to DTCs, attending required materials trainings, providing training to the districts, and assisting in the recruitment and training of blind scoring participants.

- DTCs were responsible for delivering educator and student information for enrollment, uploading student and teacher rosters into the online platform, and ensuring timelines for sample uploads and scoring were met.
- Educators followed administration and submission guidelines, attended required rubric and materials training, completed a teacher-of-record survey, ensured writing assessments were administered to students accurately and uploaded or typed directly into the platform, and supplied a teacher of record score.

In addition to implementing communication best practices, TEA used innovative communication tools to aid in effective communication and collaboration. In January 2018, TEA released a Google+ Community for ESCs to access materials and collaborate on trainings. TEA also used Remind 101 announcements for ESCs, district personnel, and educators to receive text message alerts concerning pilot deadlines and scoring reminders. A great amount of effort and planning was required to ensure effective communication among all pilot participants.

YEAR-TWO WRITING SAMPLES

The processes and procedures regarding the writing samples for year two were similar to year one. To establish a baseline of student writing, the first timed sample, TS1, was assigned. Students were given an in-class timed writing assignment and had the opportunity to choose from three prompts. While there was a time restriction (see table below), there was no length restriction. Students were free to write as much as they wanted within the given time limit. TS1 was collected at the end of September 2017 for original enrollees and in February 2018 for additional mid-year enrollees.

GRADE/COURSE	TIME LIMIT
Grade 4	35 minutes
Grade 7	45 minutes
English I and English II	60 minutes

Again, teachers worked on the instructionally based writing process samples with their students. The process samples were assigned and collected according to the appropriate grade-level genres outlined in the TEKS, as well as when campuses enrolled in the pilot.

Teachers were provided with designated timeframes and submission windows for assigning and collecting the writing-process samples (see Appendix C). Participating districts and teachers could choose the writing genre to collect during each submission window. Submission windows and choice of genre gave teachers the flexibility to fully align the assessment with local instruction and scope and sequence of curriculum. These untimed samples were evidence of the student's writing process—planning, drafting, revising, editing, and publishing.

Participating campuses were required to complete the following:

FOR AUGUST 2017 ENROLLEES	FOR JANUARY 2018 ENROLLEES
ESC Support Required	ESC Support Recommended
Timed Sample 1*	Timed Sample 1*
Process Sample 1	Not required
Process Sample 2**	Process Sample 2**
Timed Sample 2	Timed Sample 2
ESC Connection and support through an institution of higher education	ESC Connection and support through an institution of higher education
Writing Samples entered or uploaded into TEAMUp Online Platform	Writing Samples entered or uploaded into TEAMUp Online Platform

* Timed Sample 1—required to show improvement in student writing between two timed samples and necessary for exemption from the STAAR writing assessment requirements under Texas Education Code (TEC), §39.023 (a) and (c). Districts may choose not to submit this sample. However, districts will then be required to participate in the STAAR writing assessments.

** Process Sample 2—student papers must be entered or uploaded into the pilot online platform, TEA Measuring Upward Progress (TEAMUp), to qualify for exemption from the STAAR writing assessment requirements under TEC §39.023 (a) and (c).

Both timed samples (TS1 and TS2), as well as the writing-process sample (PS1 and PS2), were compiled into a student's writing portfolio.

YEAR-TWO MATERIALS AND COLLECTION

Analytic Rubric

After the completion of scoring and data analysis for year one, ESCs expressed the need for TEA and participating ESCs to collaboratively develop a rubric that allowed for a more accurate articulation of writing improvement across domains. ESC representatives and participating educators viewed the year-one holistic rubric as too similar to the STAAR rubric. The rationale from these professionals was founded on the basis that:

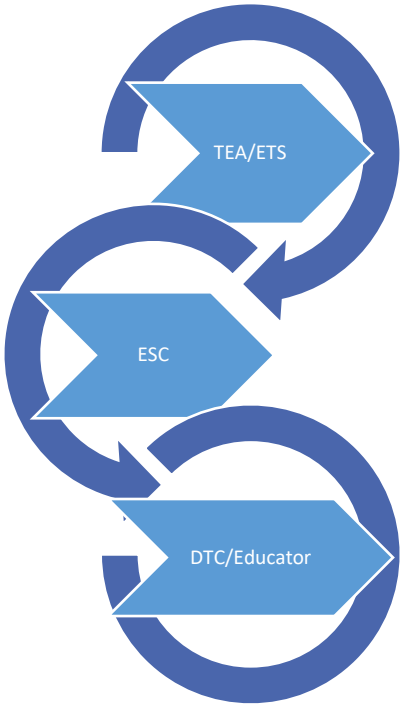
- a portfolio writing assessment has greater instructional value for educators and students through an analytic rubric using performance measurement across domains;
- an analytic rubric would allow for a better understanding and awareness of domain language when scored by several raters; and
- students would accept ownership of performance and improvement through itemized feedback of writing.

The Texas Writing Pilot analytic rubric (see Appendix D) was developed in November 2017 in coordination with updated scoring training materials. The analytic rubric measured organization, content, language, and conventions, but instead of providing an overall holistic score, each of the four

domains were scored individually across six or three levels. The rubric shifted from a 4-category holistic rubric to a 6- or 3-category analytic rubric.

Due to the updated rubric and the amount of time needed to supply a numeric score for 4 domains, TEA decided that a holistic portfolio rubric would not be used for year two. Instead, TEA requested that each teacher of record complete an online survey at the close of the 2017–2018 school year.

ESCs were required to attend trainings provided by TEA and ETS via webinars. ESCs were trained on the implementation of the analytic rubric, how to supply a numeric score, and rationale for the supplied score in the form of annotations. Through a train-the-trainer model, ESCs were then instructed to hold training sessions with educators. Two specific trainings were required: 1) for the teacher of record when he or she entered the pilot program and 2) for the ESC rater to complete blind scoring of student writing samples.



TEA Measuring Upward Progress (TEAMUp) Online Platform

Another improvement for year two was the use of an online platform, TEAMUp, which hosted student samples, scores, and prompts. With TEAMUp, educators no longer had to store paper materials in folders; instead, the system supported students who chose to type and submit their writing samples directly into the online platform. Educators who did not have ready access to computers or chose not to require students to type samples were able to scan handwritten samples and load them into the TEAMUp system.

The online platform also served as a tool for educator scoring and blind scoring, as well as a source on information for the individualized needs of educators and their students. TEAMUp required DTCs to upload educator and student rosters into the platform. Then, educators were responsible for student sample completion and scoring.

Based on initial feedback from pilot participants, ETS launched a TEAMUp application for iOS devices in March 2018. The TEAMUp app allowed educators or students to upload a handwritten sample through the use of a device's camera directly into the platform. The app then assigned the sample to the student through a numeric or QR code.

ETS provided training of the TEAMUp online platform to DTCs via webinar in August 2017 and January 2018. ETS modeled the online system and access features for DTCs, educators, and students. The training also demonstrated how to upload student and educator rosters, type or upload student samples, and supply a score in the system. This training was recorded, and with the analytic rubric, placed within the TEAMUp online platform for DTCs and educators to reference when necessary.

YEAR-TWO SCORING

For educators, the TEAMUp online platform housed an educator dashboard, student rosters, samples, and scoring access features. The TEAMUp system monitored student and educator progress for the submissions and scorings of TS1, PS1, PS2, and TS2. The teacher of record supplied numeric domain scores for each student sample and submitted the scores in the TEAMUp system following the timeline of the Genre Guide and Submission Window document.

A different education professional supplied a second blind score for the students' writing samples. ESCs recruited 596 blind scorers or raters from across the state of Texas to score student samples for the Texas Writing Pilot. The raters consisted of educators, RLA content specialists, and higher education partners. Since raters had different skill sets, TEA determined that a standard calibration set similar to ones completed by professional raters for STAAR would regulate and align raters to the skillset of the teacher of record. Therefore, TEA and ETS created and piloted a nonconsequential calibration model for raters to supply standardized training and rigor for accuracy of scoring,

Blind scoring raters attended a training session held by their ESC the week of April 30 through May 4, 2018. During training, a calibration practice set was completed by raters with the instructions that a nonconsequential calibration must be completed in the TEAMUp system to orient the rater.

Once live scoring began in the TEAMUp online platform, ESC raters completed a calibration set of five student samples. If an ESC rater changed a sample type, the calibration was repeated. If an ESC rater changed a grade level, a new calibration set for the coordinating grade level was then completed.

The raters accessed a simplified dashboard where they chose the grade level and sample type. The raters applied a numeric score for each domain or a "Skip" for purposes of illegibility, off-topic, insufficient, blank, or a cry for help. If a "Skip" was applied to a sample, TEA and ETS assessment analysts

would review the student sample for ratability. DTCs were alerted to a student cry for help writing samples that displayed any troubling content following standard TEA alert paper protocol used for STAAR. Raters were provided feedback by the assessment analysts for other skipped student samples, so a numeric score could be provided.

Raters completed scoring of student samples in the TEAMUp system May 7 through June 1, 2018. With the implementation and capabilities of the TEAMUp system, participating pilot educators were able to score remotely. Each rater was required to complete the scoring of 180 individual student samples to ensure over 90,000 total student samples were successfully scored. TEA, ETS, and ESCs monitored the blind scoring process through weekly scoring reports to ensure deadlines were met. Due to raters' substantial commitment of time and effort, TEA offered 35 hours of Continuing Professional Education (CPE) as compensation to all raters that completed the training and scoring of student samples.

YEAR-TWO SURVEY RESULTS

For year two of the Texas Writing Pilot, TEA requested that each teacher of record complete an online survey at the close of the 2017–2018 school year. The survey asked teachers about the analytic rubric, the TEAMUp system, and the pilot impact. Teachers from both pilot years were included.

Teachers were asked about their experience working with the new analytic rubric.

- 57% indicated that they used the analytic rubric during normal class instruction.
- Teachers stated, for example
    - "I showed them [the students] how their writings would be scored."
    - "I used the rubric to guide my instruction and to explain to students what their goals were."
    - "We discussed the language used on the rubric and the differences between the categories in whole group instruction. When conferencing, the student and I would look to the rubric to discuss what revisions or editing needed to be done to improve. The students needed to understand what was expected from them to know how to succeed."
- Educators who did not use the rubric during class instruction indicated that the language of the rubric was not at an appropriate level for a grade 4 student to use.

The survey indicated that most campuses did not allow students to access the TEAMUp system due to the student's age, access to technology, or a desire to ensure materials were complete and submitted without student upload errors.

- 71% indicated "*Not At All*" when responding to student input into TEAMUp.
- 69% frequently performed a teacher upload.

Educators were asked, "*How did your experience with the Texas Writing Pilot change the way you teach writing in the classroom?*"

- 35% selected, *"It allowed my classes to focus on their quality of writing through the writing process."*
- 20% selected, *"It allowed me to feel better equipped to utilize a writing rubric."*
- 24% selected, *"It allowed me to focus on multiple genres of writing instead of just one."*

Based on the survey results, educators saw the value in the analytic rubric as a tool for instruction and feedback support. The majority of teachers understood what to use for scoring and agreed that tools and trainings were sufficient. Additionally, they had recommendations for future TEAMUp advancements consisting of comprehensive access for DTCs, as well as an application for Android devices.

YEAR-TWO DATA ANALYSIS

The purpose of analyzing pilot data was to evaluate the technical quality of the locally scored writing alternative assessment method (i.e., using students' writing portfolios that were produced in the classroom), with the primary technical challenge being ensuring that the ratings (or scores) of writing samples were comparable in meaning when evaluated in different places, at different times, and by different people. After the completion of writing sample collection and scoring, the data to support the analyses had the following characteristics:

- Four writing samples were planned chronologically across the school year: TS1, PS1, PS2, and TS2.
- The final product of each writing sample in a scored student portfolio received a set of four ratings—organization, content, language, and conventions—from each type of rater.
- Three sets of ratings were independently assigned according to the rubric by three types of raters: 1) the classroom teacher of record ("Teacher"); 2) a rater recruited and trained by the ESC ("ESC"); and 3) a qualified Trained Rater ("TR1").
    - Teachers were provided scoring training and support by DTCs, ESCs, and TEA.
    - ESC raters and qualified trained raters received the same scoring training and support during their organized scoring sessions.
- Additionally, approximately 25% of the students' writing samples received an additional set of ratings from a qualified trained rater (i.e., double-scored with Trained Rater 2, "TR2") for the purpose of studying the quality of ratings assigned by the trained raters.

Writing samples scored by three or four raters were used in the data analyses. Appendix E lists the demographic distribution of the students included in the data analyses. The table below is a summary of the number of students, campuses, and regions, and their scored data to support analyses.

A Summary of Students and Writing Samples in Data Analyses

	Grade 4	Grade 7	English I	English II
Number of Participating Students	13,875	10,298	828	597
Number of Participating Campuses	162	169	13	10
Number of Participating Regions	4	4	3	3
Number of Writing Samples Scored by Three Raters	2,169	1,872	2,248	1,715
Number of Timed Writing Sample 1	603	890	681	469
Number of Process Writing Sample 1	361	383	289	431
Number of Process Writing Sample 2	602	237	597	428
Number of Timed Writing Sample 2	603	362	681	387
Number of Writing Samples Scored by Two Trained Raters	517	506	533	407
Number of Timed Writing Sample 1	124	185	147	102
Number of Process Writing Sample 1	62	139	73	77
Number of Process Writing Sample 2	200	99	165	163
Number of Timed Writing Sample 2	131	83	148	65

The rating quality of the Trained Raters were first evaluated with the double-scored students' writing samples to establish a frame of reference. The ratings by Teachers and by ESC raters were then compared with those produced by the Trained Raters. By way of explanation, the ratings by the Trained Raters were used as the criteria to evaluate how much Teachers and ESC (blind) raters agreed or disagreed with them. In addition to describing the scored data characteristics with summary statistics, other statistics were used to examine the extent to which the ratings assigned by Teachers, ESC raters, and Trained Raters were consistent, as rating reliability indicators. The other statistics consisted of:

- polychoric[1] correlations (COR);
- quadratic weighted kappa coefficients[2] (WK);
- percentages of exact agreement (EA); and
- percentages of exact or adjacent agreement (EAA) between ratings.

Key observations are summarized below, and the detailed analyses methodology and results are presented in Appendix F–L.

- The agreement between the two Trained Raters was higher than the agreement between Teachers and Trained Raters, the agreement between ESC and Trained Raters, or the agreement between Teachers and ESC raters. The two trained raters' scores in general were a little more consistent than the scores from the other rater pairs. The score agreement between Teachers

[1] Drasgow, F. (1988). Polychoric and polyserial correlations. In L. Kotz, & N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*. Vol. 7 (pp. 69-74). New York: Wiley.

[2] Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, 33,* 613–619.

and Trained Raters was the closest to that between the two Trained Raters on English I among the four grades/courses based on all writing samples. The maximum difference on weighted kappa across the four scores between Teachers versus Trained Raters and the two Trained Raters was 0.17 for grade 4 writing, 0.16 for grade 7 writing, 0.06 for English I, and 0.21 for English II.

- Based on all writing samples, Teachers gave the highest average ratings among the three or four raters across ratings and grades/courses except for organization and conventions ratings in English I, where the average ratings of ESC raters were higher than those of Teachers.
- There were some variations on score agreement among raters by grade/course, rating score, writing sample type, writing prompt, or genre. For example, between Teachers and Trained Raters, 1) the process writing samples within the analytic genre had the best agreement across ratings among the different writing genres in grade 4, while they had the worst agreement in English I; 2) the PS1 writing samples had the best agreement across ratings among the four writing samples in English I, while in grade 4 the agreement on TS1 were the best.

In the table below, the agreement statistics of constructed response ratings based on STAAR grades 4 and 7 writing, English I, and English II administered in spring 2018 were used as another frame of reference (top section). For easy comparisons, the ranges of these statistics based on all writing samples from the Texas Writing Pilot were summarized across rating scores and grades/courses (middle section), and the ranges of these statistics based on each writing sample—TS1, PS1, PS2, and TS2—at the teacher level were also summarized across writing samples, rating scores, and grades/courses (bottom section).

Rater Agreement Statistics: Spring 2018 STAAR and Texas Writing Pilot

	N	EA (%)	EAA (%)	COR	WK
STAAR (4-category rubrics)					
Grade 4 Writing	371,894	60	98	0.75	0.66
Grade 7 Writing	388,176	62	98	0.75	0.65
English I	492,315	61	98	0.80	0.72
English II	453,511	60	98	0.78	0.71
Texas Writing Pilot: All Writing Samples (6- or 3-category rubrics)					
Organization; Content; Language (6-category)		28–45	72–87	0.37–0.67	0.31–0.63
Conventions (3-category)		54–65	96–99	0.43–0.66	0.33–0.53
Texas Writing Pilot: Writing Samples by Teacher (6- or 3-category rubrics)					
Organization; Content; Language (6-category)		3–66	23–100	-0.47–0.84	-0.39–0.71
Conventions (3-category)		14–79	52–100	[a]	-0.34–0.65

Note: EA=percentage of exact agreement; EAA=percentage of exact or adjacent agreement; COR=correlation; and WK=weighted kappa with quadratic weights.

[a] Correlation was not calculated for conventions score at the class level because of instability with a small sample size.

The overall variance in the data for the writing pilot indicates a higher variance than is allowable for a standardized assessment.

- Across all ratings, rater pairs, and grades/courses, the agreement statistics of the Texas Writing Pilot ratings with 6-category rubrics were lower than those of the STAAR ratings with 4-category rubrics. The percentages of exact and adjacent agreement of the Texas Writing Pilot ratings (i.e., Conventions) with a 3-category rubric were close to or higher than those of the STAAR ratings with 4-category rubrics, while the correlations and weighted kappa coefficients were still lower.
- In some classes (defined by Teacher), the agreements among Teacher, ESC raters, and Trained Raters were high. The four agreement statistics were calculated at the class level for each writing sample in each grade/course among the ratings from the three raters—the Teacher, the ESC rater, and the Trained Rater—for each class with a sample size of at least 30. These statistics varied widely across classes, grades/courses, writing samples, and raters. However, it is encouraging to observe that the agreements in some classes among Teachers, ESC raters, and Trained Raters were close to or higher than the corresponding STAAR scoring agreement statistics.
- However, the level of agreement referenced above occurred at a low frequency amongst the population and would likely be limited to a small number of overall campuses statewide.

The polyserial correlation was estimated between the scores of each rating and the corresponding spring 2018 STAAR scale scores as external validity indicators for the rating.

- Based on all writing samples, grade 4 pilot students' rating scores across all ratings and raters had low to medium correlations (ranging from 0.25 to 0.54) with their spring 2018 STAAR grade 4 writing scale scores.
- Based on all writing samples, grade 7, English I, and English II pilot students' rating scores across all ratings and raters had medium correlations (ranging from 0.41 to 0.69) with corresponding spring 2018 STAAR scale scores.

The Texas Writing Pilot analysis results show that overall, the Teachers and ESC raters agreed less often with Trained Raters than Trained Raters agreed with each other. However, it should also be noted that the Texas Writing Pilot had impact on Teachers' behaviors in their classrooms (see section on Survey Results). The pilot demonstrated the meaningful integration of instruction and assessment. Further qualitative research, conducted in pilot districts, and engagement with educators involved in the process identifies increased volume and variety of student writing experiences, increased depth of instruction, and student growth as pilot outcomes that should be recognized. The behavioral impact of the Texas Writing Pilot has shown some evidence of stronger writing instruction in the classroom, which could have a benefit in the long-term writing abilities and engagement of students.

CONCLUSIONS AND RECOMMENDATIONS

The data from the Texas Writing Pilot suggests the following conclusions:

- The correlations and rater-agreement of scoring never reached the same level as STAAR, at scale. While there were some sporadic highlights across the population in both Year 1 and Year 2, the overwhelming variance in data suggests that appropriately training enough educators to be standardized scorers would not be possible. This is generally consistent with the broader literature base on inter-rater reliability and mass scoring.
- In order to be reliable, the project would need to be longer and allow for expected and typical adjustments and improvements to training materials, rubrics, guidance, formats and infrastructure, as well as interrater reliability comparisons across the years. The development of a new assessment with known requirements typically requires at least three years. The development of such a robust assessment at the scale of Texas would require even more time and a sizeable appropriation or grant.
- The costs of administering a statewide, authentic writing assessment would be prohibitive, even with donated teacher time. Each LEA would need to donate 25–30 hours per teacher, per year. This would account for training, calibration activities, and scoring. Cost-reduction measures such as the exploration of computer-based scoring, would significantly alleviate cost concerns if the assessment were ever to launch at scale.
- There were a number of possible variables that could have been tested related to training, structuring the study, and creating additional resources. However, no funding was appropriated, and decisions were made to accommodate the resources available.
- Teachers reported more intentional and focused writing instruction because of the Texas Writing Pilot. Further, teachers generally felt that the prompts were a more authentic assessment tool than the current version of STAAR.
- Teachers reported stronger student engagement, as a result of more intentional teaching.

The Texas Writing Pilot provided the opportunity to begin an investigation into alternative forms of writing assessment in Texas. Data collected related to student performance, as well as the implementation of the pilot from educators all contributed to the following recommendations:

- **Free materials for all LEAs to use for instruction.** Materials from the Texas Writing Pilot should be produced for LEAs and teachers to use for free through the interim assessment portal and through Texas Gateway. These resources should include the rubric, online training materials (modules, documents, and videos), calibration activities, sample annotated student writing, and an implementation guide.
- **Continue to explore further options.** TEA should continue to explore options for what authentic writing assessment could look like, and the impact of strong reading and writing instruction when paired with authentic writing assessments.

- **Consider use of additional appropriations.** Pending the availability of resources appropriated for the purpose, TEA should begin investigating the inclusion of automated scoring of writing, as a way to ensure minimum validity and reliability in scoring, and also control for the costs of implementing a statewide, authentic writing assessment.
- **Timely, definitive guidance for implementation.** Determining pilot structures, such as participant selection, writing samples to be collected, metrics to be used for student feedback, and the selection of a system for collecting student work in advance of implementation, allows for proactive planning and communication.
- **Prompt, effective professional development.** While professional development was designed and provided for ESC representatives, district leaders, campus administrators, and campus teachers involved in the statewide pilot study, there is a need for timely training, as well as increased time and depth of training related to the submission of student writing samples and teacher scoring of student responses. Due to the daily expectations of campus administrators and teachers, the provision of training in advance of the school year would allow time for campus personnel to incorporate pilot study expectations into instructional sequences. Increased time and depth of training are recommended to increase teacher knowledge of scoring instruments to increase teacher rating reliability. Providing an opportunity for educators to engage in critical conversations related to scoring will also support increased reliability.
- **Determine sites for continued pilot work.** Use data from the pilot to determine possible sites for continued development of a portfolio-based assessment model and utilize data from teachers with higher exact agreements and correlations. A smaller number of pilot sites would allow for ongoing collaboration regarding implementation and training needs.
- **Consider integrity of multiple assessment model.** If pilot districts are asked to implement both STAAR and a portfolio-based method, TEA should consider the integrity of the portfolio-based assessment and whether districts involved in the pilot are at risk of conflicting instructional practices. TEA should consider submission of a student timed writing sample and/or process sample as their state writing assessment.

While the Texas Writing Pilot was not able to validate the creation of an alternative writing assessment as outlined, the pilot reflected improved writing instruction. Educators indicated they experienced a more intentional instruction methodology and a more thorough integration of the writing standards throughout the year. Although the pilot did not prove to be a valid assessment instrument, it did demonstrate the importance of embedding strong assessment intro instruction, reflecting authenticity in daily classroom activities, and more clearly integrating instruction and the state assessment. The professional development offered through the pilot enhanced teacher understanding of the TEKS and promoted writing throughout the year. It further demonstrates that when adequate financial resources, time, and training are appropriated, assessment can be meaningfully incorporated into routine instructional practices so that student learning is reflected accurately and in alignment with classroom instruction.

APPENDICIES

APPENDIX A: YEAR-ONE TEXAS WRITING PILOT HOLISTIC RUBRIC (2016–2017)

Score Point 4 (Accomplished): The response will contain most of the following characteristics.			
Organizational Structure and Focus	**Content/Development of Ideas**	**Use of Language**	**Conventions**
• Structure is clearly appropriate to the purpose. • The writer establishes and maintains a strong focus. • Strong, meaningful transitions and idea-to-idea, sentence-to-sentence, and paragraph-to-paragraph connections are clearly evident.	• Specific, well chosen, and relevant details are clearly evident. • Ideas are clearly, thoughtfully, and effectively expressed and developed.	• Language and word choice are purposeful, precise, and enhance the writing. • Sentences are purposeful, well-constructed, and controlled. • Use of an authentic, expressive voice is clearly reflected throughout the writing.	• Although minor errors may be evident, they do not detract from the fluency or clarity of the writing. • Use of grade-appropriate spelling, capitalization, punctuation, grammar, and usage conventions is consistently demonstrated.
Score Point 3 (Satisfactory): The response will contain most of the following characteristics.			
Organizational Structure and Focus	**Content/Development of Ideas**	**Use of Language**	**Conventions**
• Structure is, for the most part, appropriate to the purpose. • The writer, for the most part, establishes and maintains focus. • Sufficient use of transitions and idea-to-idea, sentence-to-sentence, and paragraph-to-paragraph connections is somewhat evident.	• Specific, appropriate, and relevant details are somewhat evident. • Ideas are sufficiently expressed and developed.	• Language and word choice are, for the most part, clear, concise, and somewhat enhance the writing. • Sentences are somewhat purposeful and adequately constructed and controlled. • Authentic voice is somewhat evident and appropriately reflected throughout the writing.	• Minor errors create some disruption in the fluency or clarity of the writing. • Use of grade-appropriate spelling, capitalization, punctuation, grammar, and usage conventions is adequately demonstrated.

<u>Score Point 2 (Basic)</u>: The response will contain most of the following characteristics.			
Organizational Structure and Focus	**Content/Development of Ideas**	**Use of Language**	**Conventions**
• Structure is evident but may not always be appropriate to the purpose. • The writer does not effectively establish or maintain focus and may include irrelevant information. • Use of transitions, idea-to-idea, sentence-to-sentence, and paragraph-to-paragraph connections is minimal or inconsistent.	• Specific and relevant details are too brief, too vague, or are not clearly evident. • Ideas are minimally expressed and developed.	• Language and word choice are general, imprecise, or inappropriate and do not sufficiently enhance the writing. • Sentences are awkward or only somewhat controlled. • Authentic voice is inconsistent throughout the writing.	• Distracting errors create moderate disruptions in the fluency or clarity of the writing. • Use of grade-appropriate spelling, capitalization, punctuation, grammar, and usage conventions is partially demonstrated.
<u>Score Point 1 (Very Limited)</u>: The response will contain most of the following characteristics.			
Organizational Structure and Focus	**Content/Development of Ideas**	**Use of Language**	**Conventions**
• Structure is inappropriate to the purpose. • Focus is not established or maintained. • Transitions, idea-to-idea, sentence-to-sentence, and paragraph-to-paragraph connections are not evident.	• Details are inappropriate or missing. • Ideas are missing or not expressed or developed.	• Language and word choice is limited or missing and does not enhance the writing. • Sentences are simplistic or uncontrolled. • Authentic voice is missing or inappropriate to the writing task.	• Serious and persistent errors create disruptions in the fluency or clarity of the writing. • Little to no use of grade-appropriate spelling, capitalization, punctuation, grammar, and usage conventions is demonstrated.

APPENDIX B: YEAR-TWO PARTICPANTS

Below is a list of the regions, districts, and campuses that completed year two of the Texas Writing Pilot.

ESC	DISTRICT	CAMPUS
REGION 3		
3	Woodsboro ISD	Woodsboro Elementary
REGION 4		
4	Aldine ISD	Aldine Elementary
4	Aldine ISD	Aldine High School
4	Aldine ISD	Smith Elementary
4	Aldine ISD	Thompson Elementary
4	Aldine ISD	Hambrick Middle School
4	Aldine ISD	Shotwell Middle School
4	Aldine ISD	Davis 9th Grade School
4	Aldine ISD	MacArthur High School
4	Barbers Hill ISD	Barbers Hill Elementary School North
4	Barbers Hill ISD	Barbers Hill Elementary School South
4	Barbers Hill ISD	Barbers Hill Middle School North
4	Barbers Hill ISD	Barbers Hill Middle School South
4	Columbia-Brazoria ISD	Wild Peach Elementary
4	Columbia-Brazoria ISD	West Columbia Elementary
4	Columbia-Brazoria ISD	Barrow Elementary
4	Columbia-Brazoria ISD	West Brazos Junior High
4	Deer Park ISD	W.A. Carpenter Elementary
4	Deer Park ISD	J.P. Dabbs Elementary
4	Deer Park ISD	Deepwater Elementary
4	Deer Park ISD	Deer Park Elementary
4	Deer Park ISD	Fairmont Elementary
4	Deer Park ISD	San Jacinto Elementary
4	Klein ISD	Bernshausen Elementary
4	Klein ISD	Eiland Elementary
4	Klein ISD	Epps Island Elementary
4	Klein ISD	Greenwood Forest Elementary
4	Klein ISD	Kaiser Elementary
4	Klein ISD	Nitsch Elementary
4	Klein ISD	McDougle Elementary
4	Klein ISD	Mittelstadt Elementary
4	Klein ISD	Wunderlich Intermediate
4	Klein ISD	Klein Intermediate
4	Spring Branch ISD	Cedar Brook Elementary
4	Spring Branch ISD	Thornwood Elementary
REGION 5		
5	Kirbyville CISD	Kirbyville Elementary
5	Kountze ISD	Kountze Intermediate
REGION 6		
6	Calvert ISD	Calvert School

ESC	DISTRICT	CAMPUS
6	Splendora ISD	Greenleaf Elementary
6	Splendora ISD	Peach Creek Elementary
6	Splendora ISD	Piney Woods Elementary
6	Splendora ISD	Splendora Junior High
7	Fruitvale ISD	Hallie Randall Elementary
7	Fruitvale ISD	Fruitvale Middle School
7	Hawkins ISD	Hawkins Elementary
7	Hawkins ISD	Hawkins Middle School
7	Longview ISD	Judson STEAM Academy
7	Longview ISD	South Ward Elementary
7	Mineola ISD	Mineola Middle School
7	Quitman ISD	Quitman Elementary
7	Quitman ISD	Quitman Junior High School
7	Tyler ISD	Bell Elementary
7	Tyler ISD	Birdwell Elementary
7	Tyler ISD	Clarkston Elementary
7	Tyler ISD	Owens Elementary
7	Tyler ISD	Rice Elementary
7	Westwood ISD	Westwood Primary
7	Westwood ISD	Westwood Elementary
8	Jefferson ISD	Jefferson Elementary
8	Maud ISD	Maud Elementary
8	New Boston ISD	Crestview Elementary
8	New Boston ISD	New Boston Middle School
8	New Boston ISD	New Boston High School
9	Wichita Falls ISD	Cunningham School
9	Wichita Falls ISD	Southern Hills Elementary
9	Wichita Falls ISD	Fain Elementary
10	Athens ISD	Central Athens School
10	Athens ISD	South Athens Elementary
10	Athens ISD	Bel Air Elementary
10	Athens ISD	Athens Middle School
10	Blue Ridge ISD	Blue Ridge Elementary
10	Celeste ISD	Celeste Elementary
10	Crandall ISD	Barbara Walker Elementary
10	Crandall ISD	Nola Kathryn Wilson Elementary
10	Crandall ISD	W.A. Martin Elementary
10	Crandall ISD	Hollis T. Dietz Elementary
10	Crandall ISD	Crandall Middle School
10	Frisco ISD	Christie Elementary
10	Frisco ISD	Scott Elementary

ESC	DISTRICT	CAMPUS
10	Frisco ISD	Shawnee Trail Elementary
10	Frisco ISD	Miller Elementary
10	Garland ISD	Bradfield Elementary
10	Garland ISD	Club Hill Elementary
10	Garland ISD	Shorehaven Elementary
10	Garland ISD	Williams Elementary
10	Garland ISD	Kimberlin Academy
10	Garland ISD	Sellers Middle School
10	Garland ISD	Lyles Middle School
10	Garland ISD	Sam Houston Middle School
10	Garland ISD	O'Banion Middle School
10	Grand Prairie ISD	Stephen F. Austin Elementary
10	Grand Prairie ISD	James Bowie Elementary
10	Grand Prairie ISD	David Daniels Elementary Academy of Science & Math
10	Grand Prairie ISD	Florence Hill Elementary
10	Grand Prairie ISD	Ellen Ochoa STEM Academy at Ben Milam Elementary
10	Grand Prairie ISD	Dwight D. Eisenhower Elementary
10	Grand Prairie ISD	Sam Rayburn Elementary STEAM Academy
10	Grand Prairie ISD	Lorenzo De Zavala Environmental Science Academy
10	Grand Prairie ISD	Suzanna Dickinson Elementary
10	Grand Prairie ISD	Garner Fine Arts Academy
10	Grand Prairie ISD	Barbara Bush Elementary
10	Grand Prairie ISD	Colin Powell Elementary
10	Grand Prairie ISD	Hector P. Garcia Elementary
10	Grand Prairie ISD	Sallye R. Moore Elementary
10	Grand Prairie ISD	Ervin C. Whitt Elementary
10	Grand Prairie ISD	Juan N. Seguin Elementary
10	Grand Prairie ISD	Thurgood Marshall Leadership Academy
10	Grand Prairie ISD	Mike Moseley Elementary
10	Grand Prairie ISD	Hobbs Williams Elementary
10	Grand Prairie ISD	Robert E. Lee Elementary
10	Grand Prairie ISD	School for the Highly Gifted
10	Grand Prairie ISD	William B. Travis World Language Academy
10	Grand Prairie ISD	Grand Prairie Fine Arts Academy
10	Grand Prairie ISD	Grand Prairie Collegiate Institute
10	Grand Prairie ISD	John Adams Middle School
10	Grand Prairie ISD	Andrew Jackson Middle School
10	Grand Prairie ISD	Harry S. Truman Middle School
10	Grand Prairie ISD	Ronald W. Reagan Middle School
10	Grand Prairie ISD	James Fannin Middle School
10	Grand Prairie ISD	YMLA at John F. Kennedy Middle School
10	Grand Prairie ISD	YWLA at Bill Arnold
10	Prosper ISD	John A. Baker Elementary
10	Prosper ISD	Cynthia Cockrell Elementary
10	Prosper ISD	Judy Rucker Elementary
10	Prosper ISD	Steve Folsom Elementary
10	Prosper ISD	Light Farms Elementary

ESC	DISTRICT	CAMPUS
10	Prosper ISD	Windsong Ranch Elementary
10	Prosper ISD	Jim and Betty Hughes Elementary
10	Prosper ISD	Reynolds Middle School
10	Prosper ISD	Lorene Rogers Middle School
10	Royse City ISD	Royse City Middle School
10	Royse City ISD	Anita Scott Elementary
10	Royse City ISD	William R. Fort Elementary
10	Royse City ISD	Miss May Vernon Elementary
10	Sunnyvale ISD	Sunnyvale Elementary
10	Sunnyvale ISD	Sunnyvale Middle School
10	Van ISD (through region 7)	Van Middle School
10	Van ISD (through region 7)	Van Junior High
REGION 11		
11	Burleson ISD	Academy at Nola Dunn
11	Burleson ISD	Mound Elementary
11	Burleson ISD	Norwood Elementary
11	Burleson ISD	Jack Taylor Elementary
11	Burleson ISD	William Stribling Elementary
11	Burleson ISD	Bransom Elementary
11	Burleson ISD	Judy Hajek Elementary
11	Burleson ISD	Ann Brock Elementary
11	Burleson ISD	Irene Clinkscale Elementary
11	Burleson ISD	Hughes Middle School
11	Burleson ISD	Nick Kerr Middle School
11	Burleson ISD	STEAM Middle School
11	Cleburne ISD	A.D. Wheat Middle School
11	Cleburne ISD	Coleman Elementary
11	Cleburne ISD	Marti Elementary
11	Cleburne ISD	Gerard Elementary
11	Eagle Mountain-Saginaw ISD	Bryson Elementary
11	Eagle Mountain-Saginaw ISD	Gililland Elementary
11	Eagle Mountain-Saginaw ISD	Willow Creek Elementary
11	Eagle Mountain-Saginaw ISD	Ed Willkie Middle School
11	Eagle Mountain-Saginaw ISD	Wayside Middle School
11	Godley ISD	Godley Intermediate School
11	Godley ISD	Godley Middle School
11	Granbury ISD	Acton Elementary
11	Granbury ISD	Brawner Intermediate School
11	Granbury ISD	Mambrino School
11	Granbury ISD	Nettie Baccus Elementary
11	Granbury ISD	Oak Wood School
11	Granbury ISD	Acton Middle School
11	Granbury ISD	Granbury Middle School
REGION 12		
12	Malone ISD	Malone Elementary
REGION 13		

ESC	DISTRICT	CAMPUS
13	Dripping Springs ISD	Sycamore Springs Elementary
13	Dripping Springs ISD	Rooster Springs Elementary
13	Dripping Springs ISD	Walnut Springs Elementary
13	Dripping Springs ISD	Dripping Springs Elementary
13	Eanes ISD	Valley View Elementary
13	Eanes ISD	Forest Trail Elementary
13	Eanes ISD	Eanes Elementary
13	Eanes ISD	Cedar Creek Elementary
13	Eanes ISD	Bridge Point Elementary
13	Eanes ISD	Barton Creek Elementary
REGION 14		
14	Moran ISD	Moran School
14	Roscoe Collegiate ISD	Roscoe Elementary
14	Roscoe Collegiate ISD	Roscoe Collegiate High School
REGION 15		
15	Winters ISD	Winters Elementary
15	Winters ISD	Winters Junior High
REGION 16		
16	Amarillo ISD	Mesa Verde
16	Amarillo ISD	Humphries Highland
16	Amarillo ISD	Tascosa High School
16	Amarillo ISD	Palo Duro High School
16	Amarillo ISD	Travis Middle School
16	Borger ISD	Borger Middle School
16	Dalhart ISD	Dalhart Jr. High School
16	Dimmitt ISD	Dimmitt Middle School
16	Dumas ISD	Dumas High School
16	Kress ISD	Kress Elementary
16	Kress ISD	Kress Jr and Sr High School
16	Lefors ISD	Lefors School
16	Memphis ISD	Memphis High School
16	Memphis ISD	Memphis Middle School
16	Memphis ISD	Austin Elementary
16	Panhandle ISD	Panhandle Elementary
16	Plemons-Stinnett-Phillips CISD	West Texas Elementary
16	Plemons-Stinnett-Phillips CISD	West Texas Middle School
16	River Road ISD	Rolling Hills Elem/River Road Middle School
16	Spring Creek ISD	Spring Creek School
16	Sunray ISD	Sunray Elementary
16	Sunray ISD	Sunray Middle School
REGION 19		
19	Dell City ISD	Dell City School
19	Sierra Blanca ISD	Sierra Blanca School
REGION 20		
20	East Central ISD	Heritage Middle School

ESC	DISTRICT	CAMPUS
20	Jubilee Academy	Jubilee-Lake View University Prep, Jubilee Academies
20	Kerrville ISD	Tom Daniels Elementary School
20	Kerrville ISD	Starkey Elementary School
20	Northside ISD (015915)	Leon Valley Elementary
20	Northside ISD (015915)	Oak Hills Terrace Elementary
20	North East ISD	Garner Middle School
20	North East ISD	Larkspur Elementary
20	North East ISD	Ridgeview Elementary
20	School of Excellence in Education	Dr. Harmon Kelley Elementary
20	School of Excellence in Education	Dr. David Walker Elementary
20	School of Excellence in Education	Dr. Paul Saenz Junior High
20	School of Excellence in Education	Milton B. Lee Academy
20	Southwest ISD	Hidden Cove Elementary
20	Southwest ISD	Sun Valley Elementary
20	Southwest ISD	Spicewood Park Elementary
20	Southwest ISD	Bob Hope Elementary
20	Southwest ISD	Indian Creek Elementary
20	Southwest Preparatory School	SPS-Northwest
20	Southwest Preparatory School	Southwest Preparatory School
20	Southwest Preparatory School	SP Southeast Campus
20	Southwest Preparatory School	SP Northwest Elementary
20	Southwest Preparatory School	New Directions
20	Southwest Preparatory School	Seguin Elementary

APPENDIX C: YEAR-TWO GENRE GUIDE AND SUBMISSION WINDOWS

The Genre Guide and Submission Windows document served as a guideline for districts and regions to follow. Dates were extended and amended in January 2018, for mid-year additions to the Texas Writing Pilot.

Sample	4th Grade	7th Grade	English I	English II
Timed Sample I	Assignment window: **09/18/17–10/06/17** Mid-Year Additions: **1/29/18-2/16/18 (EXTENDED to 2/23/18)** Upload deadline: **10/06/17** Mid- Year Additions: **2/16/18 (EXTENDED to 2/23/18)** Scores entered deadline: **10/27/17** Mid- Year Additions: **3/9/18**			
Genre	**Personal Narrative**	**Expository**	**Expository**	**Persuasive**
Process Sample 1	Assignment window: **1/29/18-2/23/18 (UPDATED)Mid- Year Additions NOT APPLICABLE** Upload deadline: **2/23/18** Scores entered deadline: **3/9/18**			
Genre	**District Choice* Persuasive, Expository, or Analytic**	**District Choice* Persuasive, Personal Narrative, or Analytic**	**District Choice* Persuasive, Personal Narrative, or Analytic**	**District Choice* Expository, Personal Narrative, or Analytic**
Process Sample 2	Assignment window: **02/05/18 – 3/30/18** Upload deadline: **03/30/18** Scores entered deadline: **04/27/18**			
Genre	**District Choice* Persuasive, Expository, or Analytic**	**District Choice* Persuasive, Personal Narrative, or Analytic**	**District Choice* Persuasive, Personal Narrative, or Analytic**	**District Choice* Expository, Personal Narrative, or Analytic**
Timed Sample II	Assignment window: **03/28/18-4/27/18** Upload deadline: **04/30/18** Scores entered deadline: **05/18/18**			
Genre	**Personal Narrative**	**Expository**	**Expository**	**Persuasive**
***The genre selected for process samples should be different for each sample and should follow an individual district's scope and sequence.**				

Below are possible pieces of writing process evidence to include as part of the Process Samples in students' portfolios. These items should follow the natural writing process that teachers already use in classroom instruction.

	4th Grade	7th Grade	English I	English II
Possible Evidence for Process Sample 1 and 2	**Prewriting:** Brainstorm, web, graphic organizer, journal entry, etc. **Drafting, Revising and Editing:** First draft, teacher conference form, peer editing forms, self-editing forms, secondary drafts, etc. **Final Products:** Final copy, self-reflection, etc.	**Prewriting:** Brainstorm, web, graphic organizer, journal entry, outline, etc. **Drafting, Revising and Editing:** First draft, teacher conference form, peer editing forms, self-editing forms, secondary drafts, etc. **Final Products:** Final copy, self-reflection, etc.	**Prewriting:** Brainstorm, web, graphic organizer, journal entry, outline, etc. **Drafting, Revising and Editing:** First draft, teacher conference form, peer editing forms, self-editing forms, secondary drafts, etc. **Final Products:** Final copy, self-reflection, etc.	**Prewriting:** Brainstorm, web, graphic organizer, journal entry, outline, etc. **Drafting, Revising and Editing:** First draft, teacher conference form, peer editing forms, self-editing forms, secondary drafts, etc. **Final Products:** Final copy, self-reflection, etc.

Texas Education Agency

APPENDIX D: YEAR-TWO TEXAS WRITING PILOT ANALYTIC RUBRIC (2017–2018)

Very Limited	Limited	Basic	Satisfactory	Accomplished	Exceptional
1	2	3	4	5	6
ORGANIZATION: STRUCTURE, FOCUS, AND PROGRESSION					
The composition does not include a central idea, thesis, or theme. The composition lacks an organizational structure. The composition lacks a central focus and is therefore incoherent and not unified. The composition includes no evidence of connections between ideas.	The composition includes a central idea, thesis, or theme that is mostly unclear. An organizational structure may be evident, but it does not support the development of the central idea, thesis, or theme. The focus is inconsistent, causing the composition to be mostly incoherent and not unified. The sentences, paragraphs, and/or ideas are not clearly connected. Repetition of ideas causes serious disruptions in the flow of the essay.	The central idea, thesis, or theme is somewhat clear. The organizational structure only minimally supports the development of the central idea, thesis, or theme. The focus is at times inconsistent, causing lapses in the composition's coherence and unity. The sentences, paragraphs, and/or ideas are connected by mechanical, formulaic transitions. Some repetition of ideas causes minor disruptions in the flow of the essay.	The central idea, thesis, or theme is clear. The organizational structure is appropriate and adequately supports the development of the central idea, thesis, or theme. The focus is generally consistent and clear, helping the composition remain mostly coherent and unified. The sentences, paragraphs, and/or ideas are connected by logical and mostly effective transitions.	The central idea, thesis, or theme is clear and skillfully presented. The organizational structure is appropriate and effectively supports the development of the central idea, thesis, or theme. The focus is consistent and clear throughout, contributing to the composition's sustained coherence and unity. The sentences, paragraphs, and/or ideas are connected by logical, effective transitions.	The central idea, thesis, or theme is clear and thoughtful. The organizational structure enhances the development of the central idea, thesis, or theme. The focus is consistent and clear throughout, contributing to the composition's sustained coherence and unity. The sentences, paragraphs, and/or ideas are connected by purposeful, logical, and highly effective transitions.
1	2	3	4	5	6
CONTENT: SUPPORT AND ELABORATION					
The composition includes few, if any, details and/or examples related to the topic or theme. The composition may be too brief to reflect an understanding of the writing purpose and/or communicate the writer's intent.	The composition includes details and examples that are list-like and/or too vague to adequately develop the topic or theme. The composition reflects an inadequate understanding of the writing purpose and/or is unable to communicate the writer's intent.	The composition includes mostly relevant details and examples, but they are too general or partially presented to adequately develop the topic or theme. The composition reflects some understanding of the writing purpose and/or only somewhat communicates the writer's intent.	The composition includes relevant details and examples that adequately develop the topic or theme. The composition reflects an adequate understanding of the writing purpose and/or adequately communicates the writer's intent.	The composition includes relevant, specific details and examples that clearly develop the topic or theme. The composition reflects a thorough understanding of the writing purpose and/or strongly communicates the writer's intent.	The composition includes details and examples that are specific, well chosen, relevant, and enhance the development of the topic or theme. The composition reflects a thorough and insightful understanding of the writing purpose and/or clearly communicates the writer's intent in ways that are original and thoughtful.

Very Limited	Limited	Basic	Satisfactory	Accomplished	Exceptional
1	2	3	4	5	6
LANGUAGE					
The composition includes limited diction that is frequently used incorrectly and does not contribute to creating an appropriate/effective tone and style. Literary and/or rhetorical devices are typically missing. The composition includes sentences that are mostly unclear and illogical. Sentences are choppy, irregular, awkward, or incomplete and do not establish the relationships among ideas.	The composition includes simplistic diction that only minimally contributes to the writer's tone and style. Literary and/or rhetorical devices, when used, do not contribute to the quality or effectiveness of the composition. The composition includes sentences that are at times unclear and illogical. Sentences are mostly simple, may include inappropriate fragments, and may not establish the relationships among ideas.	The composition includes sometimes vague or general diction that inconsistently contributes to the writer's tone and style. Literary and/or rhetorical devices, when used, are somewhat effective in contributing to the quality or effectiveness of the composition. The composition includes sentences that are mostly clear and logical. Sentences and phrases may at times be awkward or only somewhat controlled, occasionally weakening the relationships among ideas.	The composition includes mostly appropriate diction that satisfactorily contributes to the writer's tone and style. Literary and/or rhetorical devices, when used, are effective and contribute to the quality or effectiveness of the composition. The composition includes sentences that are consistently clear and logical. Sentences and phrases are adequately controlled and usually establish the relationships among ideas.	The composition includes specific diction that consistently contributes to the writer's tone and style. Literary and/or rhetorical devices, when used, are engaging, and contribute to the quality or effectiveness of the composition. The composition includes sentences that are consistently clear, logical, and varied in structure. Sentences and phrases are skillfully controlled and effectively establish the relationships among ideas.	The composition includes purposeful and precise diction that strongly contributes to the writer's tone and style. Literary and/or rhetorical devices, when used, are effective, engaging, original, and enhance the quality or effectiveness of the composition. The composition includes sentences that are consistently clear, logical, and varied in structure. Sentences and phrases are sophisticated in construction and strongly establish the relationships among ideas.
2		4		6	
CONVENTIONS					
The composition includes a variety of errors reflecting limited or no control of basic writing conventions (spelling, capitalization, punctuation, grammar, and usage). The composition may require extensive editing for conventions errors or may be too brief to evaluate for control of conventions. The composition demonstrates limited or no control of sentence boundaries. If included, paragraph breaks interfere with meaning or demonstrate only a basic understanding of their use.		The composition demonstrates sufficient control of standard writing conventions (spelling, capitalization, punctuation, grammar, and usage). The composition may require minor to moderate editing for conventions errors. The composition demonstrates reasonable control of sentence boundaries. If included, paragraph breaks demonstrate adequate understanding of their use.		The composition demonstrates consistent command of standard writing conventions (spelling, capitalization, punctuation, grammar, and usage). The composition requires minor, if any, editing for conventions errors. The composition may contain purposeful manipulation of conventions for effect. The composition demonstrates consistent control of sentence boundaries, enhancing the composition. If included, paragraph breaks are well controlled and purposeful.	

Muy limitado	Limitado	Básico	Satisfactorio	Sobresaliente	Excepcional
1	2	3	4	5	6
ORGANIZACIÓN: ESTRUCTURA, ENFOQUE Y AVANCE PROGRESIVO					
La composición no incluye una idea principal, una tesis o un mensaje. La composición carece de una estructura organizacional. La composición carece de un enfoque central y por lo tanto no es coherente ni tiene unidad. La composición no incluye evidencia de que las ideas se conectan entre sí.	La composición incluye una idea principal, una tesis o un mensaje, que es confuso en su mayoría. Una estructura organizacional puede ser evidente, pero no apoya el desarrollo de la idea principal, la tesis o el mensaje. El enfoque es inconsistente, lo que causa que la composición sea, en su mayoría, incoherente y no tenga unidad. Las oraciones, las ideas y/o los párrafos no están conectados claramente. La repetición de ideas causa interrupciones serias en la fluidez de la composición.	La idea principal, la tesis o el mensaje es más o menos claro. La estructura organizacional apoya sólo mínimamente el desarrollo de la idea principal, la tesis o el mensaje. El enfoque es un tanto inconsistente, lo que causa fallas en la coherencia y en la unidad de la composición. Las oraciones, las ideas y/o los párrafos están conectados por transiciones mecánicas y siguen un patrón establecido. Algunas repeticiones de ideas causan interrupciones menores en la fluidez de la composición.	La idea principal, la tesis o el mensaje es claro. La estructura organizacional es apropiada y apoya adecuadamente el desarrollo de la idea principal, la tesis o el mensaje. El enfoque es consistente y claro en lo general, lo que ayuda a que la composición sea, en su mayoría, coherente y además tenga unidad. Las oraciones, las ideas y/o los párrafos están conectados usando transiciones lógicas y efectivas en su mayoría.	La idea principal, la tesis o el mensaje es claro y presenta cierta reflexión. La estructura organizacional es apropiada y apoya efectivamente el desarrollo de la idea principal, la tesis o el mensaje. El enfoque es consistente y claro en su totalidad, lo que contribuye a que la composición sea coherente y tenga unidad de principio a fin. Las oraciones, las ideas y/o los párrafos están conectados usando transiciones lógicas y efectivas.	La idea principal, la tesis o el mensaje es claro y muestra gran reflexión a lo largo de la composición. La estructura organizacional destaca el desarrollo de la idea principal, de la tesis o del mensaje. El enfoque es consistente y claro en su totalidad, lo que contribuye a que la composición sea coherente y tenga unidad de principio a fin. Las oraciones, las ideas y/o los párrafos están conectados usando transiciones lógicas y altamente efectivas que fueron seleccionadas intencionalmente.
1	2	3	4	5	6
CONTENIDO: APOYO Y DESARROLLO					
La composición incluye, si acaso, pocos detalles y/o ejemplos relacionados con el tema o el mensaje. La composición puede ser demasiado breve para demostrar comprensión del propósito de escribir un texto de acuerdo con el género asignado y/o para comunicar la intención del escritor.	La composición incluye detalles y ejemplos que parecen listados y/o son demasiado vagos para desarrollar adecuadamente el tema o el mensaje. La composición refleja una comprensión inadecuada del propósito de escribir un texto de acuerdo al género asignado y/o no logra comunicar la intención del escritor.	La composición incluye detalles y ejemplos relevantes en su mayoría, pero se presentan sólo en forma parcial o son demasiado generales para desarrollar adecuadamente el tema o el mensaje. La composición refleja cierta comprensión del propósito de escribir un texto de acuerdo al género asignado y/o comunica tan sólo parcialmente la intención del escritor.	La composición incluye detalles y ejemplos relevantes que desarrollan adecuadamente el tema o el mensaje. La composición refleja una comprensión adecuada del propósito de escribir un texto de acuerdo al género asignado y/o comunica adecuadamente la intención del escritor.	La composición incluye detalles y ejemplos específicos y relevantes que claramente desarrollan el tema o el mensaje. La composición refleja una comprensión total del propósito de escribir un texto de acuerdo al género asignado y/o comunica eficazmente la intención del escritor.	La composición incluye detalles y ejemplos que son específicos, bien seleccionados, relevantes y que destacan el desarrollo del tema o el mensaje. La composición refleja una comprensión total y profunda del propósito de escribir un texto de acuerdo con el género asignado y/o comunica claramente la intención del escritor de una manera original y mostrando una gran reflexión a lo largo de ésta.

1	2	3	4	5	6
LENGUAJE					
La composición incluye un vocabulario limitado que frecuentemente se utiliza de manera incorrecta y no contribuye a crear un tono y estilo adecuado/efectivo.	La composición incluye un vocabulario simple que sólo contribuye de manera mínima al tono y estilo del escritor.	La composición en ocasiones incluye un vocabulario vago o general que contribuye de manera inconsistente al tono y estilo del escritor.	La composición incluye un vocabulario adecuado en su mayoría, el cual contribuye satisfactoriamente al tono y estilo del escritor.	La composición incluye un vocabulario específico que contribuye consistentemente al tono y estilo del escritor.	La composición incluye un vocabulario preciso y seleccionado intencionalmente que contribuye al tono y estilo del escritor de manera sólida.
Los recursos literarios y/o retóricos están generalmente ausentes.	Los recursos literarios y/o retóricos, si se utilizan, no contribuyen a la calidad o efectividad de la composición.	Los recursos literarios y/o retóricos, si se utilizan, son un tanto efectivos al contribuir a la calidad o efectividad de la composición.	Los recursos literarios y/o retóricos, si se utilizan, son efectivos y contribuyen a la calidad o efectividad de la composición.	Los recursos literarios y/o retóricos, si se utilizan, son efectivos, interesantes y contribuyen a la calidad o efectividad de la composición.	Los recursos literarios y/o retóricos, si se utilizan, son efectivos, interesantes, originales y destacan la calidad o efectividad de la composición.
La composición incluye oraciones que son confusas e ilógicas en su mayoría.	La composición incluye oraciones que en ocasiones son confusas e ilógicas.	La composición incluye oraciones que son claras y lógicas en su mayoría.	La composición incluye oraciones que son consistentemente claras y lógicas.	La composición incluye oraciones que son consistentemente claras, lógicas y variadas en su estructura.	La composición incluye oraciones que son consistentemente claras, lógicas y variadas en su estructura.
Las oraciones son muy breves, irregulares, forzadas o incompletas, y no establecen una relación entre las ideas.	Las oraciones son simples en su mayoría, pueden incluir fragmentos inadecuados o no establecer relaciones entre las ideas.	Las oraciones y expresiones pueden ser en ocasiones forzadas o mal manejadas, lo que en ocasiones debilita la relación entre las ideas.	Las oraciones y expresiones están manejadas adecuadamente y por lo general establecen las relaciones entre las ideas.	Las oraciones y expresiones están manejadas hábilmente y establecen con efectividad las relaciones entre las ideas.	Las oraciones y expresiones son sofisticadas en su construcción y establecen fuertemente las relaciones entre las ideas.

2	4	6
CONVENCIONES		
La composición incluye una variedad de errores que reflejan poco o nulo dominio de las convenciones del lenguaje básicas (la ortografía, las mayúsculas, la puntuación, la gramática y el empleo adecuado de las palabras).	La composición demuestra un dominio suficiente de las convenciones del lenguaje básicas (la ortografía, las mayúsculas, la puntuación, la gramática y el empleo adecuado de las palabras).	La composición demuestra un dominio consistente de las convenciones del lenguaje básicas (la ortografía, las mayúsculas, la puntuación, la gramática y el empleo adecuado de las palabras).
La composición puede requerir correcciones extensas por errores en las convenciones del lenguaje o puede ser demasiado breve como para evaluar el dominio de estas convenciones.	La composición puede requerir correcciones menores o moderadas por errores en las convenciones del lenguaje.	La composición requiere correcciones menores, o ninguna, por errores en las convenciones del lenguaje. La composición puede incluir la manipulación de las convenciones del lenguaje de una manera intencionada y efectiva.
La composición demuestra poco o nulo dominio en la separación de las oraciones.	La composición demuestra un dominio razonable en la separación de las oraciones.	La composición demuestra un dominio consistente en la separación de las oraciones, lo cual hace destacar la composición.
Si se incluye, la separación de párrafos interfiere con el significado o demuestra sólo una comprensión básica de su uso.	Si se incluye, la separación de párrafos demuestra una comprensión adecuada de su uso.	Si se incluye, la separación de párrafos está bien manejada y se hace intencionalmente.

APPENDIX E: STUDENT DEMOGRAPHICS

Table E1 lists the demographic distributions for all students and by grade/course. Students who had at least one writing sample that was scored by three raters—Teacher, ESC, and TR1—had their data used in the analyses. Across all grades/courses, most key demographic groups are represented in this study, though not truly representative of the state student population due to the small sample size.

Table E1. Students' Demographics of Analysis Sample

Demographics	Value	All		Grade 4 Writing		Grade 7 Writing		English I		English II	
		N	%	N	%	N	%	N	%	N	%
Total	Total	2755	100	603	100	922	100	723	100	507	100
Region	1	1337	49	300	50	658	71	335	46	44	9
	10	414	15	256	42	158	17	0	0	0	0
	16	1004	36	47	8	106	11	388	54	463	91
Gender	Male	1420	52	308	51	456	49	390	54	266	52
	Female	1334	48	295	49	466	51	332	46	241	48
	No information provided	1	0	0	0	0	0	1	0	0	0
Ethnicity	Hispanic/Latino	1387	50	228	38	385	42	452	63	322	64
	American Indian or Alaska Native	16	1	4	1	3	0	3	0	6	1
	Asian	107	4	29	5	43	5	20	3	15	3
	Black or African American	349	13	76	13	97	11	140	19	36	7
	Native Hawaiian or Other Pacific Islander	2	0	0	0	0	0	0	0	2	0
	White	789	29	236	39	363	39	81	11	109	21
	Two or More Races	80	3	29	5	27	3	13	2	11	2
	No Information Provided	25	1	1	0	4	0	14	2	6	1
Economically Disadvantaged	Yes	1769	64	368	61	542	59	540	75	319	63
	No	986	36	235	39	380	41	183	25	188	37

Demographics	Value	All		Grade 4 Writing		Grade 7 Writing		English I		English II	
		N	%	N	%	N	%	N	%	N	%
Title I, Part A	Participants	1767	64	532	88	597	65	468	65	170	34
	Nonparticipants	988	36	71	12	325	35	255	35	337	66
Migrant	Yes	22	1	1	0	2	0	8	1	11	2
	No	2733	99	602	100	920	100	715	99	496	98
Limited English Proficient	Current LEP	368	13	96	16	135	15	86	12	51	10
	Non-LEP (Monitored 1st Year)	66	2	18	3	17	2	24	3	7	1
	Non-LEP (Monitored 2nd Year)	47	2	10	2	19	2	9	1	9	2
	Other Non-LEP student	2246	82	478	79	747	81	588	81	433	85
	No Information Provided	28	1	1	0	4	0	16	2	7	1
Bilingual	Participants	20	1	15	2	5	1	0	0	0	0
	Nonparticipants	2735	99	588	98	917	99	723	100	507	100
ESL	Participants	343	12	81	13	125	14	86	12	51	10
	Nonparticipants	2412	88	522	87	797	86	637	88	456	90
Special Education	Yes	225	8	44	7	68	7	75	10	38	7
	No	2530	92	559	93	854	93	648	90	469	93
Gifted/Talented	Participants	207	8	74	12	80	9	24	3	29	6
	Nonparticipants	2548	92	529	88	842	91	699	97	478	94
At-Risk	Yes	1572	57	273	45	453	49	535	74	311	61
	No	1183	43	330	55	469	51	188	26	196	39

APPENDIX F: MEAN RATER SCORES

The number of writing samples in each writing group ranged from 33 to 2248. The raters used all valid rating categories, with ratings concentrated on the middle scores: 2 to 5 for organization, content, and language, and 4 for conventions. This indicates that raters were able to distinguish the quality of student writings according to the rubrics. One noteworthy observation is that in general the Teacher gave the highest average scores among the four raters except for organization and conventions scores in English I where the Teacher's average scores were lower than ESC rater's and the second highest. This pattern can be observed in Figures F1–F4 that compare the average rating scores in the four categories among the four raters based on the total writing samples in the four grades/courses, respectively.

Figure F1. Grade 4 Writing Mean Rating Scores on Total Samples

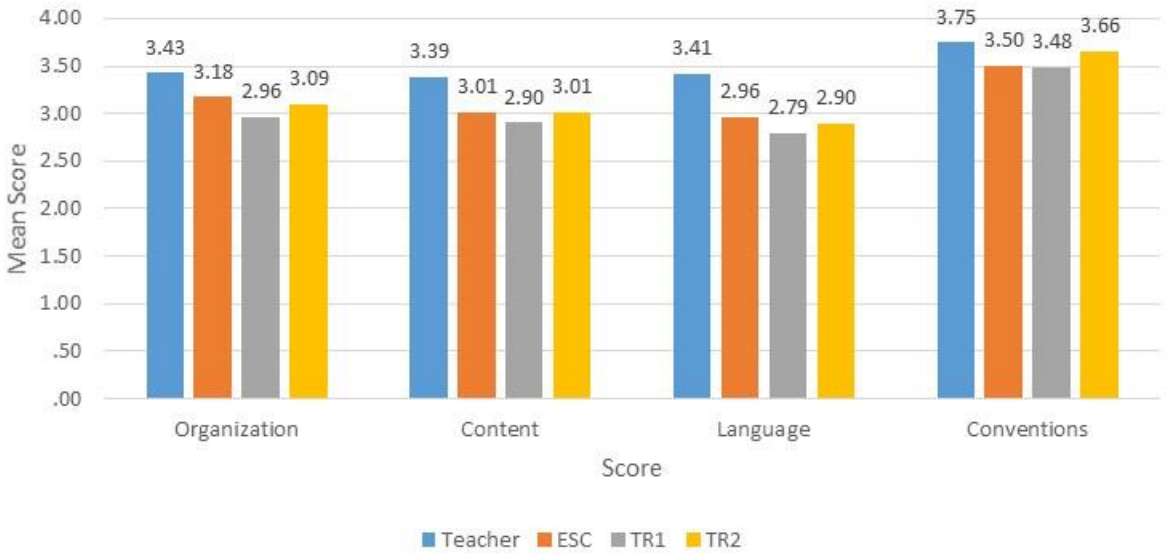


Figure F2. Grade 7 Writing Mean Rating Scores on Total Samples

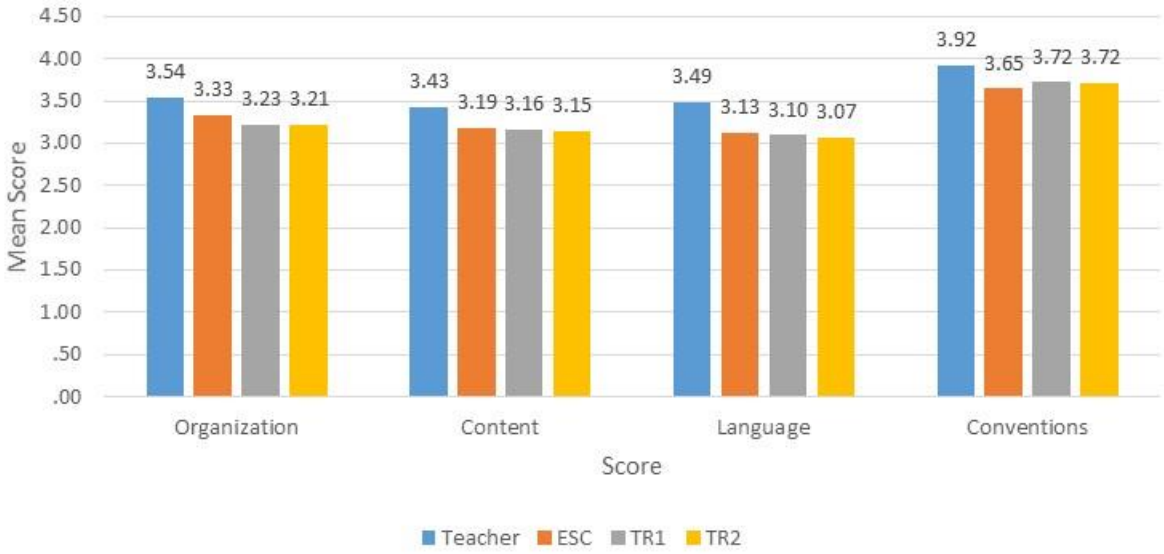


Figure F3. English I Mean Rating Scores on Total Samples

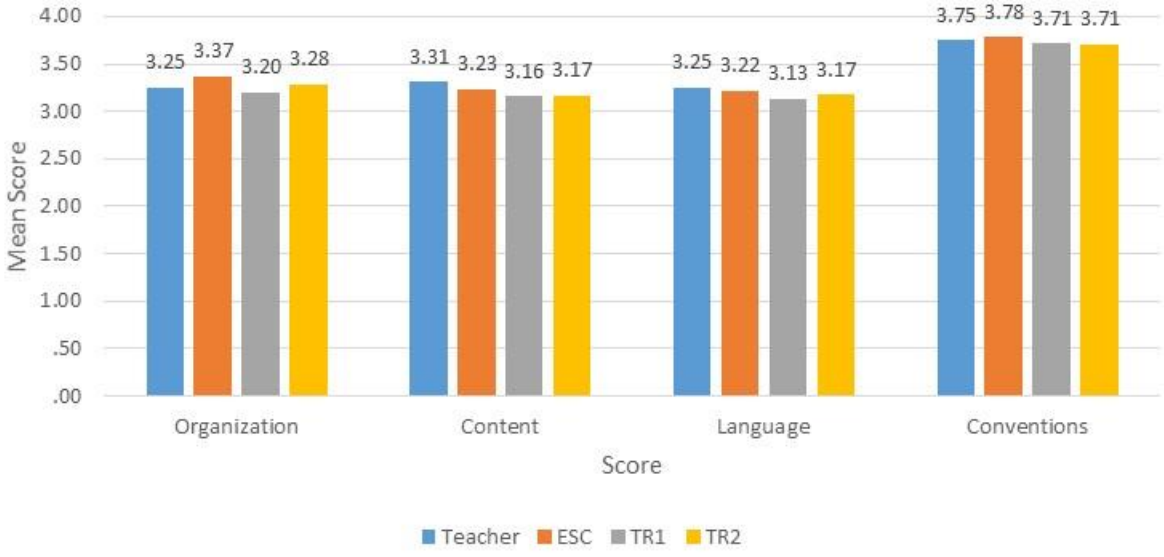
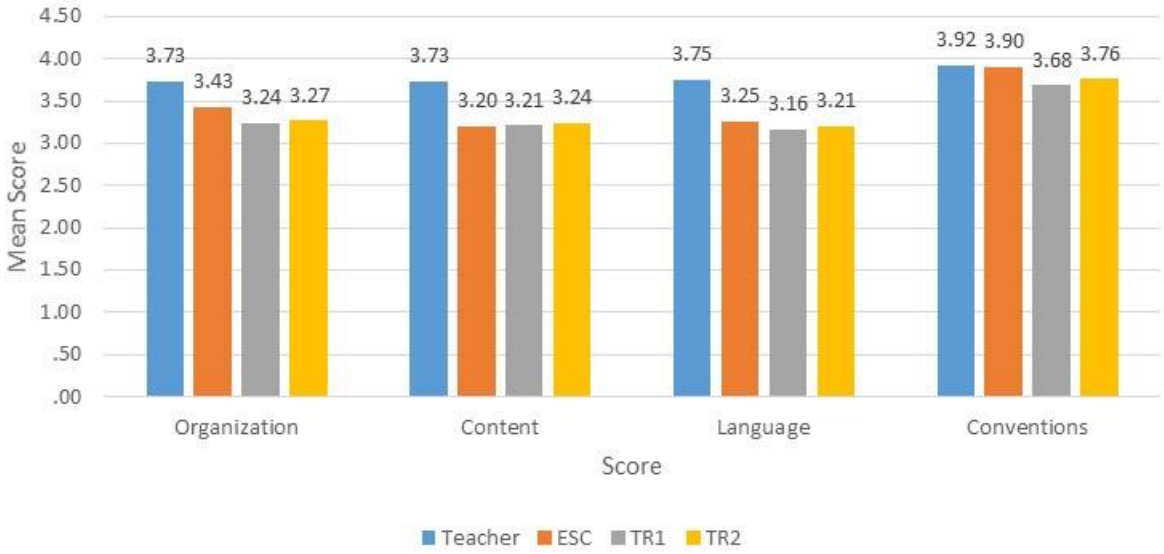


Figure F4. English II Mean Rating Scores on Total Samples


4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
.00
Mean Score
3.73
3.43
3.24
3.27
3.73
3.20
3.21
3.24
3.75
3.25
3.16
3.21
3.92
3.90
3.68
3.76
Organization
Content
Language
Conventions
Score
Teacher
ESC
TR1
TR2

APPENDIX G: RATER SCORE SUMMARY BY CATEGORY

Tables G1.1 to G1.4 show the summary statistics for scores by category: organization, content, language, and conventions, respectively, for each rater in grade 4 writing, including number of responses (N), rating score mean (Mean), standard deviation (StdDev), and the percentage of students at each score point (S1-S6). Note that Conventions only have three valid score points: 2, 4, and 6. The summary statistics were calculated for each writing sample (TS1, PS1, PS2, and TS2), each writing genre, each timed writing prompt, and the total writing samples with a sample size of at least 30. Tables G2.1 to G2.4 are the same tables for grade 7 writing, G3.1 to G3.4 are for English I, and G4.1 to G4.4 are for English II.

Table G1.1 Rater Scores Summary: Grade 4, Organization

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	603	2.89	1.11	11	26	33	22	6	1
		ESC	603	2.94	1.15	11	25	34	20	9	1
		TR1	603	2.70	1.12	14	31	32	17	5	1
		TR2	124	2.85	1.04	8	30	38	18	6	1
	PS1	Teacher	361	3.39	1.00	1	17	40	29	11	2
		ESC	361	3.14	1.17	7	24	34	24	8	3
		TR1	361	2.82	.99	9	29	38	20	4	0
		TR2	62	2.82	.91	5	32	44	15	5	0
	PS2	Teacher	602	3.66	1.12	3	10	30	35	16	5
		ESC	602	3.42	1.15	4	16	34	27	15	3
		TR1	602	3.18	1.00	5	19	39	30	6	1
		TR2	200	3.36	1.01	3	15	38	34	8	3
	TS2	Teacher	603	3.76	1.20	3	12	27	30	22	7
		ESC	603	3.18	1.17	7	23	31	27	10	2
		TR1	603	3.10	1.09	8	19	39	25	7	2
		TR2	131	3.05	1.19	11	18	44	16	9	3
Genre	Analytic	Teacher	93	2.91	1.08	12	23	32	29	4	0

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	93	3.72	1.30	5	11	30	20	27	6
		TR1	93	3.40	1.10	6	10	38	33	10	3
		TR2	60	3.50	.93	0	12	42	35	8	3
	Expository	Teacher	526	3.60	1.07	1	12	36	33	13	5
		ESC	526	3.26	1.17	6	21	33	27	10	3
		TR1	526	2.98	1.00	6	25	38	25	5	0
		TR2	83	3.11	1.00	4	22	45	23	5	2
	Other	Teacher	1								
		ESC	1								
		TR1	1								
		TR2	0								
	Personal Narrative	Teacher	79	3.42	1.23	5	20	27	27	19	3
		ESC	79	3.24	1.24	9	18	33	25	11	4
		TR1	79	2.76	1.13	15	27	32	20	6	0
		TR2	59	3.08	1.13	8	24	27	34	5	2
	Persuasive	Teacher	264	3.73	.98	1	7	34	36	19	3
		ESC	264	3.32	1.04	3	18	38	27	13	1
		TR1	264	3.13	.93	3	21	41	29	5	0
		TR2	60	3.27	.94	2	18	42	28	10	0
	Personal Narrative_TS	Teacher	1206	3.32	1.23	7	19	30	26	14	4
		ESC	1206	3.06	1.17	9	24	33	23	9	2
		TR1	1206	2.90	1.12	11	25	35	21	6	1
		TR2	255	2.95	1.12	9	24	41	17	7	2
Timed Sample Prompt	1000022	Teacher	523	3.21	1.19	7	22	31	26	12	2

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	523	3.08	1.17	8	24	35	19	11	2
		TR1	523	2.86	1.09	11	26	37	19	6	1
		TR2	116	2.82	1.03	9	28	42	16	3	2
	1000023	Teacher	293	3.37	1.28	7	20	27	25	16	4
		ESC	293	3.05	1.16	11	21	32	27	8	1
		TR1	293	2.89	1.16	13	22	37	20	6	2
		TR2	52	3.02	1.15	12	17	40	19	12	0
	1000024	Teacher	390	3.43	1.25	7	15	32	27	15	5
		ESC	390	3.05	1.17	9	26	29	26	8	2
		TR1	390	2.96	1.14	10	26	33	24	6	2
		TR2	87	3.09	1.22	9	21	39	17	10	3
Total	Total	Teacher	2169	3.43	1.17	5	16	32	29	14	4
		ESC	2169	3.18	1.17	7	22	33	25	11	2
		TR1	2169	2.96	1.08	9	24	37	23	6	1
		TR2	517	3.09	1.07	6	21	40	23	7	2

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G1.2 Rater Scores Summary: Grade 4, Content

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	603	2.82	1.10	10	31	33	17	7	0
		ESC	603	2.80	1.17	13	30	30	18	8	1
		TR1	603	2.67	1.11	14	33	31	16	5	1
		TR2	124	2.73	1.00	9	33	40	14	4	1

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
	PS1	Teacher	361	3.35	1.03	2	19	39	28	11	2
		ESC	361	2.96	1.24	11	27	30	21	7	4
		TR1	361	2.66	.98	9	38	34	16	3	0
		TR2	62	2.71	.89	3	45	32	16	3	0
	PS2	Teacher	602	3.65	1.11	3	11	31	32	19	4
		ESC	602	3.25	1.21	7	20	32	25	12	3
		TR1	602	3.05	.99	5	22	43	23	6	1
		TR2	200	3.26	1.04	5	16	40	31	8	2
	TS2	Teacher	603	3.73	1.23	4	13	26	30	21	7
		ESC	603	3.02	1.18	8	28	29	24	8	2
		TR1	603	3.12	1.10	7	21	37	25	8	2
		TR2	131	3.03	1.09	8	20	47	16	8	2
Genre	Analytic	Teacher	93	2.97	1.07	9	26	32	27	6	0
		ESC	93	3.61	1.33	6	14	27	24	23	6
		TR1	93	3.26	1.19	8	16	35	30	5	5
		TR2	60	3.48	.98	0	15	38	33	10	3
	Expository	Teacher	526	3.59	1.08	2	13	35	32	15	5
		ESC	526	3.10	1.24	10	23	30	25	8	4
		TR1	526	2.82	.99	7	32	38	18	5	0
		TR2	83	3.02	1.00	2	31	35	27	2	2
	Other	Teacher	1								
		ESC	1								
		TR1	1								
		TR2	0								

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
	Personal Narrative	Teacher	79	3.51	1.34	6	19	25	22	23	5
		ESC	79	3.08	1.32	13	22	29	24	8	5
		TR1	79	2.68	1.10	14	33	30	16	6	0
		TR2	59	2.92	1.12	15	15	37	27	5	0
	Persuasive	Teacher	264	3.64	1.00	1	10	36	32	19	2
		ESC	264	3.09	1.12	6	26	35	22	10	2
		TR1	264	3.02	.90	4	22	47	22	5	0
		TR2	60	3.13	.96	2	25	42	22	10	0
	Personal Narrative_TS	Teacher	1206	3.27	1.25	7	22	29	24	14	4
		ESC	1206	2.91	1.18	11	29	30	21	8	2
		TR1	1206	2.90	1.13	11	27	34	21	7	1
		TR2	255	2.89	1.05	8	26	43	15	6	2
Timed Sample Prompt	1000022	Teacher	523	3.15	1.21	7	27	29	22	13	2
		ESC	523	2.96	1.16	9	30	30	20	9	2
		TR1	523	2.87	1.09	9	29	35	20	6	1
		TR2	116	2.79	.98	8	28	47	11	3	2
	1000023	Teacher	293	3.34	1.31	8	19	29	21	18	4
		ESC	293	2.85	1.17	15	23	34	20	8	1
		TR1	293	2.89	1.15	12	26	34	20	7	1
		TR2	52	2.94	1.00	8	23	42	21	6	0
	1000024	Teacher	390	3.38	1.25	7	18	29	27	13	5
		ESC	390	2.89	1.21	11	33	25	22	7	2
		TR1	390	2.94	1.17	11	25	33	22	7	2
		TR2	87	2.98	1.17	9	25	38	16	9	2

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Total	Total	Teacher	2169	3.39	1.19	5	18	31	27	15	4
		ESC	2169	3.01	1.21	10	26	30	22	9	2
		TR1	2169	2.90	1.08	9	27	36	20	6	1
		TR2	517	3.01	1.05	6	24	40	21	6	2

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G1.3. Rater Scores Summary: Grade 4, Language

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	603	2.86	1.13	12	27	34	20	7	1
		ESC	603	2.73	1.14	13	33	28	19	6	1
		TR1	603	2.54	1.10	17	36	29	12	5	1
		TR2	124	2.57	.96	10	40	36	10	2	1
	PS1	Teacher	361	3.50	.93	1	14	33	40	10	2
		ESC	361	2.85	1.21	12	30	30	19	7	2
		TR1	361	2.67	1.00	10	37	32	16	4	0
		TR2	62	2.73	.93	5	40	37	13	5	0
	PS2	Teacher	602	3.59	1.11	3	13	28	36	17	3
		ESC	602	3.20	1.18	7	22	32	26	11	2
		TR1	602	2.89	.94	5	29	43	18	4	0
		TR2	200	3.10	1.01	6	20	44	24	6	2
	TS2	Teacher	603	3.75	1.23	4	11	27	28	24	6
		ESC	603	3.02	1.14	6	31	31	21	9	2
		TR1	603	3.01	1.11	9	20	41	20	7	2

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		TR2	131	2.98	1.14	10	18	49	12	8	3
Genre	Analytic	Teacher	93	3.00	1.11	8	28	31	25	8	1
		ESC	93	3.49	1.29	8	13	31	25	18	5
		TR1	93	3.06	1.12	6	25	38	20	9	2
		TR2	60	3.23	.98	0	23	42	27	5	3
	Expository	Teacher	526	3.65	.95	0	11	29	43	13	3
		ESC	526	2.97	1.20	11	27	30	22	8	2
		TR1	526	2.79	.97	7	33	37	18	5	0
		TR2	83	2.99	1.01	5	28	39	23	5	1
	Other	Teacher	1								
		ESC	1								
		TR1	1								
		TR2	0								
	Personal Narrative	Teacher	79	3.09	1.33	13	24	23	27	10	4
		ESC	79	3.13	1.32	11	22	30	22	10	5
		TR1	79	2.62	.99	11	35	38	10	5	0
		TR2	59	2.85	1.11	15	17	42	19	7	0
	Persuasive	Teacher	264	3.69	1.03	1	11	32	33	20	3
		ESC	264	3.11	1.10	6	25	34	24	11	1
		TR1	264	2.80	.87	5	33	42	18	2	0
		TR2	60	2.98	.89	2	28	47	17	7	0
	Personal Narrative_TS	Teacher	1206	3.30	1.26	8	19	30	24	15	4
		ESC	1206	2.87	1.15	10	32	30	20	7	1
		TR1	1206	2.78	1.13	13	28	35	16	6	1

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		TR2	255	2.78	1.07	10	29	43	11	5	2
Timed Sample Prompt	1000022	Teacher	523	3.18	1.19	8	20	34	23	13	2
		ESC	523	2.89	1.13	9	31	31	20	7	1
		TR1	523	2.75	1.07	11	31	36	15	6	1
		TR2	116	2.62	.95	10	34	44	9	1	2
	1000023	Teacher	293	3.31	1.33	10	18	31	19	18	4
		ESC	293	2.81	1.13	13	29	31	20	7	0
		TR1	293	2.76	1.17	15	27	35	15	6	2
		TR2	52	2.90	1.01	8	25	44	15	8	0
	1000024	Teacher	390	3.46	1.28	6	19	25	28	16	5
		ESC	390	2.91	1.17	8	36	27	19	8	2
		TR1	390	2.82	1.17	14	26	34	19	6	2
		TR2	87	2.93	1.23	11	24	40	11	9	3
Total	Total	Teacher	2169	3.41	1.18	5	17	30	30	15	3
		ESC	2169	2.96	1.18	9	29	30	21	8	2
		TR1	2169	2.79	1.06	10	30	37	17	5	1
		TR2	517	2.90	1.04	8	26	42	16	5	2

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G1.4. Rater Scores Summary: Grade 4, Conventions

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	603	3.27	1.18		42		52		6
		ESC	603	3.28	1.29		45		46		9

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		TR1	603	3.09	1.29		54		38		8
		TR2	124	3.24	1.21		44		49		6
	PS1	Teacher	361	3.90	1.04		16		73		11
		ESC	361	3.51	1.37		39		47		14
		TR1	361	3.33	1.25		42		50		8
		TR2	62	3.39	1.12		35		60		5
	PS2	Teacher	602	3.97	1.22		19		63		18
		ESC	602	3.70	1.25		28		58		13
		TR1	602	3.78	1.20		24		63		13
		TR2	200	4.00	1.20		18		64		18
	TS2	Teacher	603	3.91	1.27		22		60		18
		ESC	603	3.50	1.26		35		54		10
		TR1	603	3.67	1.33		32		53		15
		TR2	131	3.65	1.23		29		60		11
Genre	Analytic	Teacher	93	3.18	1.07		43		55		2
		ESC	93	3.91	1.35		25		55		20
		TR1	93	3.81	1.25		25		60		15
		TR2	60	3.97	1.19		18		65		17
	Expository	Teacher	526	4.00	1.08		15		71		14
		ESC	526	3.62	1.36		34		50		15
		TR1	526	3.55	1.27		34		55		11
		TR2	83	3.76	1.19		24		64		12
	Other	Teacher	1								
		ESC	1								

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		TR1	1								
		TR2	0								
	Personal Narrative	Teacher	79	3.70	1.51		37		42		22
		ESC	79	3.52	1.25		34		56		10
		TR1	79	3.49	1.22		34		57		9
		TR2	59	3.90	1.31		24		58		19
	Persuasive	Teacher	264	4.18	1.09		11		70		20
		ESC	264	3.60	1.17		29		62		9
		TR1	264	3.70	1.17		25		64		11
		TR2	60	3.83	1.18		22		65		13
	Personal Narrative_TS	Teacher	1206	3.59	1.27		32		56		12
		ESC	1206	3.39	1.27		40		50		10
		TR1	1206	3.38	1.34		43		45		12
		TR2	255	3.45	1.23		36		55		9
Timed Sample Prompt	1000022	Teacher	523	3.53	1.19		32		59		9
		ESC	523	3.38	1.28		41		50		10
		TR1	523	3.37	1.36		44		44		12
		TR2	116	3.34	1.17		39		55		6
	1000023	Teacher	293	3.57	1.37		37		48		15
		ESC	293	3.33	1.23		41		51		8
		TR1	293	3.34	1.27		42		49		9
		TR2	52	3.50	1.24		35		56		10
	1000024	Teacher	390	3.68	1.28		30		56		14
		ESC	390	3.45	1.30		39		50		11

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		TR1	390	3.43	1.38		42		44		14
		TR2	87	3.56	1.31		34		53		13
Total	Total	Teacher	2169	3.75	1.23		26		61		13
		ESC	2169	3.50	1.29		37		52		11
		TR1	2169	3.48	1.30		38		51		12
		TR2	517	3.66	1.24		29		59		12

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G2.1 Rater Scores Summary: Grade 7, Organization

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	890	3.36	1.21	7	18	30	30	13	4
		ESC	890	3.20	1.25	8	22	30	24	11	4
		TR1	890	3.12	1.17	9	21	34	25	9	2
		TR2	185	3.04	1.09	9	20	37	27	6	1
	PS1	Teacher	383	3.73	1.18	4	10	27	31	22	5
		ESC	383	3.66	1.40	7	14	25	25	17	11
		TR1	383	3.43	1.17	5	17	31	30	14	4
		TR2	139	3.24	1.43	12	21	27	21	11	9
	PS2	Teacher	237	3.66	1.33	8	12	22	30	22	7
		ESC	237	3.57	1.30	5	16	27	28	16	8
		TR1	237	3.40	1.22	7	15	30	34	8	6
		TR2	99	3.54	1.25	4	15	32	29	10	9
	TS2	Teacher	362	3.71	1.09	2	10	30	38	14	7

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	362	3.16	1.22	9	18	36	22	10	3
		TR1	362	3.15	1.09	7	18	39	26	9	1
		TR2	83	3.17	1.09	6	20	37	23	13	0
Genre	Analytic	Teacher	126	4.42	1.13	1	6	11	35	29	19
		ESC	126	4.11	1.35	2	11	19	26	23	18
		TR1	126	3.94	1.26	4	9	21	31	25	10
		TR2	59	4.17	1.43	5	10	12	29	24	20
	Expository	Teacher	60	3.10	1.23	13	15	35	22	15	0
		ESC	60	2.88	1.22	12	32	25	22	8	2
		TR1	60	2.78	.90	12	18	50	20	0	0
		TR2	60	2.52	1.03	18	30	37	12	3	0
	Other	Teacher	1								
		ESC	1								
		TR1	1								
		TR2	0								
	Personal Narrative	Teacher	316	3.62	1.16	5	11	28	32	21	3
		ESC	316	3.54	1.36	8	15	28	24	16	9
		TR1	316	3.36	1.15	5	19	29	33	10	4
		TR2	60	3.35	1.33	7	22	27	28	8	8
	Persuasive	Teacher	117	3.47	1.21	7	15	27	29	21	2
		ESC	117	3.74	1.23	3	12	26	34	15	9
		TR1	117	3.34	1.14	6	15	33	32	9	3
		TR2	59	3.44	1.12	3	12	42	29	7	7
	Expository_TS	Teacher	1252	3.46	1.19	5	15	30	32	13	5

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	1252	3.19	1.24	9	21	32	24	11	4
		TR1	1252	3.13	1.14	8	20	35	25	9	2
		TR2	268	3.08	1.09	8	20	37	26	8	1
Timed Sample Prompt	1000029	Teacher	619	3.40	1.14	5	16	32	32	12	3
		ESC	619	3.19	1.23	9	19	33	24	11	3
		TR1	619	3.17	1.12	7	19	36	26	10	2
		TR2	118	3.22	1.07	7	16	36	31	9	1
	1000030	Teacher	180	3.40	1.13	5	16	32	32	13	3
		ESC	180	3.14	1.27	9	22	33	19	12	4
		TR1	180	3.10	1.13	7	23	33	27	6	3
		TR2	33	2.94	1.09	12	21	30	33	3	0
	1000031	Teacher	453	3.58	1.26	6	15	26	33	14	7
		ESC	453	3.21	1.25	8	23	30	25	10	4
		TR1	453	3.09	1.17	10	20	35	24	9	2
		TR2	117	2.97	1.09	9	24	39	19	9	1
Total	Total	Teacher	1872	3.54	1.21	5	14	28	32	16	5
		ESC	1872	3.33	1.30	8	19	30	24	13	6
		TR1	1872	3.23	1.17	7	19	34	27	10	3
		TR2	506	3.21	1.23	8	19	33	25	9	5

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G2.2 Rater Scores Summary: Grade 7, Content

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	890	3.26	1.21	6	22	31	24	13	3
		ESC	890	3.04	1.26	11	24	34	18	9	4
		TR1	890	3.04	1.19	10	23	33	23	9	2
		TR2	185	3.02	1.09	9	21	36	26	6	1
	PS1	Teacher	383	3.59	1.17	4	11	35	24	22	4
		ESC	383	3.59	1.39	7	15	28	22	18	10
		TR1	383	3.27	1.20	6	21	32	25	13	3
		TR2	139	3.08	1.38	12	25	28	17	12	6
	PS2	Teacher	237	3.60	1.30	6	13	31	23	20	7
		ESC	237	3.42	1.36	7	21	27	24	14	8
		TR1	237	3.39	1.27	7	18	27	32	11	6
		TR2	99	3.49	1.33	4	20	30	23	12	10
	TS2	Teacher	362	3.57	1.14	3	14	33	30	17	5
		ESC	362	2.96	1.21	11	25	35	19	7	3
		TR1	362	3.19	1.10	7	17	39	25	10	2
		TR2	83	3.12	1.12	7	20	39	22	11	1
Genre	Analytic	Teacher	126	4.17	1.12	1	3	29	23	32	12
		ESC	126	3.99	1.35	2	13	24	21	25	15
		TR1	126	3.83	1.26	5	9	26	28	25	8
		TR2	59	4.05	1.33	2	12	22	25	22	17
	Expository	Teacher	60	3.08	1.21	15	10	40	22	13	0
		ESC	60	2.75	1.13	15	25	38	13	8	0
		TR1	60	2.57	.96	15	32	35	18	0	0
		TR2	60	2.32	.98	22	38	28	10	2	0

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
	Other	Teacher	1								
		ESC	1								
		TR1	1								
		TR2	0								
	Personal Narrative	Teacher	316	3.49	1.17	5	13	35	26	18	3
		ESC	316	3.50	1.37	8	17	27	25	16	9
		TR1	316	3.24	1.18	6	22	32	28	9	4
		TR2	60	3.27	1.31	8	22	28	23	13	5
	Persuasive	Teacher	117	3.49	1.24	4	18	32	21	22	3
		ESC	117	3.51	1.36	5	20	27	25	13	10
		TR1	117	3.36	1.23	7	19	26	31	14	3
		TR2	59	3.39	1.26	3	20	37	20	10	8
	Expository_TS	Teacher	1252	3.35	1.20	5	20	32	26	14	4
		ESC	1252	3.02	1.25	11	24	34	18	9	4
		TR1	1252	3.08	1.16	9	21	35	24	9	2
		TR2	268	3.05	1.10	9	21	37	25	7	1
Timed Sample Prompt	1000029	Teacher	619	3.29	1.16	5	22	31	27	12	3
		ESC	619	3.01	1.26	11	25	33	18	9	4
		TR1	619	3.14	1.12	7	21	36	25	10	2
		TR2	118	3.18	1.06	6	20	33	32	8	1
	1000030	Teacher	180	3.34	1.11	4	17	36	25	16	1
		ESC	180	3.05	1.28	11	23	34	17	11	4
		TR1	180	3.02	1.11	8	23	37	23	7	2
		TR2	33	2.94	1.20	15	18	30	33	0	3

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
	1000031	Teacher	453	3.44	1.29	6	17	30	25	16	6
		ESC	453	3.02	1.21	10	23	36	20	7	4
		TR1	453	3.02	1.24	13	20	32	23	9	2
		TR2	117	2.96	1.10	9	22	43	15	9	1
Total	Total	Teacher	1872	3.43	1.21	5	17	32	25	16	4
		ESC	1872	3.19	1.31	9	22	32	20	11	6
		TR1	1872	3.16	1.19	8	21	33	25	10	3
		TR2	506	3.15	1.24	9	22	33	23	9	4

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G2.3 Rater Scores Summary: Grade 7, Language

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	890	3.31	1.17	6	19	32	28	13	3
		ESC	890	2.97	1.26	12	25	32	20	8	4
		TR1	890	3.00	1.13	9	24	35	21	8	1
		TR2	185	2.96	1.03	8	24	37	25	5	1
	PS1	Teacher	383	3.66	1.14	4	10	28	36	18	4
		ESC	383	3.46	1.39	9	15	28	23	16	8
		TR1	383	3.19	1.23	9	21	32	24	12	3
		TR2	139	2.96	1.42	16	27	25	15	12	5
	PS2	Teacher	237	3.62	1.27	5	14	26	28	21	5
		ESC	237	3.38	1.31	6	21	28	25	13	7
		TR1	237	3.30	1.26	8	19	28	31	7	6

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		TR2	99	3.42	1.31	5	19	33	22	11	9
	TS2	Teacher	362	3.65	1.10	2	12	31	35	15	6
		ESC	362	3.01	1.18	10	25	32	24	6	3
		TR1	362	3.10	1.08	8	18	41	23	7	2
		TR2	83	3.05	1.06	6	23	42	19	8	1
Genre	Analytic	Teacher	126	4.24	1.08	0	7	15	37	28	13
		ESC	126	3.92	1.34	3	12	25	25	21	14
		TR1	126	3.78	1.31	6	10	28	25	22	10
		TR2	59	4.03	1.35	2	12	25	20	24	17
	Expository	Teacher	60	2.98	1.27	18	12	35	25	8	2
		ESC	60	2.52	1.13	22	28	32	13	5	0
		TR1	60	2.52	1.03	18	33	27	22	0	0
		TR2	60	2.17	1.03	30	37	22	10	2	0
	Other	Teacher	1								
		ESC	1								
		TR1	1								
		TR2	0								
	Personal Narrative	Teacher	316	3.58	1.10	4	11	31	35	17	3
		ESC	316	3.41	1.33	8	16	29	25	15	7
		TR1	316	3.14	1.19	8	21	34	25	8	3
		TR2	60	3.17	1.29	8	27	25	23	13	3
	Persuasive	Teacher	117	3.49	1.23	5	18	26	26	21	3
		ESC	117	3.41	1.34	7	21	26	26	13	8
		TR1	117	3.24	1.17	7	21	26	35	7	3

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		TR2	59	3.27	1.22	5	19	42	19	8	7
	Expository_TS	Teacher	1252	3.41	1.16	5	17	31	30	13	4
		ESC	1252	2.98	1.24	12	25	32	21	8	3
		TR1	1252	3.03	1.12	9	23	37	22	8	2
		TR2	268	2.99	1.04	7	24	39	23	6	1
Timed Sample Prompt	1000029	Teacher	619	3.35	1.12	4	20	31	31	12	3
		ESC	619	2.95	1.23	12	25	32	19	9	3
		TR1	619	3.08	1.10	8	21	39	22	9	1
		TR2	118	3.14	.98	5	19	42	27	8	0
	1000030	Teacher	180	3.41	1.15	6	14	32	32	13	3
		ESC	180	2.98	1.25	14	21	33	23	6	4
		TR1	180	2.98	1.10	7	29	33	24	5	2
		TR2	33	2.94	1.14	12	21	33	30	0	3
	1000031	Teacher	453	3.49	1.20	5	15	32	28	15	5
		ESC	453	3.02	1.24	10	25	31	23	7	4
		TR1	453	2.98	1.15	11	23	36	21	8	2
		TR2	117	2.85	1.06	9	30	38	17	6	1
Total	Total	Teacher	1872	3.49	1.17	5	15	30	31	15	4
		ESC	1872	3.13	1.30	11	22	30	22	10	5
		TR1	1872	3.10	1.17	9	22	35	24	9	2
		TR2	506	3.07	1.22	9	24	34	21	9	4

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G2.4 Rater Scores Summary: Grade 7, Conventions

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	890	3.75	1.35		30		53		17
		ESC	890	3.53	1.31		36		51		12
		TR1	890	3.66	1.22		29		60		11
		TR2	185	3.71	1.19		26		63		11
	PS1	Teacher	383	4.06	1.33		21		56		24
		ESC	383	3.98	1.41		25		51		24
		TR1	383	3.70	1.40		33		49		18
		TR2	139	3.53	1.50		42		39		19
	PS2	Teacher	237	3.98	1.43		26		49		25
		ESC	237	3.86	1.38		27		52		20
		TR1	237	3.76	1.31		28		56		16
		TR2	99	3.94	1.38		25		53		22
	TS2	Teacher	362	4.15	1.28		17		59		24
		ESC	362	3.45	1.27		38		52		10
		TR1	362	3.87	1.31		25		57		18
		TR2	83	3.81	1.42		30		49		20
Genre	Analytic	Teacher	126	4.68	1.24		8		50		42
		ESC	126	4.33	1.36		16		52		33
		TR1	126	4.38	1.38		16		49		35
		TR2	59	4.58	1.44		15		41		44
	Expository	Teacher	60	3.43	1.28		38		52		10
		ESC	60	3.17	1.18		47		48		5

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		TR1	60	2.97	1.19		57		38		5
		TR2	60	2.73	1.10		67		30		3
	Other	Teacher	1								
		ESC	1								
		TR1	1								
		TR2	0								
	Personal Narrative	Teacher	316	3.92	1.33		24		56		20
		ESC	316	3.92	1.38		26		52		22
		TR1	316	3.61	1.34		34		51		15
		TR2	60	3.60	1.37		35		50		15
	Persuasive	Teacher	117	3.91	1.40		26		51		22
		ESC	117	3.95	1.43		26		50		24
		TR1	117	3.73	1.23		26		61		13
		TR2	59	3.90	1.31		24		58		19
	Expository_TS	Teacher	1252	3.87	1.34		26		55		19
		ESC	1252	3.50	1.30		37		52		12
		TR1	1252	3.72	1.25		27		59		13
		TR2	268	3.74	1.26		27		59		14
Timed Sample Prompt	1000029	Teacher	619	3.86	1.36		27		54		20
		ESC	619	3.51	1.33		37		50		13
		TR1	619	3.76	1.24		26		60		14
		TR2	118	3.85	1.29		25		58		17
	1000030	Teacher	180	3.89	1.32		24		57		19
		ESC	180	3.54	1.33		36		51		13

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		TR1	180	3.61	1.22		30		59		11
		TR2	33	3.52	1.12		30		64		6
	1000031	Teacher	453	3.87	1.32		25		56		19
		ESC	453	3.48	1.25		36		54		10
		TR1	453	3.70	1.26		29		58		13
		TR2	117	3.69	1.28		29		57		14
Total	Total	Teacher	1872	3.92	1.35		25		54		21
		ESC	1872	3.65	1.35		33		51		15
		TR1	1872	3.72	1.29		29		57		15
		TR2	506	3.72	1.36		31		52		17

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G3.1 Rater Scores Summary: English I, Organization

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	681	2.82	1.03	10	28	37	19	5	0
		ESC	681	3.14	1.16	8	20	35	25	10	2
		TR1	681	3.14	1.14	7	23	33	25	10	2
		TR2	147	3.05	1.11	7	25	36	22	7	2
	PS1	Teacher	289	3.20	1.12	4	23	37	21	13	1
		ESC	289	3.56	1.23	3	19	25	32	13	7
		TR1	289	3.22	1.22	6	22	35	21	11	4
		TR2	73	3.49	1.23	4	18	30	25	19	4
	PS2	Teacher	597	3.50	1.00	2	13	33	39	11	2

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	597	3.77	1.16	3	10	23	39	17	7
		TR1	597	3.36	1.09	3	19	33	30	12	3
		TR2	165	3.38	1.07	4	16	33	34	10	2
	TS2	Teacher	681	3.46	1.12	6	12	32	33	15	2
		ESC	681	3.17	1.06	6	19	41	24	10	1
		TR1	681	3.12	1.02	6	21	38	29	5	1
		TR2	148	3.28	.98	4	16	37	36	6	1
Genre	Analytic	Teacher	115	3.61	1.01	0	12	37	34	13	4
		ESC	115	3.83	1.22	2	14	22	35	18	10
		TR1	115	3.30	1.14	3	20	42	22	8	6
		TR2	59	3.59	1.19	3	14	32	27	19	5
	Expository	Teacher	306	3.12	1.12	5	25	38	18	13	1
		ESC	306	3.38	1.22	4	23	27	29	11	6
		TR1	306	3.04	1.16	7	26	37	19	8	3
		TR2	60	3.13	1.20	7	25	33	22	10	3
	Other	Teacher	59	4.17	1.05	0	5	22	34	29	10
		ESC	59	4.47	1.16	0	5	14	36	20	25
		TR1	59	4.25	1.09	2	0	25	31	29	14
		TR2	0								
	Personal Narrative	Teacher	320	3.47	.93	3	11	31	45	10	0
		ESC	320	3.77	1.08	4	7	23	44	18	4
		TR1	320	3.36	1.04	3	18	32	33	13	1
		TR2	60	3.53	1.00	3	8	37	37	13	2
	Persuasive	Teacher	86	3.37	.88	0	17	36	40	6	1

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	86	3.87	1.05	1	7	28	37	21	6
		TR1	86	3.51	.99	0	19	28	38	14	1
		TR2	59	3.41	1.04	2	20	27	39	10	2
	Expository_TS	Teacher	1362	3.14	1.13	8	20	35	26	10	1
		ESC	1362	3.16	1.11	7	20	38	24	10	2
		TR1	1362	3.13	1.08	6	22	35	27	8	1
		TR2	295	3.17	1.05	5	20	37	29	7	2
Timed Sample Prompt	1000032	Teacher	399	3.13	1.13	8	21	34	26	11	1
		ESC	399	3.25	1.12	6	18	38	26	11	2
		TR1	399	3.22	1.14	6	23	32	28	11	2
		TR2	81	3.33	1.11	4	19	35	31	9	4
	1000033	Teacher	415	3.02	1.08	8	23	37	23	9	0
		ESC	415	3.06	1.15	9	23	33	24	9	2
		TR1	415	3.11	1.12	7	22	34	27	9	1
		TR2	93	3.06	1.06	6	25	32	30	5	1
	21000001	Teacher	548	3.24	1.15	8	17	34	29	11	2
		ESC	548	3.17	1.07	6	18	42	23	9	2
		TR1	548	3.08	1.02	6	21	39	28	5	1
		TR2	121	3.13	1.01	6	18	41	27	7	1
Total	Total	Teacher	2248	3.25	1.10	6	19	34	29	11	1
		ESC	2248	3.37	1.17	5	17	33	29	12	4
		TR1	2248	3.20	1.11	5	21	35	27	9	2
		TR2	533	3.28	1.09	5	19	35	30	10	2

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G3.2 Rater Scores Summary: English I, Content

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	681	2.93	1.07	8	25	41	16	8	1
		ESC	681	2.96	1.16	10	26	34	20	8	2
		TR1	681	3.07	1.12	7	25	34	23	9	1
		TR2	147	2.95	1.18	11	26	31	24	5	3
	PS1	Teacher	289	3.23	1.14	4	22	40	16	16	2
		ESC	289	3.52	1.35	6	17	28	23	17	8
		TR1	289	3.25	1.19	4	23	34	23	10	4
		TR2	73	3.42	1.19	5	14	37	25	15	4
	PS2	Teacher	597	3.66	1.04	2	12	29	36	19	3
		ESC	597	3.62	1.23	5	12	28	32	16	7
		TR1	597	3.37	1.08	3	18	36	28	12	3
		TR2	165	3.33	1.08	4	15	39	30	8	4
	TS2	Teacher	681	3.41	1.11	5	14	34	32	13	2
		ESC	681	3.05	1.11	6	27	35	22	8	2
		TR1	681	3.01	1.00	5	24	41	23	4	1
		TR2	148	3.07	1.01	5	25	36	28	5	1
Genre	Analytic	Teacher	115	3.66	1.04	0	12	35	33	15	5
		ESC	115	3.76	1.29	3	15	28	23	22	10
		TR1	115	3.33	1.12	3	17	46	22	6	7
		TR2	59	3.58	1.21	3	15	31	27	19	5
	Expository	Teacher	306	3.12	1.12	5	24	41	15	14	1

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	306	3.28	1.28	7	21	33	20	14	5
		TR1	306	3.07	1.11	5	27	38	20	8	3
		TR2	60	3.10	1.10	7	20	42	23	5	3
	Other	Teacher	59	4.19	.96	0	3	19	42	27	8
		ESC	59	4.44	1.30	0	7	20	25	17	31
		TR1	59	4.31	1.05	0	2	24	32	27	15
		TR2	0								
	Personal Narrative	Teacher	320	3.73	1.05	2	11	27	36	23	2
		ESC	320	3.65	1.18	6	9	24	40	16	5
		TR1	320	3.38	1.06	3	17	34	32	13	2
		TR2	60	3.50	1.03	5	5	42	35	10	3
	Persuasive	Teacher	86	3.50	.84	0	12	36	44	7	1
		ESC	86	3.66	1.15	3	10	31	30	20	5
		TR1	86	3.42	1.00	0	22	29	34	15	0
		TR2	59	3.27	1.06	3	19	39	29	7	3
	Expository_TS	Teacher	1362	3.17	1.12	7	19	38	24	10	2
		ESC	1362	3.00	1.13	8	26	35	21	8	2
		TR1	1362	3.04	1.06	6	25	38	23	7	1
		TR2	295	3.01	1.10	8	25	34	26	5	2
Timed Sample Prompt	1000032	Teacher	399	3.14	1.05	6	19	41	25	8	1
		ESC	399	3.06	1.13	9	23	35	23	10	1
		TR1	399	3.15	1.08	4	25	35	25	10	2
		TR2	81	3.25	1.15	2	27	30	30	6	5
	1000033	Teacher	415	3.10	1.12	7	23	39	18	13	1

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	415	2.94	1.14	9	27	35	20	6	2
		TR1	415	3.07	1.10	7	25	33	26	8	1
		TR2	93	2.90	1.19	16	18	32	27	5	1
	21000001	Teacher	548	3.25	1.16	7	18	34	28	10	3
		ESC	548	3.01	1.13	7	28	35	20	8	2
		TR1	548	2.94	1.01	7	24	43	21	4	1
		TR2	121	2.93	.98	5	30	38	22	4	1
Total	Total	Teacher	2248	3.31	1.12	5	18	36	26	13	2
		ESC	2248	3.23	1.22	7	21	32	24	11	4
		TR1	2248	3.16	1.09	5	23	37	25	8	2
		TR2	533	3.17	1.12	6	21	36	27	7	3

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G3.3 Rater Scores Summary: English I, Language

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	681	2.88	1.06	10	26	37	21	6	0
		ESC	681	2.98	1.11	9	26	34	25	6	2
		TR1	681	3.04	1.11	7	26	35	22	9	1
		TR2	147	2.98	1.10	7	28	33	25	5	2
	PS1	Teacher	289	3.31	1.11	4	18	39	23	13	2
		ESC	289	3.44	1.21	5	18	30	28	15	4
		TR1	289	3.22	1.16	5	22	36	23	10	4
		TR2	73	3.48	1.19	4	15	36	23	18	4

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
	PS2	Teacher	597	3.53	1.03	2	14	30	39	13	2
		ESC	597	3.46	1.15	6	12	34	33	11	5
		TR1	597	3.34	1.09	3	20	35	28	12	3
		TR2	165	3.25	1.03	4	17	41	28	8	2
	TS2	Teacher	681	3.35	1.07	7	11	36	33	11	1
		ESC	681	3.15	1.04	5	20	38	27	7	1
		TR1	681	2.99	1.00	6	24	41	23	4	1
		TR2	148	3.13	1.03	5	22	36	30	7	1
Genre	Analytic	Teacher	115	3.48	1.05	0	17	40	28	10	5
		ESC	115	3.62	1.16	3	13	29	32	18	4
		TR1	115	3.27	1.12	3	20	44	19	9	5
		TR2	59	3.51	1.17	3	15	34	25	19	3
	Expository	Teacher	306	3.26	1.11	5	19	39	23	12	2
		ESC	306	3.21	1.15	6	22	35	25	11	3
		TR1	306	3.03	1.09	5	28	38	20	7	3
		TR2	60	3.17	1.12	5	22	40	22	8	3
	Other	Teacher	59	4.25	1.04	0	3	22	32	31	12
		ESC	59	4.22	1.18	0	5	25	31	20	19
		TR1	59	4.34	1.04	0	2	22	32	29	15
		TR2	0								
	Personal Narrative	Teacher	320	3.53	.96	3	11	30	43	13	0
		ESC	320	3.46	1.14	7	10	32	38	9	4
		TR1	320	3.35	1.05	3	18	34	32	12	1
		TR2	60	3.50	.98	5	5	38	40	10	2

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
	Persuasive	Teacher	86	3.37	.98	0	26	21	45	7	1
		ESC	86	3.58	1.08	3	9	35	34	15	3
		TR1	86	3.41	.99	0	22	29	35	14	0
		TR2	59	3.10	1.03	3	24	44	19	8	2
	Expository_TS	Teacher	1362	3.11	1.09	8	19	37	27	9	1
		ESC	1362	3.06	1.08	7	23	36	26	7	1
		TR1	1362	3.01	1.06	7	25	38	23	7	1
		TR2	295	3.05	1.06	6	25	34	27	6	1
Timed Sample Prompt	1000032	Teacher	399	3.14	1.08	7	20	36	28	9	1
		ESC	399	3.15	1.09	6	22	34	29	8	2
		TR1	399	3.11	1.12	6	26	33	24	9	2
		TR2	81	3.26	1.10	4	21	37	25	11	2
	1000033	Teacher	415	2.99	1.08	9	22	37	23	8	0
		ESC	415	2.93	1.09	9	27	34	25	4	2
		TR1	415	3.01	1.07	7	25	37	23	7	1
		TR2	93	2.98	1.09	10	24	32	29	4	1
	21000001	Teacher	548	3.19	1.09	8	15	37	30	9	1
		ESC	548	3.11	1.05	6	21	40	25	7	1
		TR1	548	2.94	.99	7	25	42	22	4	1
		TR2	121	2.98	1.00	6	28	34	28	3	1
Total	Total	Teacher	2248	3.25	1.09	6	17	35	30	10	1
		ESC	2248	3.22	1.13	6	20	35	28	9	3
		TR1	2248	3.13	1.09	5	23	37	24	8	2
		TR2	533	3.17	1.08	5	21	36	27	8	2

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G3.4 Rater Scores Summary: English I, Conventions

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	681	3.34	1.28		42		48		9
		ESC	681	3.65	1.26		30		57		13
		TR1	681	3.65	1.28		31		56		13
		TR2	147	3.63	1.22		29		60		11
	PS1	Teacher	289	3.90	1.34		25		55		20
		ESC	289	4.09	1.40		22		51		27
		TR1	289	3.83	1.32		26		56		18
		TR2	73	3.81	1.34		27		55		18
	PS2	Teacher	597	3.87	1.15		20		66		13
		ESC	597	3.77	1.29		27		57		16
		TR1	597	3.81	1.19		23		64		13
		TR2	165	3.75	1.06		21		70		8
	TS2	Teacher	681	3.99	1.30		21		58		21
		ESC	681	3.79	.99		18		75		7
		TR1	681	3.64	1.16		28		63		9
		TR2	148	3.69	1.11		24		67		9
Genre	Analytic	Teacher	115	3.77	1.18		23		64		12
		ESC	115	4.05	1.46		25		47		28
		TR1	115	3.95	1.20		19		64		17
		TR2	59	3.93	1.28		22		59		19
	Expository	Teacher	306	3.94	1.37		25		53		22

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	306	3.82	1.40		29		51		20
		TR1	306	3.60	1.25		31		57		11
		TR2	60	3.53	1.24		33		57		10
	Other	Teacher	59	4.92	1.19		5		44		51
		ESC	59	4.41	1.22		10		59		31
		TR1	59	4.92	1.07		2		51		47
		TR2	0								
	Personal Narrative	Teacher	320	3.78	.98		18		75		7
		ESC	320	3.72	1.18		25		63		11
		TR1	320	3.76	1.17		24		64		12
		TR2	60	3.90	1.00		15		75		10
	Persuasive	Teacher	86	3.47	1.12		33		62		6
		ESC	86	4.05	1.41		23		51		26
		TR1	86	3.84	1.16		21		66		13
		TR2	59	3.69	1.04		22		71		7
	Expository_TS	Teacher	1362	3.66	1.33		32		53		15
		ESC	1362	3.72	1.13		24		66		10
		TR1	1362	3.64	1.22		29		59		11
		TR2	295	3.66	1.16		27		63		10
Timed Sample Prompt	1000032	Teacher	399	3.69	1.35		32		52		16
		ESC	399	3.80	1.10		21		69		11
		TR1	399	3.73	1.25		27		59		14
		TR2	81	3.83	1.15		21		67		12
	1000033	Teacher	415	3.57	1.37		36		49		15

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	415	3.59	1.21		31		59		10
		TR1	415	3.63	1.26		31		57		12
		TR2	93	3.44	1.16		34		59		6
	21000001	Teacher	548	3.71	1.27		28		57		14
		ESC	548	3.76	1.09		22		69		9
		TR1	548	3.59	1.17		29		62		9
		TR2	121	3.72	1.16		25		64		11
Total	Total	Teacher	2248	3.75	1.29		28		57		15
		ESC	2248	3.78	1.22		25		62		14
		TR1	2248	3.71	1.23		27		60		13
		TR2	533	3.71	1.16		25		64		11

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G4.1 Rater Scores Summary: English II, Organization

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	469	3.35	1.08	3	20	29	34	12	1
		ESC	469	3.48	1.29	6	16	31	27	12	9
		TR1	469	3.24	1.08	5	20	33	32	7	3
		TR2	102	3.19	1.11	5	22	38	24	9	3
	PS1	Teacher	431	3.98	1.12	1	7	25	36	20	10
		ESC	431	3.40	.99	2	14	42	30	9	3
		TR1	431	3.45	1.09	2	18	33	31	13	3
		TR2	77	3.39	1.02	0	22	31	35	9	3

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
	PS2	Teacher	428	3.77	1.03	1	10	26	39	20	3
		ESC	428	3.44	1.11	2	21	29	32	13	3
		TR1	428	3.29	.98	3	16	43	28	8	2
		TR2	163	3.39	1.05	2	15	42	30	6	6
	TS2	Teacher	387	3.87	1.16	3	11	19	36	26	5
		ESC	387	3.39	1.04	3	14	39	32	10	3
		TR1	387	2.95	1.05	8	24	38	24	4	2
		TR2	65	2.95	1.16	14	17	38	23	6	2
Genre	Analytic	Teacher	160	3.67	.94	0	11	31	39	18	1
		ESC	160	3.36	1.12	1	24	31	28	12	4
		TR1	160	3.75	1.02	1	9	31	38	18	4
		TR2	60	3.78	1.11	0	12	28	40	10	10
	Expository	Teacher	183	4.12	1.19	2	8	19	35	23	14
		ESC	183	3.56	1.05	1	13	39	28	15	4
		TR1	183	3.42	1.18	2	24	29	25	16	4
		TR2	60	3.18	1.02	0	28	37	27	5	3
	Other	Teacher	161	4.17	1.15	0	7	22	31	25	14
		ESC	161	3.32	.94	3	14	41	34	7	1
		TR1	161	3.43	1.09	4	17	29	36	12	2
		TR2	0								
	Personal Narrative	Teacher	98	3.76	1.21	5	10	22	33	26	4
		ESC	98	3.80	1.13	1	12	24	39	15	8
		TR1	98	3.49	.88	1	8	44	37	8	2
		TR2	60	3.50	.98	2	10	42	33	10	3

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
	Persuasive	Teacher	257	3.69	.89	1	7	32	44	14	2
		ESC	257	3.28	1.01	3	20	36	32	9	1
		TR1	257	3.01	.85	4	19	52	21	3	0
		TR2	60	3.08	.91	3	20	47	27	2	2
	Persuasive_TS	Teacher	856	3.59	1.15	3	16	25	35	18	3
		ESC	856	3.44	1.18	5	15	34	29	11	6
		TR1	856	3.11	1.08	7	22	36	29	5	2
		TR2	167	3.10	1.14	8	20	38	23	8	2
Timed Sample Prompt	1000035	Teacher	284	3.46	1.17	4	20	26	30	17	3
		ESC	284	3.29	1.18	7	15	36	29	9	4
		TR1	284	3.02	1.10	8	23	35	26	5	2
		TR2	62	2.85	1.02	10	26	39	21	5	0
	1000036	Teacher	308	3.63	1.08	2	15	24	41	15	4
		ESC	308	3.47	1.18	4	15	36	28	10	7
		TR1	308	3.14	1.03	5	23	36	31	5	2
		TR2	59	3.20	1.20	8	15	42	19	12	3
	1000037	Teacher	264	3.67	1.19	4	13	23	33	23	3
		ESC	264	3.56	1.18	3	16	31	30	14	6
		TR1	264	3.17	1.11	7	19	36	29	6	3
		TR2	46	3.28	1.17	7	17	33	33	7	4
Total	Total	Teacher	1715	3.73	1.12	2	12	25	36	19	5
		ESC	1715	3.43	1.12	3	16	35	30	11	5
		TR1	1715	3.24	1.07	4	19	37	29	8	2
		TR2	407	3.27	1.09	4	18	38	28	7	4

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G4.2 Rater Scores Summary: English II, Content

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	469	3.39	1.15	4	21	27	31	16	2
		ESC	469	3.16	1.29	9	23	36	16	11	6
		TR1	469	3.14	1.15	6	24	34	25	9	3
		TR2	102	3.13	1.19	7	25	31	24	10	3
	PS1	Teacher	431	3.90	1.15	1	9	29	27	26	8
		ESC	431	3.24	1.07	3	23	35	27	9	3
		TR1	431	3.45	1.10	2	17	33	33	10	4
		TR2	77	3.42	1.10	0	26	25	35	10	4
	PS2	Teacher	428	3.83	.99	0	9	26	39	23	3
		ESC	428	3.42	1.07	1	17	41	25	12	4
		TR1	428	3.29	.99	2	17	45	25	9	2
		TR2	163	3.37	1.02	2	16	41	29	8	4
	TS2	Teacher	387	3.85	1.19	4	9	21	36	24	6
		ESC	387	2.96	1.10	9	23	41	19	5	2
		TR1	387	2.93	1.15	10	26	37	18	7	2
		TR2	65	2.89	1.17	14	22	34	26	2	3
Genre	Analytic	Teacher	160	3.61	.90	0	10	36	39	13	2
		ESC	160	3.51	.98	0	14	37	36	8	4
		TR1	160	3.74	1.02	1	9	33	36	16	5
		TR2	60	3.73	1.06	0	12	30	38	13	7
	Expository	Teacher	183	3.97	1.17	2	11	20	22	42	2

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	183	3.42	1.14	2	20	37	22	15	4
		TR1	183	3.51	1.22	2	21	29	27	14	7
		TR2	60	3.25	1.11	0	30	32	27	7	5
	Other	Teacher	161	4.19	1.16	0	7	22	33	21	17
		ESC	161	3.18	1.04	5	22	31	34	7	1
		TR1	161	3.39	1.08	4	17	30	37	10	2
		TR2	0								
	Personal Narrative	Teacher	98	3.74	1.12	1	13	29	28	27	3
		ESC	98	3.69	1.15	2	11	35	24	22	5
		TR1	98	3.54	.89	1	7	43	37	10	2
		TR2	60	3.50	1.00	2	13	35	35	13	2
	Persuasive	Teacher	257	3.80	.97	1	6	32	39	19	4
		ESC	257	3.10	1.00	2	25	46	18	6	3
		TR1	257	2.96	.82	3	22	54	18	2	1
		TR2	60	3.05	.91	3	22	47	25	2	2
	Persuasive_TS	Teacher	856	3.60	1.19	4	16	24	33	20	4
		ESC	856	3.07	1.21	9	23	38	18	8	4
		TR1	856	3.05	1.16	8	25	35	21	8	3
		TR2	167	3.04	1.19	10	24	32	25	7	3
Timed Sample Prompt	1000035	Teacher	284	3.46	1.18	4	20	24	31	18	2
		ESC	284	2.93	1.21	11	26	35	18	6	4
		TR1	284	2.96	1.13	10	25	35	23	6	2
		TR2	62	2.76	1.07	13	31	26	29	2	0
	1000036	Teacher	308	3.62	1.11	3	14	27	36	18	3

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	308	3.11	1.22	8	21	40	18	8	5
		TR1	308	3.05	1.13	6	27	36	22	6	3
		TR2	59	3.15	1.24	8	20	37	20	8	5
	1000037	Teacher	264	3.72	1.27	5	13	21	32	23	6
		ESC	264	3.17	1.21	7	22	39	17	11	4
		TR1	264	3.14	1.22	8	22	35	20	12	3
		TR2	46	3.26	1.22	7	20	35	24	11	4
Total	Total	Teacher	1715	3.73	1.14	2	12	26	33	22	5
		ESC	1715	3.20	1.15	5	21	38	22	9	4
		TR1	1715	3.21	1.11	5	21	37	25	9	3
		TR2	407	3.24	1.12	5	21	34	29	8	3

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G4.3 Rater Scores Summary: English II, Language

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	469	3.44	1.10	3	19	27	35	14	2
		ESC	469	3.25	1.25	7	19	35	24	8	6
		TR1	469	3.12	1.11	6	22	38	23	8	2
		TR2	102	3.13	1.11	5	23	42	20	7	4
	PS1	Teacher	431	3.95	1.13	0	9	29	30	23	10
		ESC	431	3.32	.99	2	17	39	30	9	2
		TR1	431	3.42	1.02	1	17	36	32	11	3
		TR2	77	3.38	.97	0	19	38	30	12	1

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
	PS2	Teacher	428	3.78	.96	1	7	28	43	19	2
		ESC	428	3.39	1.05	2	18	38	28	12	3
		TR1	428	3.27	.98	2	17	45	26	7	2
		TR2	163	3.33	1.05	2	18	42	28	6	5
	TS2	Teacher	387	3.85	1.14	4	8	19	42	22	5
		ESC	387	3.02	1.12	8	25	36	24	4	3
		TR1	387	2.81	1.13	13	25	39	16	5	2
		TR2	65	2.82	1.09	14	22	38	23	2	2
Genre	Analytic	Teacher	160	3.71	.93	2	4	34	42	15	3
		ESC	160	3.43	1.04	1	19	31	36	9	3
		TR1	160	3.76	1.01	1	9	29	41	15	5
		TR2	60	3.70	1.14	0	13	33	33	10	10
	Expository	Teacher	183	4.21	1.13	1	8	19	25	38	10
		ESC	183	3.41	1.09	1	19	38	26	11	4
		TR1	183	3.44	1.14	1	22	34	25	14	5
		TR2	60	3.18	.93	0	23	45	23	7	2
	Other	Teacher	161	4.09	1.16	0	8	24	34	19	15
		ESC	161	3.32	.95	4	14	37	36	9	0
		TR1	161	3.37	1.01	2	19	33	35	11	1
		TR2	0								
	Personal Narrative	Teacher	98	3.64	1.05	2	13	24	41	17	2
		ESC	98	3.64	1.09	2	9	38	30	16	5
		TR1	98	3.46	.92	1	10	45	32	10	2
		TR2	60	3.45	1.08	2	17	37	28	13	3

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
	Persuasive	Teacher	257	3.66	.91	1	8	35	40	16	1
		ESC	257	3.18	.95	2	21	46	22	9	1
		TR1	257	2.96	.79	4	20	55	20	1	0
		TR2	60	3.05	.83	3	20	47	28	2	0
	Persuasive_TS	Teacher	856	3.63	1.13	3	14	23	38	18	3
		ESC	856	3.15	1.20	8	22	35	24	6	5
		TR1	856	2.98	1.13	9	23	38	20	7	2
		TR2	167	3.01	1.11	8	22	41	21	5	3
Timed Sample Prompt	1000035	Teacher	284	3.50	1.17	5	18	23	36	16	3
		ESC	284	3.06	1.20	10	21	37	23	5	5
		TR1	284	2.88	1.14	10	29	34	19	6	2
		TR2	62	2.81	1.04	10	27	42	16	3	2
	1000036	Teacher	308	3.64	1.04	2	13	26	42	16	3
		ESC	308	3.14	1.21	7	22	37	21	7	5
		TR1	308	2.99	1.09	8	24	40	20	6	2
		TR2	59	3.12	1.15	8	17	42	22	7	3
	1000037	Teacher	264	3.74	1.19	4	13	20	36	23	5
		ESC	264	3.25	1.18	5	22	32	28	7	5
		TR1	264	3.09	1.16	10	17	42	20	9	3
		TR2	46	3.13	1.15	7	22	37	26	4	4
Total	Total	Teacher	1715	3.75	1.10	2	11	26	37	20	5
		ESC	1715	3.25	1.12	5	20	37	27	8	4
		TR1	1715	3.16	1.08	5	20	39	24	8	2
		TR2	407	3.21	1.07	4	20	41	25	7	3

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

Table G4.4 Rater Scores Summary: English II, Conventions

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
Writing Sample	TS1	Teacher	469	3.67	1.18		27		63		10
		ESC	469	3.86	1.35		26		54		19
		TR1	469	3.57	1.29		34		54		12
		TR2	102	3.69	1.22		27		61		12
	PS1	Teacher	431	4.04	1.30		20		58		22
		ESC	431	3.92	1.06		16		72		12
		TR1	431	4.00	1.16		17		66		17
		TR2	77	3.90	1.21		21		64		16
	PS2	Teacher	428	3.95	1.13		17		68		15
		ESC	428	3.96	1.17		18		66		16
		TR1	428	3.83	1.05		18		72		10
		TR2	163	3.91	1.10		17		70		13
	TS2	Teacher	387	4.07	1.31		20		57		23
		ESC	387	3.87	1.34		26		55		19
		TR1	387	3.31	1.27		43		48		9
		TR2	65	3.32	1.19		40		54		6
Genre	Analytic	Teacher	160	4.33	1.18		11		63		27
		ESC	160	4.01	1.11		15		69		16
		TR1	160	4.34	1.13		9		66		26
		TR2	60	4.13	1.21		15		63		22
	Expository	Teacher	183	4.34	1.54		22		38		39

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	183	4.00	1.22		19		63		19
		TR1	183	3.91	1.24		21		62		17
		TR2	60	3.83	1.18		22		65		13
	Other	Teacher	161	3.83	1.10		20		69		11
		ESC	161	4.01	.88		9		81		10
		TR1	161	3.93	1.12		17		69		14
		TR2	0								
	Personal Narrative	Teacher	98	3.71	1.07		22		69		8
		ESC	98	4.06	1.16		15		66		18
		TR1	98	4.00	.95		11		78		11
		TR2	60	4.00	1.10		15		70		15
	Persuasive	Teacher	257	3.76	.98		19		75		7
		ESC	257	3.76	1.14		23		66		11
		TR1	257	3.63	.96		23		74		4
		TR2	60	3.67	.99		22		73		5
	Persuasive_TS	Teacher	856	3.85	1.26		24		60		16
		ESC	856	3.86	1.34		26		55		19
		TR1	856	3.45	1.28		38		51		11
		TR2	167	3.54	1.22		32		58		10
Timed Sample Prompt	1000035	Teacher	284	3.75	1.30		28		56		15
		ESC	284	3.76	1.32		29		55		17
		TR1	284	3.32	1.26		42		49		8
		TR2	62	3.29	1.09		39		58		3
	1000036	Teacher	308	3.86	1.18		21		65		14

Group	Sample[a]	Rater	N	Mean	StdDev	S1 (%)	S2 (%)	S3 (%)	S4 (%)	S5 (%)	S6 (%)
		ESC	308	3.86	1.29		24		58		18
		TR1	308	3.44	1.26		38		52		10
		TR2	59	3.76	1.24		25		61		14
	1000037	Teacher	264	3.95	1.29		22		58		20
		ESC	264	3.97	1.42		26		50		24
		TR1	264	3.60	1.33		34		52		14
		TR2	46	3.61	1.31		33		54		13
Total	Total	Teacher	1715	3.92	1.24		21		62		17
		ESC	1715	3.90	1.23		22		62		17
		TR1	1715	3.68	1.22		28		60		12
		TR2	407	3.76	1.18		24		64		12

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; the numbers are the prompt IDs of timed samples.

APPENDIX H: SCORE CORRELATIONS BETWEEN RATERS

Figures H1.1–H1.4 compare the percentages of exact agreement (EA), percentages of exact or adjacent agreement (EAA), polychoric correlations (Cor), and quadratic weighted kappa coefficients (WKC), respectively, among the four raters on the four rating scores based on the total writing samples in grade 4. Figures H2.1–H2.4 are the same plots for grade 7 writing, Figures H3.1–H3.4 for English I, and Figures H4.1–H4.4 for English II. Across all rating scores, rater pairs, and the four tests, the range of exact agreement based on all writing samples was between 28% and 45% (except for Conventions score), the range of adjacent agreement was between 72% and 87% (except for Conventions score), the range of polychoric correlations was between 0.38 and 0.67, and the range of kappa coefficients was between 0.32 and 0.63. For Conventions score based on all writing samples, the range of exact agreement was between 54% and 65% and the range of adjacent agreement was between 96% and 99%, which were higher than those of the other scores because it has three score categories while the other scores have six score categories.

Figure H1.1. Percentage of Exact Agreement between Raters on Total Samples: Grade 4 Writing.

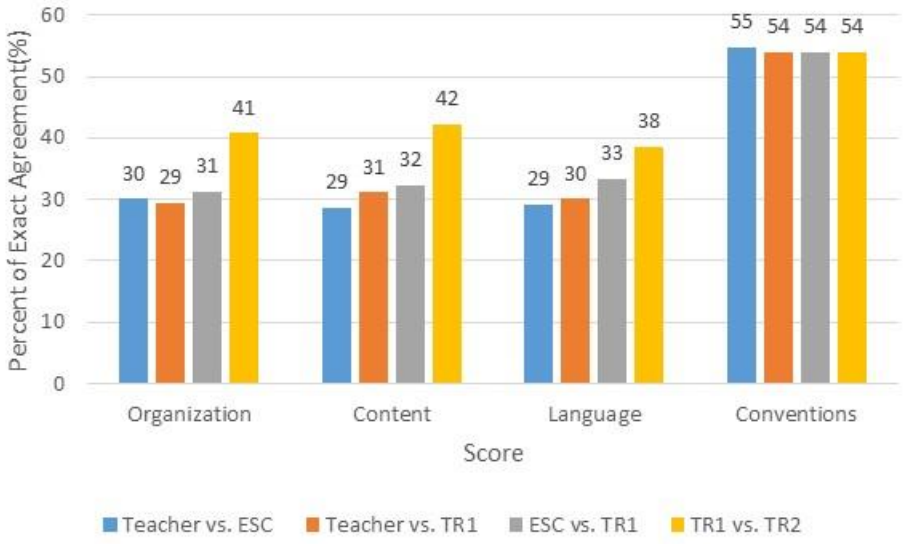


Figure H1.2. Percentage of Adjacent Agreement between Raters on Total Samples: Grade 4 Writing.

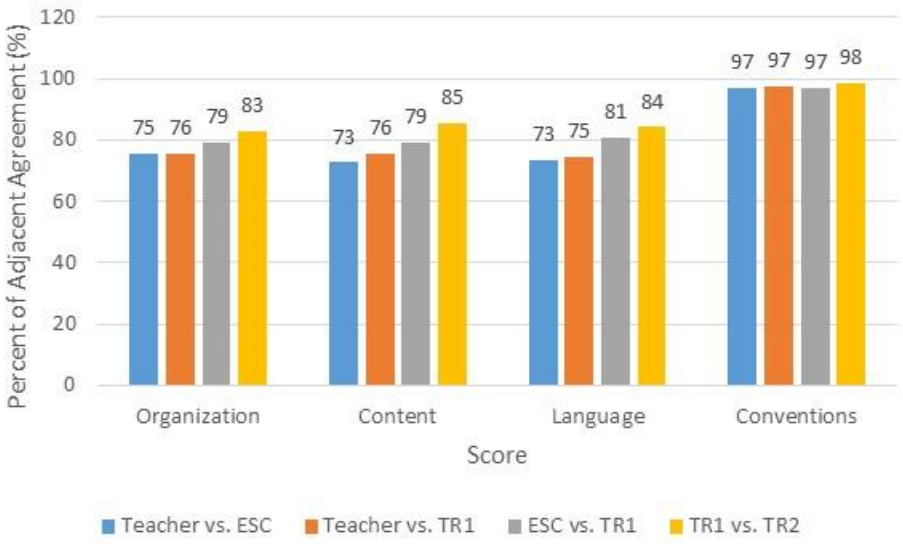


Figure H1.3. Score Correlation between Raters on Total Samples: Grade 4 Writing.

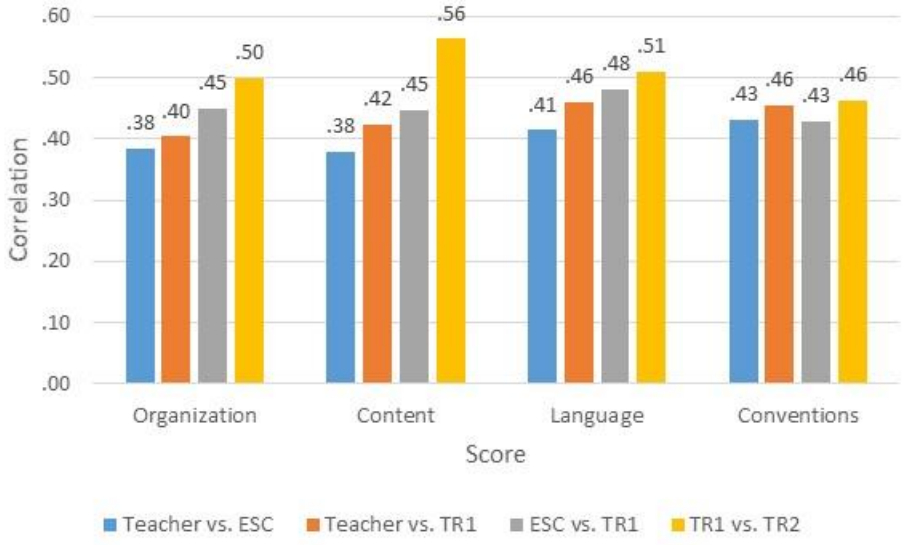


Figure H1.4. Weighted Kappa Coefficient between Raters on Total Samples: Grade 4 Writing.

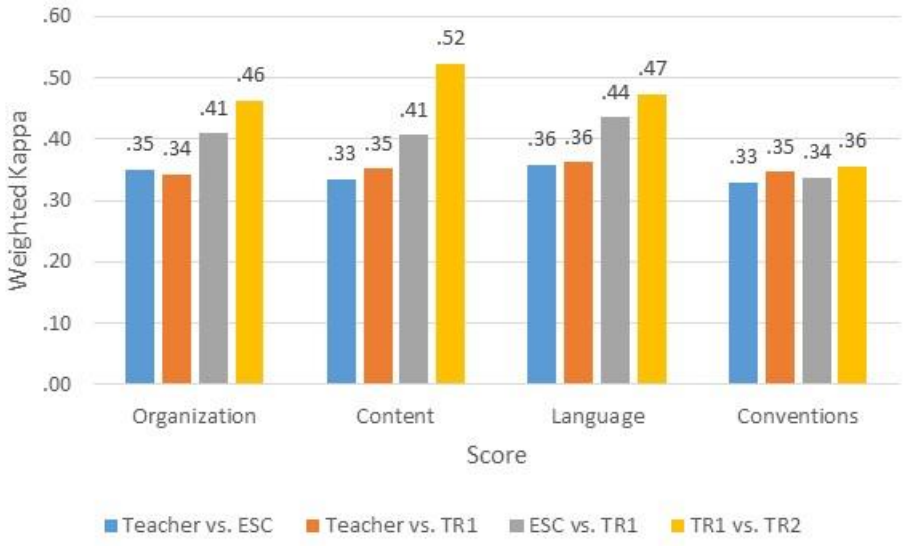


Figure H2.1. Percentage of Exact Agreement between Raters on Total Samples: Grade 7 Writing.

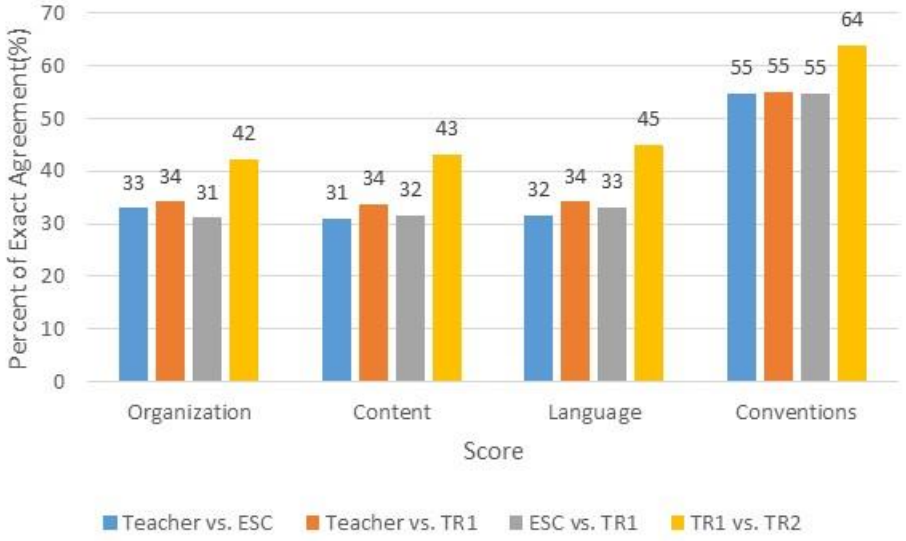


Figure H2.2. Percentage of Adjacent Agreement between Raters on Total Samples: Grade 7 Writing.

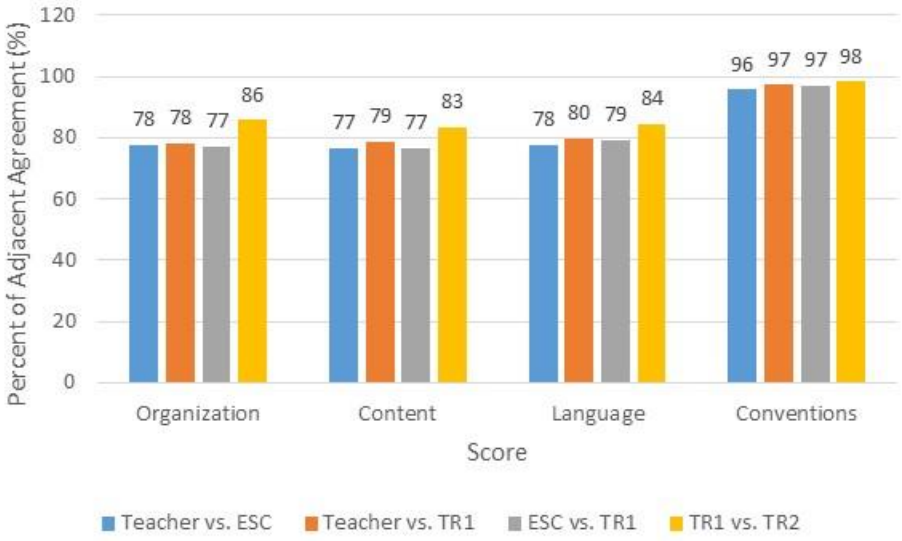

120
100
80
60
40
20
0
Percent of Adjacent Agreement (%)
78 78 77 86
77 79 77 83
78 80 79 84
96 97 97 98
Organization
Content
Language
Conventions
Score
Teacher vs. ESC
Teacher vs. TR1
ESC vs. TR1
TR1 vs. TR2

Figure H2.3. Score Correlation between Raters on Total Samples: Grade 7 Writing.

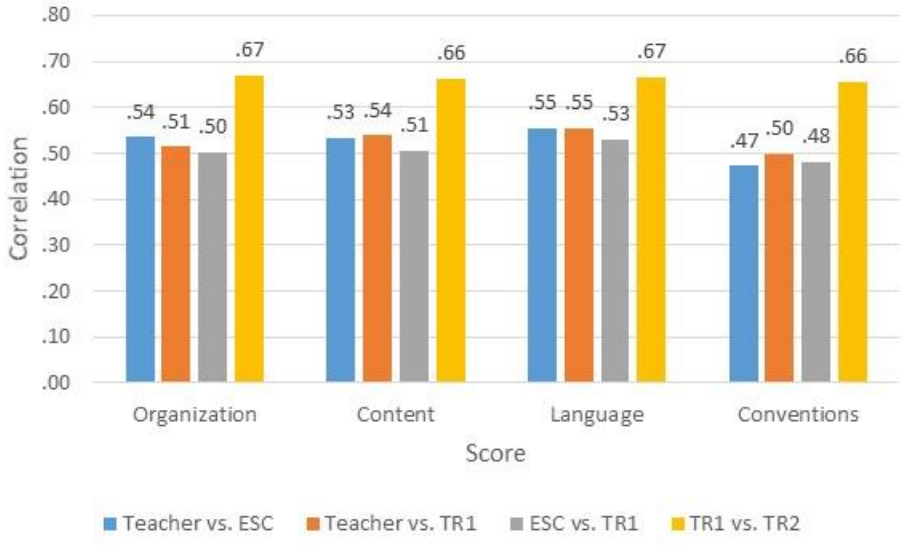


Figure H2.4. Weighted Kappa Coefficient between Raters on Total Samples: Grade 7 Writing.

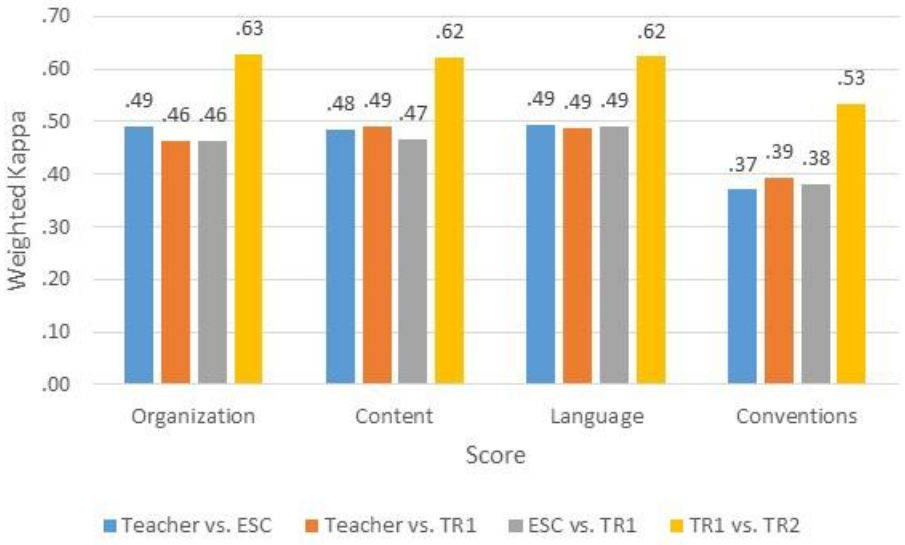


Figure H3.1. Percentage of Exact Agreement between Raters on Total Samples: English I.

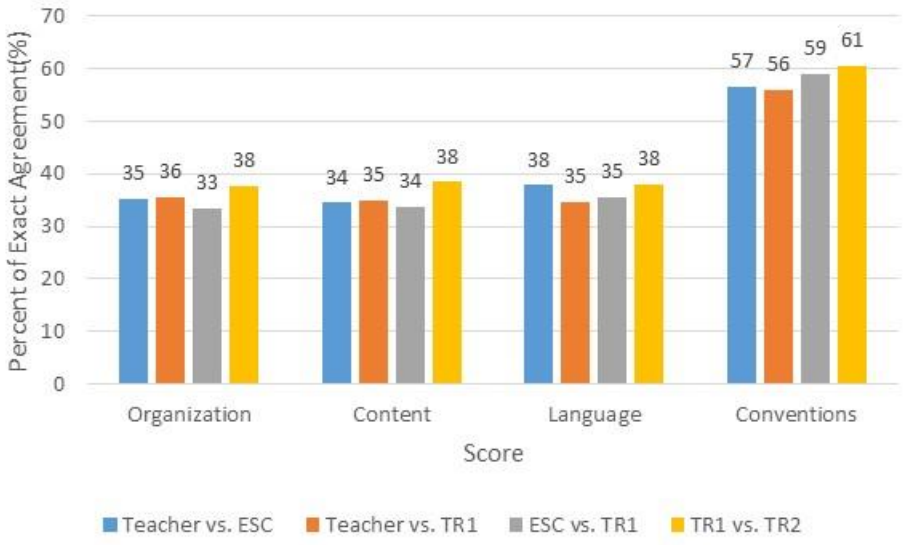


Figure H3.2. Percentage of Adjacent Agreement between Raters on Total Samples: English I.

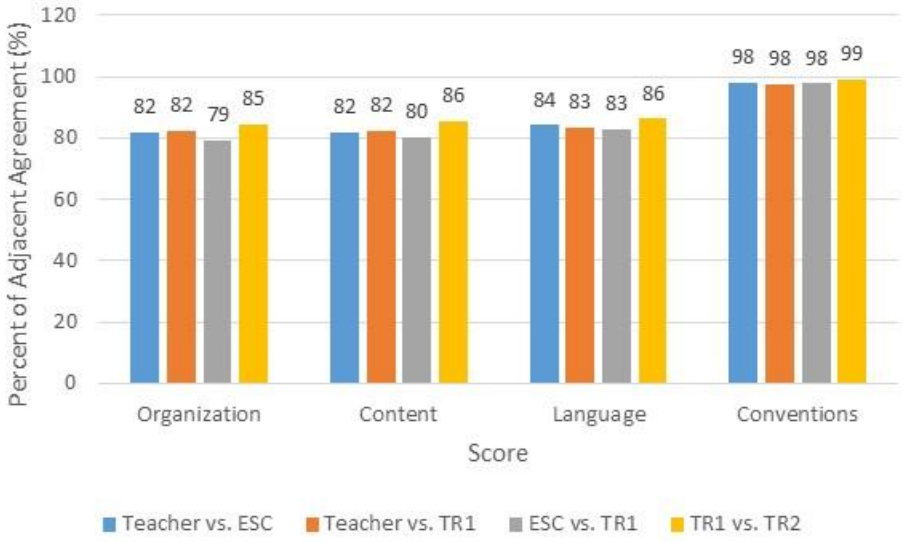


Figure H3.3. Score Correlation between Raters on Total Samples: English I.

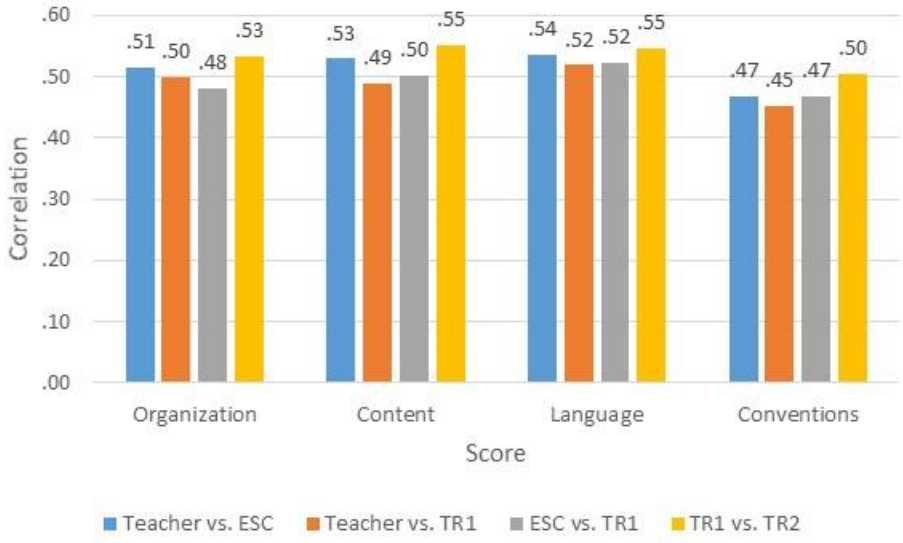


Figure H3.4. Weighted Kappa Coefficient between Raters on Total Samples: English I.

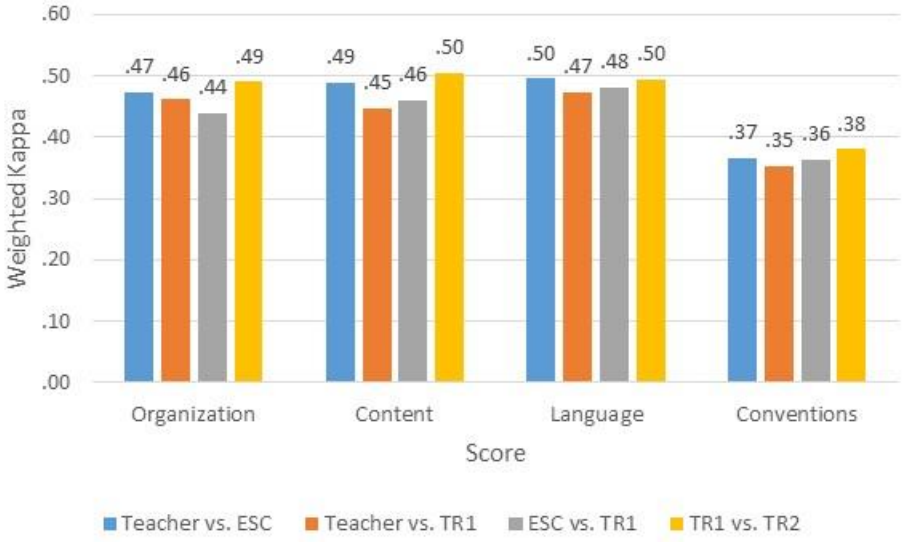


Figure H4.1. Percentage of Exact Agreement between Raters on Total Samples: English II.

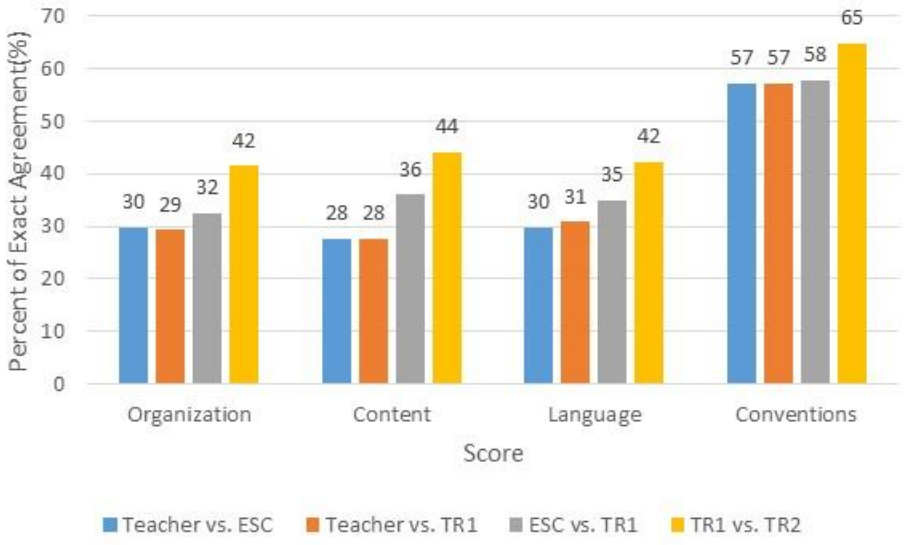


Figure H4.2. Percentage of Adjacent Agreement between Raters on Total Samples: English II.

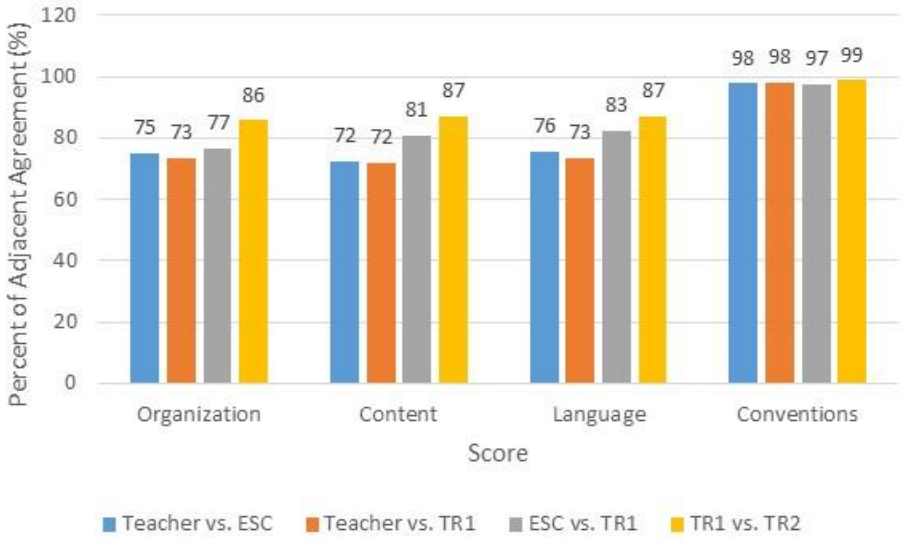


Figure H4.3. Score Correlation between Raters on Total Samples: English II.

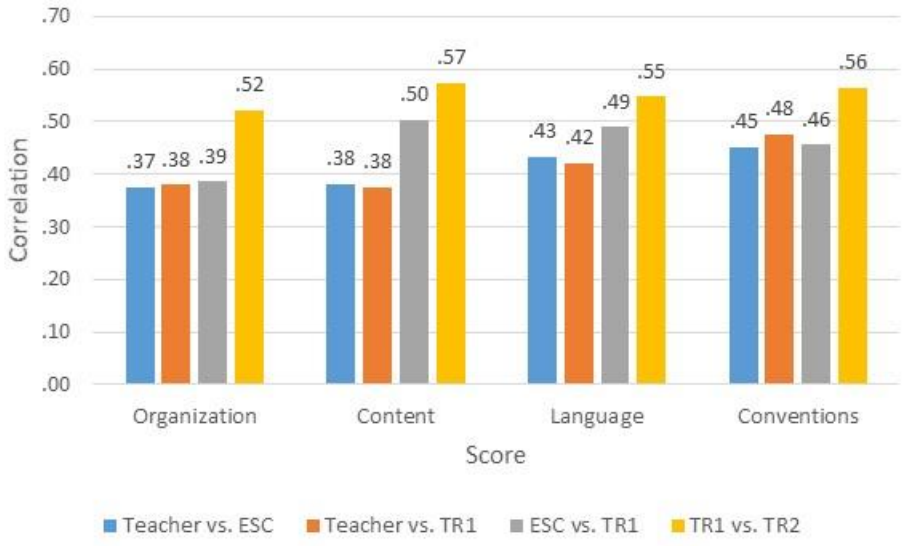
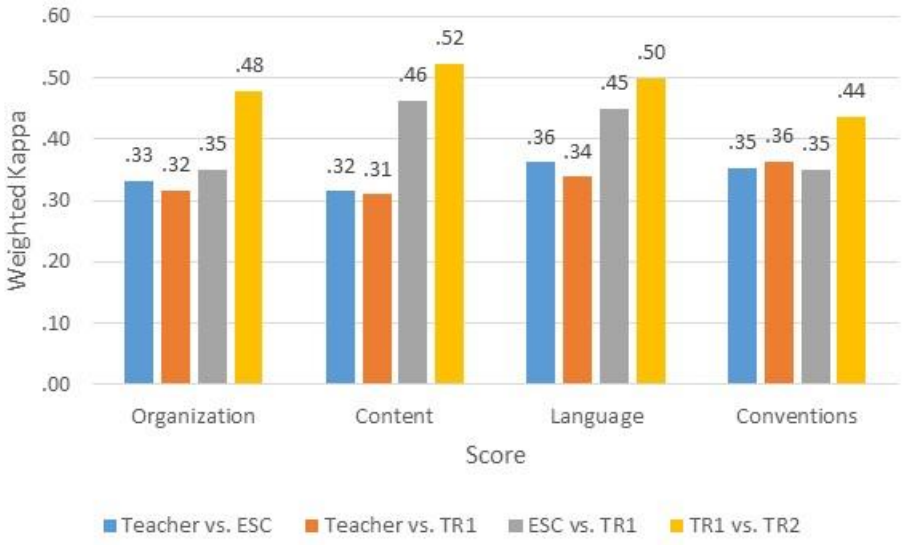


Figure H4.4. Weighted Kappa Coefficient between Raters on Total Samples: English II.



APPENDIX I: RATER SCORE CONSISTENCY SUMMARY

Tables I1–I4 report the sample sizes, percentages of exact agreement (EA), percentages of exact or adjacent agreement (EAA), polychoric correlations (Cor), and quadratic weighted kappa coefficients (WKC) for grades 4 and 7 writing, English I, and English II, respectively, between rating scores

- from ESC raters and from Teachers,
- from Trained Rater 1 and from Teachers,
- from Trained Rater 1 and from ESC raters,
- from Trained Rater 1 and from Trained Rater 2.

These statistics are calculated for each writing sample (TS1, PS1, PS2, and TS2), each writing genre, each timed writing prompt (represented by prompt ID), and the total writing samples with a sample size of at least 30. These statistics are used to examine the extent to which the ratings assigned by teachers, ESC raters, and trained raters are consistent, as rating reliability indicators.

Polychoric correlation is suitable for the case where both variables are ordered categorical variables (Drasgow, 1988[3]), like rating scores in this study. Polychoric correlation assumes there is a continuous variable underlying each categorical variable and the two continuous variables follow a binormal distribution. The polychoric correlation is the correlation between the two variables in the binormal distribution. Polychoric correlation is estimated by the maximum likelihood estimation. Compared to Pearson correlation, polychoric correlation more accurately reflects the true relationship between two ordered categorical variables if the assumptions hold, while Pearson correlation tends to underestimate the association.

The kappa with quadratic weights (Fleiss & Cohen, 1973[4]) is a commonly used weighted kappa statistic for summarizing inter-rater agreement on an ordinal scale. The kappa coefficient (Cohen 1968[5]) is a chance-adjusted index of agreement, which assumes nominal categories. It is extended to non-nominal categories through weighting, which indicates that some categories are more similar than others, and, thus, mismatching pairs of categories deserve varying degrees of partial credit. Quadratic weight is one of the popular ways of determining how much partial credit to assign to each mismatched pair of categories, because the kappa with quadratic weights is equivalent to the intraclass reliability as demonstrated in Fleiss and Cohen, and, thus, it can be interpreted and evaluated as a reliability index.

Based on Tables I1–I4, these statistics had some variations across writing sample groups, rating scores, rater pairs and tests. The two trained raters' scores in general were a little more consistent than the scores from the other rater pairs. The score agreement between Teachers and Trained Rater 1 is closest to that between the two trained raters on English I among the four tests. For example, based on all writing samples the maximum difference on weighted kappa across the four scores between Teachers versus Trained Rater 1 and the two trained raters was 0.17 for grade 4 writing, 0.16 for grade 7 writing, 0.06 for English I, and 0.21 for English II.

---

[3] Drasgow, F. (1988). Polychoric and polyserial correlations. In L. Kotz, & N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*. Vol. 7 (pp. 69-74). New York: Wiley.
[4] Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, 33,* 613–619.
[5] Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220.

Table I1. Rater Score Consistency: Grade 4 Writing

Score	Sample[a]	N[b]	Teacher vs. ESC[c]				Teacher vs. TR1[c]				ESC vs. TR1[c]				TR1 vs. TR2[c]				
			EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	N	EA (%)	EAA (%)	COR	WK
Organization	TS1	603	30	77	.36	.33	31	80	.45	.41	34	81	.50	.45	124	37	85	.53	.49
	PS1	361	32	76	.28	.24	27	77	.23	.17	27	76	.37	.33	62	45	81	.24	.20
	PS2	602	30	77	.42	.38	30	76	.36	.30	31	80	.45	.40	200	45	84	.50	.45
	TS2	603	29	72	.38	.32	28	69	.38	.30	31	79	.40	.37	131	37	80	.50	.47
	Analytic	93	26	70	.57	.43	43	81	.57	.47	22	70	.39	.33	60	45	77	.27	.25
	Expository	526	31	77	.35	.30	30	76	.32	.24	30	78	.44	.39	83	49	88	.60	.54
	Personal Narrative	79	34	75	.49	.44	22	70	.26	.21	27	73	.51	.43	59	37	75	.29	.24
	Persuasive	264	31	80	.45	.38	25	77	.41	.31	32	84	.37	.34	60	45	92	.65	.57
	Personal Narrative_TS	1206	30	74	.38	.35	30	75	.45	.39	32	80	.46	.42	255	37	83	.52	.49
	1000022	523	28	77	.41	.38	30	77	.44	.38	33	81	.48	.43	116	36	82	.42	.39
	1000023	293	32	76	.42	.38	31	73	.45	.39	31	81	.48	.43	52	35	79	.41	.37
	1000024	390	30	70	.33	.29	28	73	.45	.38	33	78	.43	.40	87	39	86	.67	.64
	Total	2169	30	75	.38	.35	29	76	.40	.34	31	79	.45	.41	517	41	83	.50	.46
Content	TS1	603	32	77	.39	.35	35	82	.52	.46	35	81	.52	.46	124	38	90	.55	.50
	PS1	361	27	72	.24	.21	31	73	.25	.18	29	76	.37	.32	62	50	84	.30	.25
	PS2	602	30	73	.38	.33	28	74	.33	.25	33	79	.43	.38	200	44	85	.59	.53
	TS2	603	25	68	.40	.31	30	73	.39	.32	30	79	.42	.38	131	40	82	.59	.55
	Analytic	93	28	80	.55	.44	43	87	.63	.55	25	73	.41	.34	60	37	85	.52	.47
	Expository	526	27	73	.33	.28	29	72	.34	.24	31	79	.44	.39	83	53	90	.65	.56
	Personal Narrative	79	30	66	.46	.39	22	59	.20	.14	32	73	.50	.44	59	36	69	.40	.34
	Persuasive	264	33	73	.35	.28	28	75	.29	.21	34	81	.31	.28	60	53	92	.68	.59
	Personal Narrative_TS	1206	28	73	.40	.35	33	77	.49	.43	33	80	.47	.43	255	39	86	.58	.54
	1000022	523	29	75	.38	.35	34	79	.49	.43	34	82	.48	.44	116	39	86	.48	.43

Score	Sample[a]	N[b]	Teacher vs. ESC[c]				Teacher vs. TR1[c]				ESC vs. TR1[c]				TR1 vs. TR2[c]				
			EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	N	EA (%)	EAA (%)	COR	WK
	1000023	293	29	71	.44	.38	33	77	.50	.44	36	79	.49	.44	52	29	85	.53	.47
	1000024	390	27	70	.40	.34	30	76	.48	.41	29	79	.46	.43	87	45	86	.71	.67
	Total	2169	29	73	.38	.33	31	76	.42	.35	32	79	.45	.41	517	42	85	.56	.52
Language	TS1	603	34	79	.44	.40	35	82	.56	.48	38	82	.54	.49	124	42	90	.65	.59
	PS1	361	27	72	.31	.24	27	71	.29	.19	30	77	.39	.34	62	40	81	.26	.21
	PS2	602	27	73	.39	.34	28	73	.38	.27	30	81	.45	.38	200	37	87	.53	.46
	TS2	603	28	70	.42	.33	29	70	.43	.33	34	82	.47	.44	131	37	77	.45	.43
	Analytic	93	26	76	.57	.48	40	83	.58	.50	23	72	.44	.37	60	33	78	.41	.35
	Expository	526	27	72	.37	.27	26	71	.35	.22	32	81	.48	.42	83	43	90	.59	.50
	Personal Narrative	79	23	71	.47	.43	33	71	.28	.22	23	73	.40	.30	59	37	78	.45	.39
	Persuasive	264	28	72	.40	.32	25	71	.37	.23	30	80	.35	.29	60	35	92	.47	.38
	Personal Narrative_TS	1206	31	74	.44	.39	32	76	.53	.44	36	82	.52	.48	255	39	83	.55	.52
	1000022	523	34	78	.45	.41	33	77	.52	.44	36	82	.54	.49	116	42	89	.54	.48
	1000023	293	34	75	.52	.44	28	76	.58	.48	35	80	.50	.45	52	25	81	.41	.36
	1000024	390	25	69	.39	.32	32	75	.50	.41	37	82	.51	.48	87	44	77	.62	.60
	Total	2169	29	73	.41	.36	30	75	.46	.36	33	81	.48	.44	517	38	84	.51	.47
Conventions	TS1	603	59	98	.48	.37	59	99	.55	.41	58	98	.53	.41	124	60	98	.58	.44
	PS1	361	53	96	.38	.25	49	96	.21	.13	47	96	.30	.23	62	58	97	.22	.17
	PS2	602	54	97	.40	.30	52	97	.31	.23	54	96	.31	.24	200	52	99	.36	.26
	TS2	603	52	97	.41	.30	53	98	.50	.39	54	97	.46	.36	131	50	99	.46	.35
	Analytic	93	57	98	.67	.43	56	98	.51	.34	58	96	.44	.35	60	57	98	.39	.29
	Expository	526	54	96	.39	.27	51	97	.29	.20	47	96	.31	.24	83	55	100	.45	.31
	Personal Narrative	79	44	96	.38	.30	41	94	.21	.16	57	96	.36	.28	59	47	95	.24	.17
	Persuasive	264	55	98	.43	.28	54	97	.27	.18	57	96	.25	.19	60	52	98	.33	.25
	Personal Narrative_TS	1206	56	97	.45	.35	56	98	.56	.43	56	97	.50	.40	255	55	99	.54	.42

Score	Sample[a]	N[b]	Teacher vs. ESC[c]				Teacher vs. TR1[c]				ESC vs. TR1[c]				TR1 vs. TR2[c]				
			EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	N	EA (%)	EAA (%)	COR	WK
	1000022	523	58	97	.46	.35	56	98	.53	.40	57	98	.53	.42	116	59	99	.53	.40
	1000023	293	56	97	.48	.37	60	99	.65	.51	59	97	.48	.37	52	46	100	.47	.37
	1000024	390	52	97	.42	.33	54	97	.53	.41	52	97	.48	.38	87	54	98	.57	.45
	Total	2169	55	97	.43	.33	54	97	.46	.35	54	97	.43	.34	517	54	98	.46	.36

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; The numbers are the prompt IDs of timed samples.

[b] N is the same for the comparisons among Teacher, ESC, and TR1 raters; N for the comparison between TR1 and TR2 raters is different.

[c] EA=Percentage of exact agreement, EAA=Percentage of exact or adjacent agreement, COR=Correlation, WK=Weighted Kappa with quadratic weights.

Table I2. Rater Score Consistency: Grade 7 Writing

Score	Sample[a]	N[b]	Teacher vs. ESC[c]				Teacher vs. TR1[c]				ESC vs. TR1[c]				TR1 vs. TR2[c]				
			EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	N	EA (%)	EAA (%)	COR	WK
Organization	TS1	890	34	79	.53	.49	34	78	.50	.45	31	77	.46	.43	185	38	84	.56	.51
	PS1	383	34	78	.59	.54	35	80	.55	.49	28	74	.52	.47	139	53	86	.79	.74
	PS2	237	27	70	.43	.40	35	77	.54	.49	30	74	.45	.41	99	41	86	.66	.62
	TS2	362	34	80	.58	.48	33	78	.49	.39	35	83	.58	.53	83	36	89	.66	.60
	Analytic	126	36	79	.60	.53	37	79	.54	.47	25	71	.45	.43	59	49	86	.83	.76
	Expository	60	28	70	.45	.41	33	73	.42	.35	27	73	.24	.18	60	55	87	.62	.51
	Personal Narrative	316	31	74	.50	.45	34	79	.48	.43	29	72	.45	.40	60	53	85	.74	.70
	Persuasive	117	31	76	.44	.40	36	81	.57	.53	32	81	.59	.52	59	34	86	.52	.48
	Expository_TS	1252	34	79	.54	.48	34	78	.49	.44	32	78	.49	.45	268	37	85	.58	.54
	1000029	619	34	81	.56	.51	38	82	.54	.49	34	79	.51	.47	118	35	83	.52	.48
	1000030	180	37	79	.55	.49	28	81	.46	.41	26	76	.43	.39	33	45	100	.84	.77
	1000031	453	32	76	.50	.45	31	71	.46	.39	33	79	.50	.46	117	38	84	.56	.52
	Total	1872	33	78	.54	.49	34	78	.51	.46	31	77	.50	.46	506	42	86	.67	.63

Score	Sample[a]	N[b]	Teacher vs. ESC[c]				Teacher vs. TR1[c]				ESC vs. TR1[c]				TR1 vs. TR2[c]				
			EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	N	EA (%)	EAA (%)	COR	WK
Content	TS1	890	33	76	.52	.47	32	77	.50	.45	32	76	.45	.41	185	38	83	.59	.54
	PS1	383	31	80	.61	.56	32	81	.58	.53	28	77	.59	.53	139	53	82	.75	.70
	PS2	237	28	74	.48	.45	37	79	.59	.54	32	75	.52	.49	99	39	84	.69	.65
	TS2	362	28	76	.51	.42	37	80	.54	.47	35	79	.54	.48	83	41	86	.63	.57
	Analytic	126	33	82	.59	.53	38	82	.60	.54	27	76	.52	.50	59	51	85	.82	.75
	Expository	60	30	73	.47	.41	27	75	.54	.43	30	77	.31	.25	60	53	78	.40	.33
	Personal Narrative	316	30	79	.56	.52	33	81	.53	.49	28	74	.52	.48	60	52	80	.70	.66
	Persuasive	117	26	74	.46	.42	37	79	.58	.55	35	81	.61	.56	59	34	88	.64	.59
	Expository_TS	1252	32	76	.51	.45	34	78	.51	.46	33	77	.47	.43	268	39	84	.60	.55
	1000029	619	32	77	.52	.45	37	81	.54	.50	31	76	.44	.40	118	38	85	.57	.52
	1000030	180	28	72	.49	.42	37	83	.58	.51	31	77	.45	.41	33	48	94	.83	.74
	1000031	453	32	76	.53	.46	28	71	.47	.41	35	78	.52	.47	117	38	80	.55	.51
	Total	1872	31	77	.53	.48	34	79	.54	.49	32	77	.51	.47	506	43	83	.66	.62
Language	TS1	890	31	77	.53	.47	34	81	.54	.48	33	79	.48	.44	185	42	86	.59	.53
	PS1	383	35	82	.65	.59	33	80	.61	.52	28	75	.54	.50	139	52	80	.74	.68
	PS2	237	33	78	.54	.50	35	79	.60	.53	34	76	.56	.51	99	40	83	.67	.63
	TS2	362	29	73	.50	.39	35	77	.50	.41	40	85	.61	.56	83	47	90	.72	.65
	Analytic	126	32	80	.58	.52	29	80	.58	.50	31	72	.53	.50	59	51	86	.85	.77
	Expository	60	35	78	.51	.44	43	77	.60	.50	28	67	.18	.14	60	50	77	.40	.34
	Personal Narrative	316	36	82	.62	.56	34	79	.55	.47	29	76	.50	.46	60	53	77	.63	.59
	Persuasive	117	30	79	.55	.51	35	82	.59	.53	33	83	.64	.57	59	34	85	.60	.56
	Expository_TS	1252	31	76	.52	.45	34	80	.53	.46	35	81	.51	.47	268	43	87	.62	.57
	1000029	619	32	77	.52	.45	36	83	.55	.49	38	81	.52	.47	118	39	88	.60	.52
	1000030	180	31	74	.49	.42	32	77	.54	.45	31	78	.48	.43	33	55	91	.83	.71
	1000031	453	29	76	.53	.46	33	77	.51	.43	32	81	.53	.48	117	44	85	.60	.55

Score	Sample[a]	N[b]	Teacher vs. ESC[c]				Teacher vs. TR1[c]				ESC vs. TR1[c]				TR1 vs. TR2[c]				
			EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	N	EA (%)	EAA (%)	COR	WK
	Total	1872	32	78	.55	.49	34	80	.55	.49	33	79	.53	.49	506	45	84	.67	.62
Conventions	TS1	890	54	96	.46	.36	58	98	.51	.40	58	97	.49	.38	185	60	99	.52	.40
	PS1	383	58	98	.60	.49	53	99	.62	.48	52	97	.55	.44	139	72	99	.84	.72
	PS2	237	50	98	.53	.43	52	97	.50	.40	53	96	.44	.35	99	63	97	.61	.50
	TS2	362	54	91	.31	.21	52	95	.33	.25	52	97	.46	.34	83	60	96	.57	.46
	Analytic	126	56	99	.62	.47	54	98	.51	.39	52	98	.53	.43	59	66	100	.82	.67
	Expository	60	63	97	.52	.39	53	100	.59	.42	53	97	.28	.19	60	72	97	.59	.42
	Personal Narrative	316	55	97	.53	.43	52	98	.54	.42	52	97	.49	.38	60	75	98	.80	.67
	Persuasive	117	50	98	.55	.45	54	97	.49	.38	53	97	.46	.36	59	59	98	.56	.44
	Expository_TS	1252	54	94	.41	.31	56	97	.46	.36	56	97	.47	.37	268	60	98	.53	.42
	1000029	619	57	94	.44	.33	55	97	.44	.34	56	97	.49	.37	118	57	96	.39	.31
	1000030	180	49	94	.30	.23	54	97	.41	.31	50	97	.34	.26	33	61	100	.54	.37
	1000031	453	53	96	.42	.32	59	97	.51	.41	59	98	.51	.40	117	63	100	.68	.53
	Total	1872	55	96	.47	.37	55	97	.50	.39	55	97	.48	.38	506	64	98	.66	.53

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; The numbers are the prompt IDs of timed samples.

[b] N is the same for the comparisons among Teacher, ESC, and TR1 raters; N for the comparison between TR1 and TR2 raters is different.

[c] EA=Percentage of exact agreement, EAA=Percentage of exact or adjacent agreement, COR=Correlation, WK=Weighted Kappa with quadratic weights.

Table I3. Rater Score Consistency: English I

Score	Sample[a]	N[b]	Teacher vs. ESC[c]				Teacher vs. TR1[c]				ESC vs. TR1[c]				TR1 vs. TR2[c]				
			EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	N	EA (%)	EAA (%)	COR	WK
Organization	TS1	681	35	82	.51	.45	37	82	.52	.46	33	77	.45	.41	147	36	82	.56	.51
	PS1	289	34	83	.59	.52	35	84	.57	.54	35	79	.58	.52	73	38	81	.63	.57
	PS2	597	37	80	.47	.42	36	85	.49	.44	31	78	.49	.42	165	38	86	.52	.46
	TS2	681	35	83	.52	.46	34	81	.49	.42	35	82	.44	.40	148	39	88	.45	.40

Score	Sample[a]	N[b]	Teacher vs. ESC[c]				Teacher vs. TR1[c]				ESC vs. TR1[c]				TR1 vs. TR2[c]				
			EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	N	EA (%)	EAA (%)	COR	WK
	Analytic	115	32	71	.31	.25	32	76	.32	.25	31	76	.39	.30	59	36	83	.55	.46
	Expository	306	34	85	.58	.53	36	86	.58	.55	36	78	.57	.51	60	40	80	.59	.54
	Other	59	29	73	.45	.39	32	88	.53	.47	31	81	.48	.41	0				
	Personal Narrative	320	41	85	.46	.42	36	87	.49	.44	33	80	.46	.41	60	33	85	.45	.41
	Persuasive	86	36	76	.35	.27	38	79	.30	.26	20	77	.37	.31	59	44	90	.57	.54
	Expository_TS	1362	35	82	.49	.46	36	81	.48	.44	34	80	.44	.41	295	37	85	.51	.47
	1000032	399	31	82	.48	.44	32	79	.41	.38	32	76	.39	.35	81	35	75	.37	.35
	1000033	415	39	85	.57	.53	40	85	.58	.53	35	79	.48	.44	93	40	91	.67	.62
	21000001	548	34	81	.45	.42	35	80	.47	.42	35	83	.45	.41	121	37	86	.47	.42
	Total	2248	35	82	.51	.47	36	82	.50	.46	33	79	.48	.44	533	38	85	.53	.49
Content	TS1	681	39	84	.53	.48	39	83	.53	.49	35	81	.49	.45	147	40	84	.63	.58
	PS1	289	35	81	.62	.56	36	86	.58	.55	33	80	.58	.53	73	40	86	.65	.59
	PS2	597	34	80	.45	.41	31	81	.42	.37	33	79	.50	.45	165	39	88	.53	.47
	TS2	681	30	81	.51	.44	33	81	.45	.39	34	81	.42	.38	148	35	84	.38	.33
	Analytic	115	23	65	.13	.10	30	69	.21	.13	31	76	.47	.39	59	41	86	.59	.48
	Expository	306	38	85	.64	.60	36	88	.59	.56	33	82	.55	.51	60	45	88	.61	.59
	Other	59	36	86	.68	.57	37	90	.53	.45	36	80	.57	.49	0				
	Personal Narrative	320	34	79	.41	.38	29	80	.40	.35	34	80	.47	.42	60	37	85	.48	.41
	Persuasive	86	35	81	.42	.35	35	84	.36	.29	23	74	.38	.33	59	36	90	.57	.51
	Expository_TS	1362	35	83	.51	.47	36	82	.48	.43	34	81	.46	.42	295	38	84	.53	.48
	1000032	399	33	81	.45	.41	35	82	.41	.37	33	79	.40	.37	81	36	74	.32	.30
	1000033	415	40	87	.63	.58	41	84	.57	.53	34	80	.49	.45	93	43	90	.74	.67
	21000001	548	31	80	.47	.43	33	81	.47	.41	36	82	.47	.43	121	35	86	.46	.42
	Total	2248	34	82	.53	.49	35	82	.49	.45	34	80	.50	.46	533	38	86	.55	.50
Language	TS1	681	38	86	.53	.48	37	84	.53	.47	35	82	.50	.46	147	37	85	.63	.56

Score	Sample[a]	N[b]	Teacher vs. ESC[c]				Teacher vs. TR1[c]				ESC vs. TR1[c]				TR1 vs. TR2[c]				
			EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	N	EA (%)	EAA (%)	COR	WK
	PS1	289	39	84	.58	.54	35	84	.57	.54	37	82	.59	.54	73	36	86	.64	.59
	PS2	597	39	80	.46	.42	35	83	.48	.43	36	82	.53	.48	165	39	87	.51	.45
	TS2	681	37	86	.54	.49	32	83	.52	.45	35	84	.47	.43	148	39	87	.40	.36
	Analytic	115	27	73	.25	.21	30	71	.26	.20	37	75	.40	.35	59	42	83	.62	.53
	Expository	306	40	87	.59	.56	37	86	.60	.55	37	84	.57	.52	60	37	90	.65	.59
	Other	59	31	81	.53	.48	36	86	.54	.47	29	78	.41	.37	0				
	Personal Narrative	320	43	81	.45	.42	36	84	.45	.41	40	85	.57	.52	60	38	87	.50	.45
	Persuasive	86	37	78	.36	.32	34	80	.39	.31	23	78	.39	.33	59	36	86	.48	.40
	Expository_TS	1362	37	86	.54	.49	34	84	.50	.46	35	83	.49	.44	295	38	86	.53	.47
	1000032	399	31	86	.50	.46	33	80	.44	.40	33	79	.43	.39	81	36	78	.38	.34
	1000033	415	39	88	.59	.54	36	88	.57	.52	35	84	.52	.48	93	44	91	.73	.66
	21000001	548	40	85	.52	.48	34	84	.51	.45	36	86	.51	.46	121	34	88	.45	.39
	Total	2248	38	84	.54	.50	35	83	.52	.47	35	83	.52	.48	533	38	86	.55	.50
Conventions	TS1	681	55	98	.50	.38	56	97	.49	.38	55	97	.42	.33	147	58	100	.56	.42
	PS1	289	54	98	.55	.44	53	98	.50	.40	59	97	.58	.46	73	62	99	.64	.52
	PS2	597	56	97	.38	.29	59	98	.43	.32	57	98	.46	.36	165	66	99	.56	.41
	TS2	681	59	99	.49	.34	54	97	.44	.32	64	99	.49	.34	148	57	98	.26	.19
	Analytic	115	46	95	.29	.22	48	98	.25	.19	58	97	.53	.41	59	75	98	.73	.61
	Expository	306	55	98	.57	.47	53	97	.48	.36	60	97	.57	.45	60	63	100	.65	.47
	Other	59	51	97	.35	.24	66	100	.65	.46	49	95	.11	.07	0				
	Personal Narrative	320	63	99	.42	.30	63	99	.45	.32	60	99	.49	.37	60	55	100	.39	.27
	Persuasive	86	49	92	.22	.15	53	98	.30	.21	48	98	.37	.28	59	66	98	.48	.36
	Expository_TS	1362	57	98	.49	.37	55	97	.45	.35	60	98	.44	.33	295	57	99	.43	.32
	1000032	399	51	98	.41	.30	53	97	.42	.33	58	98	.42	.32	81	59	100	.48	.33
	1000033	415	59	98	.56	.44	58	97	.50	.40	59	97	.44	.34	93	61	100	.60	.44

Score	Sample[a]	N[b]	Teacher vs. ESC[c]				Teacher vs. TR1[c]				ESC vs. TR1[c]				TR1 vs. TR2[c]				
			EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	N	EA (%)	EAA (%)	COR	WK
	21000001	548	60	98	.48	.36	55	98	.42	.32	61	99	.46	.34	121	53	98	.27	.20
	Total	2248	57	98	.47	.37	56	98	.45	.35	59	98	.47	.36	533	61	99	.50	.38

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; The numbers are the prompt IDs of timed samples.

[b] N is the same for the comparisons among Teacher, ESC, and TR1 raters; N for the comparison between TR1 and TR2 raters is different.

[c] EA=Percentage of exact agreement, EAA=Percentage of exact or adjacent agreement, COR=Correlation, WK=Weighted Kappa with quadratic weights.

Table I4. Rater Score Consistency: English II

Score	Sample[a]	N[b]	Teacher vs. ESC[c]				Teacher vs. TR1[c]				ESC vs. TR1[c]				TR1 vs. TR2[c]				
			EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	N	EA (%)	EAA (%)	COR	WK
Organization	TS1	469	27	72	.32	.29	33	78	.38	.34	30	76	.42	.37	102	40	79	.44	.40
	PS1	431	28	73	.37	.29	31	74	.38	.31	36	79	.41	.37	77	43	91	.67	.61
	PS2	428	31	78	.43	.37	29	76	.32	.26	31	75	.30	.27	163	45	88	.48	.43
	TS2	387	34	78	.51	.42	24	64	.49	.34	33	78	.45	.38	65	32	83	.55	.49
	Analytic	160	34	74	.32	.26	38	81	.34	.30	34	75	.39	.32	60	38	88	.57	.51
	Expository	183	28	73	.45	.36	32	73	.50	.39	33	78	.51	.45	60	43	88	.62	.57
	Other	161	20	64	.28	.19	26	63	.31	.23	29	76	.31	.27	0				
	Personal Narrative	98	31	80	.53	.48	28	70	.15	.15	30	73	.26	.22	60	48	87	.28	.25
	Persuasive	257	33	84	.45	.36	26	81	.40	.28	37	78	.25	.22	60	48	93	.52	.43
	Persuasive_TS	856	30	74	.37	.34	29	72	.38	.32	32	77	.43	.38	167	37	81	.49	.45
	1000035	284	30	76	.44	.40	31	71	.39	.32	35	80	.48	.43	62	37	81	.36	.31
	1000036	308	27	70	.28	.25	28	73	.37	.30	30	76	.41	.36	59	41	81	.60	.55
	1000037	264	34	77	.39	.35	28	72	.38	.33	30	74	.39	.33	46	33	80	.53	.47
	Total	1715	30	75	.37	.33	29	73	.38	.32	32	77	.39	.35	407	42	86	.52	.48
Content	TS1	469	26	69	.30	.27	28	74	.33	.29	31	78	.48	.45	102	41	81	.54	.48

Score	Sample[a]	N[b]	Teacher vs. ESC[c]				Teacher vs. TR1[c]				ESC vs. TR1[c]				TR1 vs. TR2[c]				
			EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	N	EA (%)	EAA (%)	COR	WK
	PS1	431	29	71	.32	.26	32	73	.39	.33	38	78	.47	.43	77	49	88	.64	.57
	PS2	428	32	80	.44	.36	29	73	.27	.21	39	84	.47	.41	163	47	91	.57	.51
	TS2	387	24	69	.53	.37	22	66	.51	.36	36	83	.56	.52	65	34	83	.55	.49
	Analytic	160	41	91	.58	.52	35	86	.39	.34	36	84	.45	.39	60	42	93	.65	.56
	Expository	183	28	70	.36	.31	29	74	.49	.41	36	80	.58	.53	60	55	88	.69	.62
	Other	161	20	58	.18	.12	27	61	.32	.23	34	74	.35	.32	0				
	Personal Narrative	98	40	82	.55	.51	27	77	.20	.16	37	76	.27	.23	60	43	87	.40	.36
	Persuasive	257	28	77	.45	.32	31	70	.34	.20	46	86	.42	.35	60	52	93	.43	.33
	Persuasive_TS	856	25	69	.38	.31	25	71	.38	.32	33	80	.52	.48	167	38	82	.54	.49
	1000035	284	26	70	.43	.35	27	72	.40	.33	34	82	.52	.48	62	35	87	.48	.40
	1000036	308	23	68	.27	.22	25	72	.39	.31	35	81	.52	.48	59	44	81	.62	.56
	1000037	264	27	70	.42	.35	23	67	.36	.31	30	78	.51	.48	46	35	76	.55	.50
	Total	1715	28	72	.38	.32	28	72	.38	.31	36	81	.50	.46	407	44	87	.57	.52
Language	TS1	469	27	74	.36	.33	34	75	.38	.34	33	79	.49	.45	102	39	80	.43	.39
	PS1	431	29	74	.37	.29	33	75	.42	.34	37	85	.44	.40	77	55	91	.64	.58
	PS2	428	35	83	.53	.44	33	81	.42	.33	38	82	.44	.39	163	40	90	.57	.50
	TS2	387	29	72	.57	.41	22	61	.51	.33	31	85	.54	.49	65	38	86	.51	.46
	Analytic	160	43	91	.69	.59	37	91	.51	.46	39	81	.46	.39	60	38	92	.67	.57
	Expository	183	27	69	.46	.32	29	73	.57	.42	40	85	.57	.50	60	58	88	.63	.57
	Other	161	20	65	.24	.17	30	66	.34	.26	27	82	.29	.27	0				
	Personal Narrative	98	29	82	.49	.44	33	77	.23	.21	39	84	.41	.36	60	30	88	.43	.40
	Persuasive	257	37	83	.47	.37	36	81	.41	.27	42	84	.38	.32	60	52	92	.46	.39
	Persuasive_TS	856	28	73	.43	.36	29	69	.40	.32	32	82	.51	.47	167	39	83	.48	.44
	1000035	284	29	74	.49	.40	31	70	.44	.35	30	81	.52	.48	62	35	85	.34	.28
	1000036	308	25	71	.34	.29	31	71	.41	.32	34	82	.50	.45	59	49	85	.68	.62

Score	Sample[a]	N[b]	Teacher vs. ESC[c]				Teacher vs. TR1[c]				ESC vs. TR1[c]				TR1 vs. TR2[c]				
			EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	N	EA (%)	EAA (%)	COR	WK
	1000037	264	30	74	.45	.37	23	66	.35	.28	31	81	.52	.48	46	30	76	.38	.33
	Total	1715	30	76	.43	.36	31	73	.42	.34	35	83	.49	.45	407	42	87	.55	.50
Conventions	TS1	469	55	97	.44	.34	54	97	.39	.30	58	96	.52	.40	102	64	99	.58	.44
	PS1	431	55	98	.37	.27	58	98	.49	.38	61	98	.35	.26	77	74	99	.73	.60
	PS2	428	63	98	.44	.34	65	99	.51	.37	61	99	.41	.30	163	62	100	.45	.31
	TS2	387	57	98	.57	.45	51	97	.60	.40	50	98	.52	.38	65	63	95	.49	.37
	Analytic	160	66	97	.49	.37	64	99	.57	.43	61	98	.43	.31	60	65	100	.64	.45
	Expository	183	46	97	.47	.35	55	99	.70	.52	59	98	.50	.39	60	73	100	.82	.64
	Other	161	58	99	.16	.10	55	97	.15	.11	60	98	.08	.05	0				
	Personal Narrative	98	57	99	.42	.30	61	98	.23	.15	57	99	.27	.18	60	58	100	.31	.19
	Persuasive	257	64	98	.39	.28	70	100	.49	.33	65	98	.38	.27	60	67	98	.22	.16
	Persuasive_TS	856	56	98	.49	.39	53	97	.45	.34	54	97	.52	.39	167	63	98	.54	.43
	1000035	284	56	98	.51	.41	51	96	.38	.29	58	98	.61	.45	62	69	98	.55	.42
	1000036	308	54	98	.39	.30	53	99	.49	.35	54	97	.47	.35	59	66	98	.64	.52
	1000037	264	58	97	.55	.45	53	97	.49	.37	51	96	.47	.37	46	52	96	.36	.28
	Total	1715	57	98	.45	.35	57	98	.48	.36	58	97	.46	.35	407	65	99	.56	.44

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; The numbers are the prompt IDs of timed samples.

[b] N is the same for the comparisons among Teacher, ESC, and TR1 raters; N for the comparison between TR1 and TR2 raters is different.

[c] EA=Percentage of exact agreement, EAA=Percentage of exact or adjacent agreement, COR=Correlation, WK=Weighted Kappa with quadratic weights.

APPENDIX J: RATER SCORE CONSISTENCY BY CLASS

Because the students in a class took the same set of writing prompts, the percentages of exact agreement (EA), percentages of exact or adjacent agreement (EAA), polychoric correlations (Cor), and quadratic weighted kappa coefficients (WKC) were calculated at the class level for each writing sample score in each test among rating scores from the three raters—Teacher, ESC rater, and Trained Rater 1—for each class with a sample size of at least 30. Polychoric correlation was not calculated for Conventions score at the class level because sometimes it might not be stable with a small sample size. Tables J1–J4 report the summaries (N, mean, standard deviation, max, and min) of these statistics across classes. These statistics varied across classes and have some variations across writing samples, scores, rater pairs, and tests. Overall, the range of exact agreement at the class level was between 3% and 66% (except for Conventions score), the range of adjacent agreement was between 23% and 100% (except for Conventions score), the range of polychoric correlations was between -0.47 and 0.84 (except for Conventions score), and the range of kappa coefficients was between -0.39 and 0.71. For Conventions score, the range of exact agreement was between 14% and 79% and the range of adjacent agreement was between 52% and 100%.

Table J1. Summary of Rater Score Consistency by Class: Grade 4 Writing

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
Organization	TS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	34	28	73	.34	.26	29	79	.53	.39	37	83	.47	.41
		StdDev	4	9	12	.11	.15	7	8	.15	.14	12	7	.12	.11
		Max	40	38	85	.49	.46	40	93	.76	.63	50	91	.70	.60
		Min	30	18	55	.17	.05	23	73	.31	.24	20	70	.37	.26
	PS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	34	34	76	.36	.25	28	80	.29	.18	26	73	.24	.19
		StdDev	4	8	15	.26	.21	6	15	.21	.16	12	5	.18	.15
		Max	40	45	94	.66	.48	37	89	.56	.41	39	80	.37	.30
		Min	30	25	53	.03	.01	20	50	.03	.00	10	68	-.12	-.10
	PS2	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	34	31	79	.37	.23	37	84	.50	.31	33	85	.44	.33
		StdDev	4	7	8	.19	.13	20	8	.12	.10	10	6	.12	.09

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
		Max	40	37	87	.56	.39	66	94	.65	.44	42	93	.60	.47
		Min	30	22	69	.05	.03	12	73	.32	.15	19	78	.27	.24
	TS2	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	34	28	77	.50	.32	24	69	.52	.28	34	82	.43	.35
		StdDev	4	15	17	.12	.12	12	13	.23	.11	7	5	.17	.15
		Max	40	53	92	.71	.45	44	88	.81	.41	43	87	.70	.57
		Min	30	10	47	.40	.12	10	53	.16	.11	25	76	.24	.20
Content	TS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	34	28	70	.40	.29	33	78	.52	.40	33	78	.35	.30
		StdDev	4	11	10	.10	.14	11	8	.18	.16	11	9	.21	.16
		Max	40	48	85	.59	.52	45	90	.70	.57	48	91	.58	.45
		Min	30	18	57	.32	.12	20	71	.29	.14	19	68	.09	.09
	PS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	34	28	77	.42	.29	32	74	.31	.23	24	74	.27	.22
		StdDev	4	7	9	.21	.19	11	15	.21	.16	14	10	.14	.11
		Max	40	39	88	.66	.55	47	89	.59	.46	45	83	.49	.36
		Min	30	18	63	.07	-.01	18	50	-.02	.01	7	61	.07	.03
	PS2	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	34	30	74	.35	.21	25	76	.30	.19	36	83	.32	.24
		StdDev	4	9	7	.31	.15	11	14	.26	.16	8	6	.18	.13
		Max	40	39	84	.73	.33	41	88	.69	.40	48	90	.50	.41
		Min	30	18	67	-.17	-.05	7	58	.02	.00	28	74	.05	.06
	TS2	N	6	6	6	6	6	6	6	6	6	6	6	6	6

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
		Mean	34	24	66	.49	.28	29	71	.53	.28	30	82	.41	.34
		StdDev	4	15	19	.19	.13	10	11	.25	.16	10	7	.30	.25
		Max	40	47	89	.66	.47	44	88	.69	.45	47	93	.78	.67
		Min	30	3	40	.21	.09	17	60	.02	-.01	18	73	-.14	-.07
Language	TS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	34	28	76	.48	.30	31	80	.59	.41	41	85	.53	.45
		StdDev	4	7	9	.15	.11	13	9	.16	.12	4	6	.20	.17
		Max	40	34	85	.69	.45	50	88	.78	.58	45	94	.72	.65
		Min	30	18	60	.31	.18	18	67	.39	.23	34	78	.18	.18
	PS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	34	26	73	.34	.21	26	69	.28	.16	29	75	.28	.23
		StdDev	4	11	13	.24	.18	11	18	.19	.14	9	7	.14	.11
		Max	40	41	88	.59	.41	41	84	.54	.33	38	83	.45	.38
		Min	30	15	53	-.07	-.03	10	33	-.04	-.01	17	63	.10	.08
	PS2	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	34	29	73	.41	.21	23	68	.37	.16	36	83	.41	.30
		StdDev	4	15	12	.20	.05	12	20	.21	.10	12	9	.20	.17
		Max	40	47	89	.81	.28	38	84	.65	.30	50	95	.70	.57
		Min	30	10	53	.25	.14	6	37	.09	.03	16	72	.15	.10
	TS2	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	34	21	65	.55	.29	24	64	.54	.26	32	81	.41	.33
		StdDev	4	13	28	.14	.13	17	19	.24	.15	5	6	.22	.19
		Max	40	39	97	.73	.47	45	88	.79	.45	40	93	.65	.56

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
		Min	30	3	23	.36	.08	7	40	.08	.05	28	75	.01	.00
Conventions	TS1	N	6	6	6		6	6	6		6	6	6		6
		Mean	34	56	99		.26	53	99		.28	61	98		.37
		StdDev	4	11	1		.22	13	1		.15	11	3		.17
		Max	40	75	100		.65	72	100		.47	76	100		.65
		Min	30	42	97		.03	30	97		.02	47	92		.18
	PS1	N	6	6	6		6	6	6		6	6	6		6
		Mean	34	53	98		.18	48	96		.10	42	96		.10
		StdDev	4	5	4		.21	15	7		.17	14	4		.25
		Max	40	60	100		.44	66	100		.34	63	100		.34
		Min	30	47	90		-.10	30	83		-.07	23	90		-.34
	PS2	N	6	6	6		6	6	6		6	6	6		6
		Mean	34	61	99		.21	56	98		.11	56	96		.14
		StdDev	4	15	2		.19	13	3		.12	12	4		.26
		Max	40	75	100		.52	69	100		.26	73	100		.53
		Min	30	37	97		.00	33	93		-.03	40	90		-.26
	TS2	N	6	6	6		6	6	6		6	6	6		6
		Mean	34	50	99		.13	54	100		.18	53	98		.21
		StdDev	4	13	3		.04	9	0		.13	9	3		.19
		Max	40	66	100		.19	66	100		.33	64	100		.37
		Min	30	36	93		.10	40	100		.02	39	93		-.11

[a] TS1=Timed Sample 1, PS1=Process Sample 1, PS2=Process Sample 2, TS2=Timed Sample 2.

[b] EA=Percentage of exact agreement, EAA=Percentage of exact or adjacent agreement, COR=Correlation, WK=Weighted Kappa with quadratic weights.

Table J2. Summary of Rater Score Consistency by Class: Grade 7 Writing

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
Organization	TS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	44	32	80	.46	.38	31	76	.43	.32	34	80	.36	.32
		StdDev	18	9	8	.10	.11	12	10	.07	.08	8	4	.26	.23
		Max	79	42	88	.61	.48	52	90	.52	.42	47	85	.62	.55
		Min	31	21	70	.36	.23	19	62	.30	.20	25	75	-.13	-.09
	PS1	N	2	2	2	2	2	2	2	2	2	2	2	2	2
		Mean	55	35	80	.50	.42	36	90	.56	.46	32	78	.44	.36
		StdDev	33	3	4	.05	.08	9	5	.05	.00	4	1	.14	.10
		Max	78	37	83	.53	.47	42	94	.60	.46	35	79	.54	.43
		Min	31	32	77	.46	.36	29	86	.52	.46	29	77	.34	.29
	TS2	N	3	3	3	3	3	3	3	3	3	3	3	3	3
		Mean	41	25	73	.47	.25	32	74	.52	.30	38	85	.48	.41
		StdDev	2	11	11	.21	.02	10	8	.07	.06	5	5	.13	.11
		Max	44	38	83	.72	.27	38	83	.58	.36	41	91	.62	.53
		Min	40	16	61	.32	.22	20	66	.45	.24	33	80	.37	.33
Content	TS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	44	35	78	.51	.39	31	72	.40	.31	35	80	.39	.33
		StdDev	18	5	8	.12	.13	14	11	.07	.05	5	6	.21	.17
		Max	79	45	90	.65	.55	58	87	.49	.38	40	87	.53	.46
		Min	31	29	64	.33	.23	19	57	.29	.24	26	71	.03	.03
	PS1	N	2	2	2	2	2	2	2	2	2	2	2	2	2

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
		Mean	55	32	86	.63	.52	31	82	.42	.35	33	74	.49	.41
		StdDev	33	1	7	.01	.03	7	7	.19	.19	13	5	.06	.01
		Max	78	32	90	.63	.54	36	87	.56	.49	42	77	.53	.42
		Min	31	31	81	.62	.50	26	77	.28	.22	23	71	.44	.41
	TS2	N	3	3	3	3	3	3	3	3	3	3	3	3	3
		Mean	41	21	68	.46	.27	30	76	.60	.41	37	79	.37	.30
		StdDev	2	11	13	.11	.03	11	9	.09	.09	7	4	.11	.08
		Max	44	30	83	.58	.30	43	83	.69	.47	45	83	.50	.40
		Min	40	9	57	.39	.24	23	66	.52	.30	33	75	.29	.24
Language	TS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	44	32	79	.62	.43	32	75	.43	.31	36	87	.48	.41
		StdDev	18	12	11	.16	.11	10	14	.12	.12	4	7	.13	.11
		Max	79	45	94	.78	.62	42	94	.59	.45	42	94	.69	.58
		Min	31	12	68	.44	.31	17	55	.25	.17	32	77	.33	.30
	PS1	N	2	2	2	2	2	2	2	2	2	2	2	2	2
		Mean	55	43	89	.71	.60	33	81	.46	.36	31	81	.52	.44
		StdDev	33	8	2	.14	.09	1	4	.00	.00	6	4	.09	.10
		Max	78	48	90	.81	.67	33	84	.47	.36	35	84	.59	.51
		Min	31	37	87	.61	.53	32	78	.46	.36	27	78	.46	.37
	TS2	N	3	3	3	3	3	3	3	3	3	3	3	3	3
		Mean	41	18	58	.47	.20	21	63	.61	.30	42	86	.47	.39
		StdDev	2	10	12	.09	.03	7	16	.13	.03	4	3	.09	.09
		Max	44	25	68	.56	.23	28	73	.76	.33	45	89	.54	.47

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
		Min	40	7	45	.39	.17	14	45	.51	.28	38	83	.37	.30
Conventions	TS1	N	6	6	6		6	6	6		6	6	6		6
		Mean	44	56	97		.33	57	97		.32	61	98		.31
		StdDev	18	12	6		.11	12	6		.16	5	3		.17
		Max	79	70	100		.45	71	100		.53	68	100		.47
		Min	31	36	86		.17	40	86		.12	53	92		-.02
	PS1	N	2	2	2		2	2	2		2	2	2		2
		Mean	55	61	98		.43	56	99		.34	48	97		.30
		StdDev	33	9	3		.00	7	1		.03	15	0		.04
		Max	78	68	100		.44	61	100		.36	58	97		.33
		Min	31	55	96		.43	51	99		.32	37	97		.27
	TS2	N	3	3	3		3	3	3		3	3	3		3
		Mean	41	44	80		.17	45	92		.25	49	96		.16
		StdDev	2	27	24		.12	18	11		.09	8	1		.08
		Max	44	68	98		.29	60	100		.36	55	98		.23
		Min	40	14	52		.05	25	80		.20	40	95		.08

[a] TS1=Timed Sample 1, PS1=Process Sample 1, PS2=Process Sample 2, TS2=Timed Sample 2.

[b] EA=Percentage of exact agreement, EAA=Percentage of exact or adjacent agreement, COR=Correlation, WK=Weighted Kappa with quadratic weights.

Table J3. Summary of Rater Score Consistency by Class: English I

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
Organization	TS1	N	8	8	8	8	8	8	8	8	8	8	8	8	8
		Mean	78	32	80	.39	.31	36	80	.41	.33	32	76	.32	.29
		StdDev	41	7	8	.22	.20	7	8	.19	.18	10	6	.34	.30
		Max	148	42	91	.65	.59	47	89	.73	.66	44	86	.61	.56
		Min	34	24	67	.00	.00	26	68	.21	.13	12	68	-.47	-.39
	PS1	N	5	5	5	5	5	5	5	5	5	5	5	5	5
		Mean	52	36	82	.35	.27	34	83	.46	.33	36	78	.39	.32
		StdDev	19	7	8	.31	.21	10	4	.08	.09	3	7	.18	.15
		Max	81	44	90	.62	.45	49	89	.56	.46	41	89	.63	.56
		Min	34	25	71	-.13	-.06	22	78	.38	.21	31	69	.19	.17
	PS2	N	8	8	8	8	8	8	8	8	8	8	8	8	8
		Mean	70	34	77	.41	.32	36	84	.47	.37	28	77	.33	.26
		StdDev	44	11	7	.22	.16	9	9	.14	.13	9	8	.25	.21
		Max	148	47	91	.75	.56	49	100	.77	.62	40	84	.58	.52
		Min	30	13	69	.00	.01	23	73	.33	.20	17	60	-.24	-.21
	TS2	N	8	8	8	8	8	8	8	8	8	8	8	8	8
		Mean	78	34	83	.56	.42	33	81	.48	.38	35	82	.42	.36
		StdDev	40	8	7	.14	.13	5	7	.12	.10	6	7	.17	.14
		Max	142	46	91	.77	.59	41	94	.65	.56	44	94	.80	.67
		Min	33	20	68	.35	.27	25	71	.25	.25	25	73	.25	.22
Content	TS1	N	8	8	8	8	8	8	8	8	8	8	8	8	8

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
		Mean	78	38	82	.42	.35	38	83	.39	.34	35	79	.38	.34
		StdDev	41	6	8	.20	.18	9	7	.24	.22	4	7	.24	.21
		Max	148	49	91	.72	.64	51	95	.73	.66	43	90	.63	.57
		Min	34	30	65	.12	.09	26	71	.15	.13	29	68	-.17	-.14
	PS1	N	5	5	5	5	5	5	5	5	5	5	5	5	5
		Mean	52	32	79	.40	.29	34	84	.41	.28	32	79	.38	.29
		StdDev	19	13	12	.18	.18	8	7	.16	.16	7	9	.17	.16
		Max	81	44	94	.66	.54	43	93	.65	.55	41	90	.63	.57
		Min	34	15	62	.16	.05	25	74	.22	.15	25	69	.19	.14
	PS2	N	8	8	8	8	8	8	8	8	8	8	8	8	8
		Mean	70	34	79	.44	.35	34	83	.43	.33	31	79	.36	.30
		StdDev	44	3	10	.20	.14	7	11	.13	.12	6	6	.21	.17
		Max	148	38	90	.70	.56	46	92	.58	.46	39	88	.56	.50
		Min	30	30	66	.08	.05	25	61	.18	.14	21	69	-.03	-.03
	TS2	N	8	8	8	8	8	8	8	8	8	8	8	8	8
		Mean	78	30	82	.57	.43	34	82	.48	.37	34	81	.38	.33
		StdDev	40	7	8	.13	.13	8	8	.12	.11	5	5	.18	.17
		Max	142	40	92	.74	.64	51	94	.65	.60	40	87	.71	.66
		Min	33	18	70	.31	.20	25	71	.32	.27	26	71	.17	.13
Language	TS1	N	8	8	8	8	8	8	8	8	8	8	8	8	8
		Mean	78	37	85	.45	.36	37	82	.38	.31	35	82	.39	.34
		StdDev	41	5	4	.20	.16	7	8	.23	.19	5	5	.24	.21
		Max	148	43	90	.71	.64	47	91	.72	.64	41	89	.62	.55

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
		Min	34	29	78	.20	.15	28	65	.07	.08	29	74	-.16	-.13
	PS1	N	5	5	5	5	5	5	5	5	5	5	5	5	5
		Mean	52	38	85	.36	.27	35	83	.42	.31	36	81	.31	.27
		StdDev	19	9	9	.24	.19	9	9	.21	.19	6	7	.17	.16
		Max	81	49	92	.63	.50	49	90	.64	.51	42	89	.54	.50
		Min	34	26	73	-.01	-.01	24	71	.19	.11	30	74	.10	.10
	PS2	N	8	8	8	8	8	8	8	8	8	8	8	8	8
		Mean	70	36	79	.40	.32	35	83	.47	.36	34	80	.34	.29
		StdDev	44	7	5	.24	.20	4	8	.10	.09	8	7	.25	.22
		Max	148	53	89	.72	.50	41	91	.60	.45	45	88	.67	.61
		Min	30	28	74	-.05	-.05	27	67	.33	.21	23	69	-.13	-.12
	TS2	N	8	8	8	8	8	8	8	8	8	8	8	8	8
		Mean	78	33	86	.60	.46	31	84	.52	.41	35	84	.45	.39
		StdDev	40	10	7	.12	.10	6	6	.13	.11	5	5	.12	.10
		Max	142	43	97	.78	.65	40	97	.71	.59	41	89	.62	.51
		Min	33	12	73	.45	.34	22	76	.32	.29	27	75	.22	.19
Conventions	TS1	N	8	8	8		8	8	8		8	8	8		8
		Mean	78	53	98		.28	56	97		.31	54	96		.24
		StdDev	41	10	4		.18	10	3		.22	7	3		.18
		Max	148	69	100		.59	71	100		.56	69	100		.53
		Min	34	41	88		.07	44	94		.00	48	92		.05
	PS1	N	5	5	5		5	5	5		5	5	5		5
		Mean	52	51	97		.15	53	98		.18	57	96		.20

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
		StdDev	19	14	3		.14	8	1		.09	8	2		.14
		Max	81	69	100		.33	62	100		.29	67	98		.44
		Min	34	38	93		-.04	42	96		.07	46	94		.10
	PS2	N	8	8	8		8	8	8		8	8	8		8
		Mean	70	52	96		.23	57	98		.23	55	97		.20
		StdDev	44	10	3		.15	9	3		.14	8	2		.21
		Max	148	66	99		.41	67	100		.45	63	100		.44
		Min	30	40	90		-.03	43	90		.05	40	94		-.14
	TS2	N	8	8	8		8	8	8		8	8	8		8
		Mean	78	60	98		.29	55	98		.27	66	99		.29
		StdDev	40	15	2		.12	14	3		.13	7	1		.13
		Max	142	77	100		.50	71	100		.47	79	100		.49
		Min	33	35	94		.14	33	90		.09	55	97		.13

[a] TS1=Timed Sample 1, PS1=Process Sample 1, PS2=Process Sample 2, TS2=Timed Sample 2.

[b] EA=Percentage of exact agreement, EAA=Percentage of exact or adjacent agreement, COR=Correlation, WK=Weighted Kappa with quadratic weights.

Table J4. Summary of Rater Score Consistency by Class: English II

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
Organization	TS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	69	28	72	.34	.27	32	79	.38	.31	28	75	.31	.26
		StdDev	22	7	10	.14	.13	9	8	.18	.17	4	6	.09	.10
		Max	95	38	84	.52	.47	43	94	.70	.62	35	81	.48	.43
		Min	32	21	55	.20	.13	18	73	.21	.17	24	67	.22	.16
	PS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	65	31	73	.36	.26	33	74	.33	.24	37	80	.31	.27
		StdDev	17	13	13	.11	.13	11	14	.09	.05	10	5	.10	.08
		Max	82	51	93	.51	.46	46	95	.44	.31	51	86	.48	.38
		Min	35	18	62	.24	.16	18	57	.22	.18	26	72	.18	.14
	PS2	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	62	32	77	.44	.36	29	75	.36	.25	31	74	.28	.21
		StdDev	21	8	7	.17	.16	9	11	.13	.10	7	7	.14	.11
		Max	87	44	82	.69	.62	39	85	.48	.36	42	81	.40	.34
		Min	34	25	65	.21	.19	20	56	.10	.08	21	66	.08	.07
	TS2	N	5	5	5	5	5	5	5	5	5	5	5	5	5
		Mean	67	36	79	.54	.38	24	64	.50	.30	34	80	.43	.37
		StdDev	20	8	8	.13	.17	13	14	.07	.15	7	6	.11	.12
		Max	93	47	92	.68	.57	45	84	.59	.54	42	87	.57	.51
		Min	38	25	71	.38	.20	14	52	.43	.16	25	71	.28	.25
Content	TS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
		Mean	69	28	69	.37	.27	29	74	.36	.27	30	76	.36	.32
		StdDev	22	12	16	.09	.12	12	16	.16	.16	5	6	.10	.09
		Max	95	41	85	.47	.39	40	87	.59	.51	38	84	.49	.45
		Min	32	12	41	.24	.11	9	46	.18	.12	22	69	.24	.23
	PS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	65	28	72	.31	.22	33	74	.32	.22	39	79	.30	.26
		StdDev	17	12	17	.18	.18	11	16	.15	.13	10	6	.17	.13
		Max	82	49	93	.61	.50	51	96	.46	.42	55	88	.43	.36
		Min	35	17	55	.10	.06	19	59	.07	.07	30	73	-.03	.01
	PS2	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	62	33	81	.53	.40	27	74	.36	.23	37	82	.37	.28
		StdDev	21	14	11	.08	.13	11	18	.22	.15	8	9	.12	.08
		Max	87	44	92	.65	.57	42	95	.53	.42	46	90	.54	.41
		Min	34	8	60	.42	.18	13	42	.00	.02	26	65	.25	.19
	TS2	N	5	5	5	5	5	5	5	5	5	5	5	5	5
		Mean	67	25	69	.58	.36	21	65	.57	.34	36	84	.55	.50
		StdDev	20	14	17	.18	.21	9	16	.12	.17	6	7	.10	.09
		Max	93	47	95	.78	.65	32	89	.67	.60	43	94	.68	.62
		Min	38	10	49	.39	.20	8	44	.39	.18	29	78	.42	.39
Language	TS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	69	26	72	.34	.26	32	74	.30	.24	31	76	.38	.32
		StdDev	22	4	13	.15	.15	11	8	.12	.12	7	4	.11	.12
		Max	95	31	82	.59	.49	47	86	.48	.42	42	81	.53	.49

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
		Min	32	21	49	.21	.13	21	66	.17	.11	24	72	.27	.19
	PS1	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	65	30	73	.36	.25	33	75	.38	.26	39	85	.31	.26
		StdDev	17	18	16	.19	.19	11	16	.24	.17	12	6	.20	.18
		Max	82	57	94	.70	.57	54	100	.66	.55	51	95	.52	.45
		Min	35	12	56	.14	.08	20	56	.02	.04	23	79	-.01	-.06
	PS2	N	6	6	6	6	6	6	6	6	6	6	6	6	6
		Mean	62	34	83	.53	.40	33	81	.41	.28	37	81	.39	.30
		StdDev	21	9	6	.15	.14	8	12	.17	.15	9	6	.09	.07
		Max	87	49	92	.68	.62	41	97	.56	.48	50	88	.49	.35
		Min	34	25	78	.28	.22	22	64	.14	.09	24	74	.23	.18
	TS2	N	5	5	5	5	5	5	5	5	5	5	5	5	5
		Mean	67	30	73	.58	.38	23	60	.50	.28	31	85	.49	.43
		StdDev	20	13	13	.18	.21	11	16	.15	.15	7	7	.13	.11
		Max	93	50	95	.84	.71	42	84	.65	.53	39	93	.69	.61
		Min	38	18	61	.41	.20	13	43	.32	.15	21	76	.37	.34
Conventions	TS1	N	6	6	6		6	6	6		6	6	6		6
		Mean	69	54	98		.21	54	97		.18	57	96		.30
		StdDev	22	9	2		.16	6	3		.08	10	1		.12
		Max	95	62	100		.45	59	100		.33	65	97		.47
		Min	32	41	94		.00	45	93		.13	38	94		.11
	PS1	N	6	6	6		6	6	6		6	6	6		6
		Mean	65	54	98		.16	58	98		.23	63	97		.18

Score	Sample[a]	Stat	N	Teacher vs. ESC[b]				Teacher vs. TR1[b]				ESC vs. TR1[b]			
				EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK	EA (%)	EAA (%)	COR	WK
		StdDev	17	16	2		.09	11	2		.17	3	2		.14
		Max	82	69	100		.29	76	100		.45	68	98		.36
		Min	35	25	94		.06	42	96		-.01	59	94		.05
	PS2	N	6	6	6		6	6	6		6	6	6		6
		Mean	62	62	98		.29	63	99		.24	60	99		.26
		StdDev	21	7	3		.15	11	1		.11	5	1		.15
		Max	87	73	100		.50	77	100		.44	69	100		.41
		Min	34	56	92		.10	47	97		.13	57	98		.03
	TS2	N	5	5	5		5	5	5		5	5	5		5
		Mean	67	56	98		.30	51	96		.30	49	98		.30
		StdDev	20	8	4		.16	15	3		.17	6	2		.08
		Max	93	65	100		.52	68	100		.51	55	100		.41
		Min	38	43	90		.11	38	92		.09	41	96		.21

[a] TS1=Timed Sample 1, PS1=Process Sample 1, PS2=Process Sample 2, TS2=Timed Sample 2.

[b] EA=Percentage of exact agreement, EAA=Percentage of exact or adjacent agreement, COR=Correlation, WK=Weighted Kappa with quadratic weights.

APPENDIX K: SUMMARY OF CORRELATIONS BETWEEN WRITING PILOT AND STAAR WRITING SCORES

The polyserial correlations were calculated between the rating scores and the corresponding spring 2018 STAAR scale scores for each test, score, and rater group. A sample size of at least 30 was required for each calculation. This correlation can serve as an external validity indicator for a rater score.

Students who participated in the Texas Writing Pilot and also completed STAAR writing received a comparison of skill assessment where skills assessed were in alignment. The correlations below are not a performance indicator, but instead identify measurement of skills. It is important to note that the STAAR writing assessment only evaluates one mode of writing while the Texas Writing Pilot evaluates multiple modes of writing.

Polyserial correlation (Drasgow, 1988) is appropriate for the case where one variable is an ordered categorical variable and the other is a continuous variable. Like polychoric correlation, polyserial correlation assumes a continuous variable underlying the categorical variable and the two continuous variables follow a binormal distribution. Polyserial correlation is estimated by the maximum likelihood estimation. If the assumptions hold, polyserial correlation more accurately reflects the association between one ordered categorical variable and one continuous variable, while Pearson correlation tends to underestimate the association. For the sums of the seven portfolio scores, their correlations are Pearson correlations because both variables are considered to be continuous.

The correlations on the total writing samples for the four tests in Figures K1–K4 were plotted to better understand the variations across raters and scores. For grade 4 writing, ESC rater scores had the highest correlations from 0.48 to 0.54 with the STAAR scale scores, and Teacher had the lowest correlations from 0.25 to 0.38 except for Language score. For grade 7 writing, Teacher had the highest correlations from 0.60 to 0.69, and ESC rater and Trained Rater 1 had similar correlations from 0.48 to 0.56. For English I, Teacher had the highest correlations from 0.50 to 0.59, and ESC rater and Trained Rater 1 had similar correlations from 0.45 to 0.51. For English II, all raters had the similar correlations from 0.41 to 0.50 except for Conventions score where Teacher had the correlation of 0.61, while ESC rater and Trained Rater 1 had the same correlation of 0.45. Overall, in grade 4 writing the rating scores had low to medium correlations with the STAAR scale scores, and in the other tests they had medium correlations, which provide some evidence to support the validity of these rating scores.

Figure K1. Correlations between Rating Scores and STAAR Scale Scores on Total Writing Samples: Grade 4 Writing.

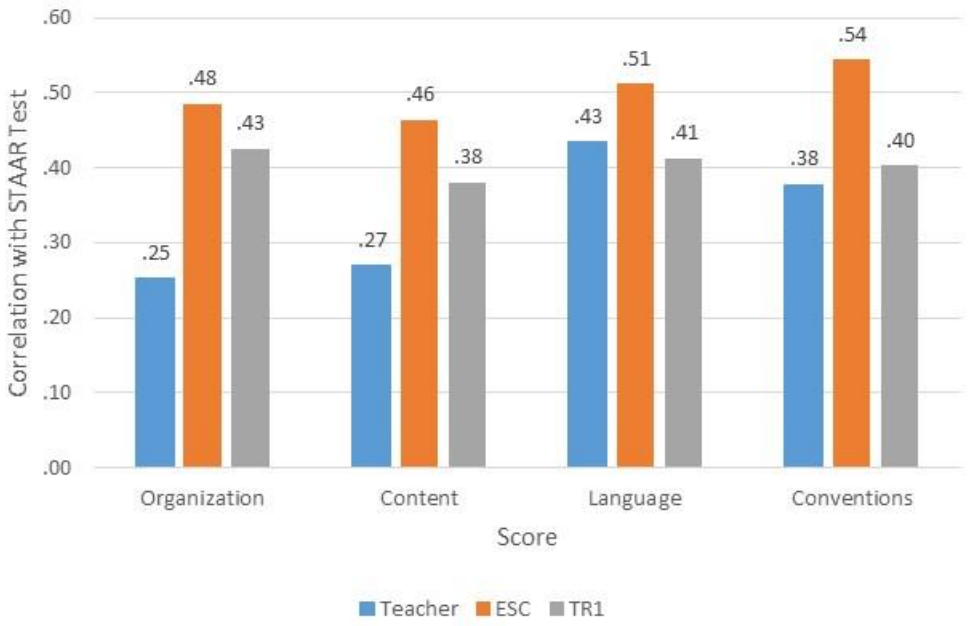


Figure K2. Correlations between Rating Scores and STAAR Scale Scores on Total Writing Samples: Grade 7 Writing.

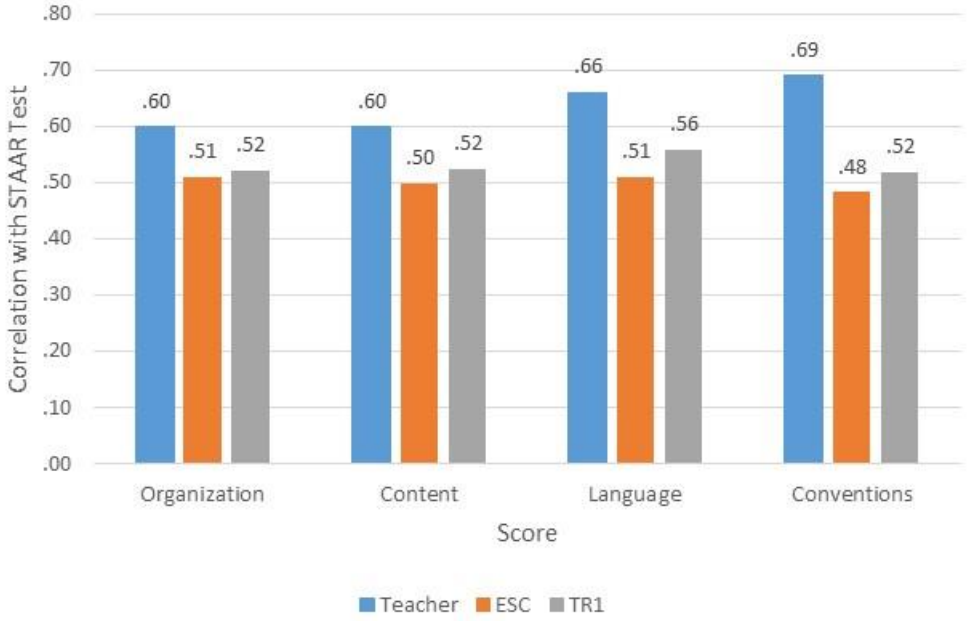


Figure K3. Correlations between Rating Scores and STAAR Scale Scores on Total Writing Samples: English I.

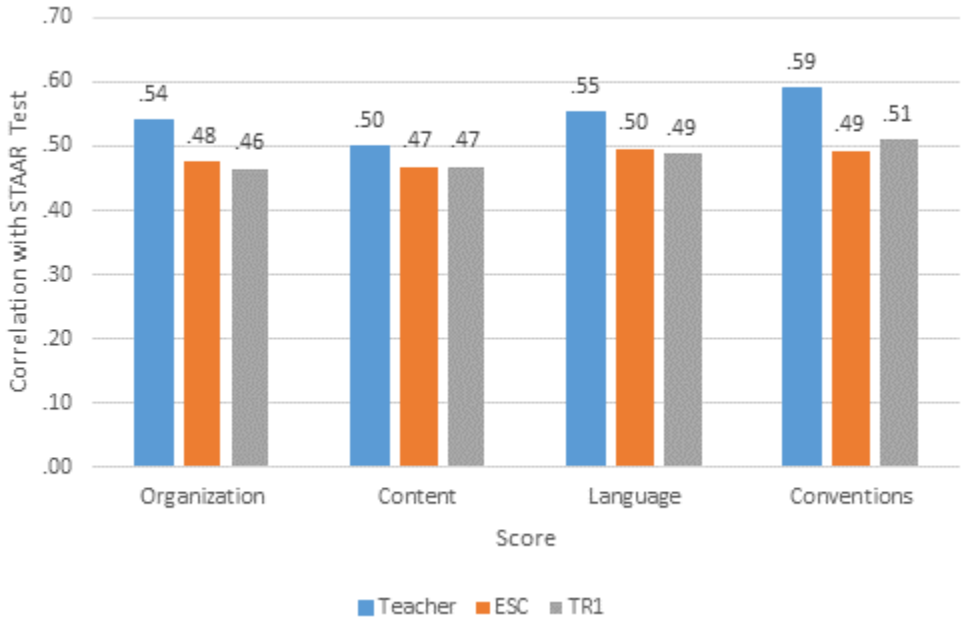
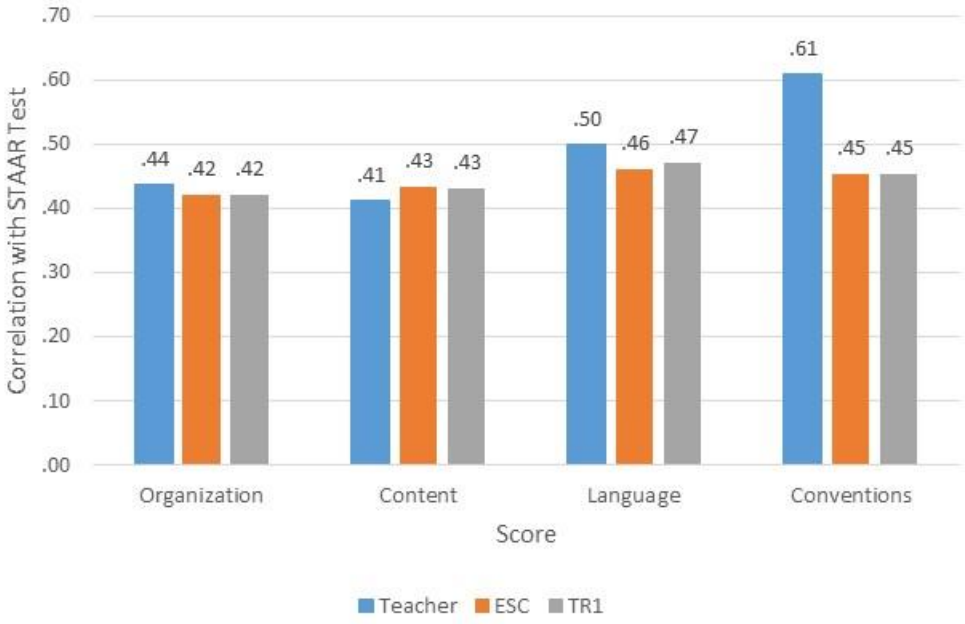


Figure K4. Correlations between Rating Scores and STAAR Scale Scores on Total Writing Samples: English II.



APPENDIX L: CORRELATIONS BETWEEN WRITING PILOT AND STAAR WRITING SCORES

Tables L1–L4 list the correlations for all rating scores and the three raters (Teacher, ESC rater, and Trained Rater 1) on each writing sample (TS1, PS1, PS2, and TS2), each writing genre, each timed writing prompt, and the total writing samples in the four tests, respectively. Within a test, the correlations varied across raters, scores, and sample groups.

Table L1. Correlations between Rating Scores and STAAR Scale Scores: Grade 4 Writing

Score	Sample[a]	N	Teacher	ESC	TR1
Organization	TS1	132	.23	.53	.49
	PS1	129	.27	.50	.48
	PS2	131	.25	.37	.26
	TS2	132	.41	.57	.47
	Expository	246	.25	.43	.40
	Personal Narrative_TS	264	.26	.54	.47
	1000022	146	.23	.54	.42
	1000023	42	.32	.57	.50
	1000024	76	.25	.49	.55
	Total	524	.25	.48	.43
Content	TS1	132	.26	.49	.50
	PS1	129	.30	.49	.40
	PS2	131	.24	.40	.26
	TS2	132	.42	.51	.40
	Expository	246	.25	.45	.34
	Personal Narrative_TS	264	.29	.49	.44
	1000022	146	.25	.49	.42
	1000023	42	.31	.52	.44
	1000024	76	.33	.46	.44
	Total	524	.27	.46	.38
Language	TS1	132	.40	.55	.51
	PS1	129	.51	.51	.38
	PS2	131	.44	.46	.29
	TS2	132	.58	.58	.47
	Expository	246	.44	.49	.35
	Personal Narrative_TS	264	.43	.55	.48
	1000022	146	.44	.54	.47
	1000023	42	.58	.59	.55

Score	Sample[a]	N	Teacher	ESC	TR1
	1000024	76	.35	.54	.45
	Total	524	.43	.51	.41
Conventions	TS1	132	.56	.63	.57
	PS1	129	.36	.54	.40
	PS2	131	.24	.45	.20
	TS2	132	.42	.58	.46
	Expository	246	.31	.50	.32
	Personal Narrative_TS	264	.45	.59	.50
	1000022	146	.43	.52	.52
	1000023	42	.53	.60	.43
	1000024	76	.44	.68	.46
	Total	524	.38	.54	.40

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; The numbers are the prompt IDs of timed samples.

Table L2. Correlations between Rating Scores and STAAR Scale Scores: Grade 7 Writing

Score	Sample[a]	N	Teacher	ESC	TR1
Organization	TS1	142	.68	.63	.55
	PS1	93	.50	.39	.52
	PS2	47	.65	.51	.52
	Expository	52	.68	.43	.60
	Personal Narrative	61	.50	.54	.47
	Expository_TS	164	.67	.60	.53
	1000029	108	.70	.66	.51
	1000031	42	.55	.44	.61
	Total	304	.60	.51	.52
Content	TS1	142	.67	.59	.54
	PS1	93	.50	.40	.56
	PS2	47	.65	.52	.55
	Expository	52	.68	.36	.64
	Personal Narrative	61	.47	.61	.54
	Expository_TS	164	.66	.57	.51
	1000029	108	.70	.60	.48
	1000031	42	.55	.51	.57

Score	Sample[a]	N	Teacher	ESC	TR1
	Total	304	.60	.50	.52
Language	TS1	142	.74	.62	.57
	PS1	93	.60	.36	.58
	PS2	47	.74	.49	.61
	Expository	52	.66	.38	.66
	Personal Narrative	61	.61	.57	.56
	Expository_TS	164	.70	.61	.54
	1000029	108	.72	.64	.51
	1000031	42	.63	.60	.61
	Total	304	.66	.51	.56
Conventions	TS1	142	.73	.65	.53
	PS1	93	.58	.26	.48
	PS2	47	.76	.50	.59
	Expository	52	.67	.15	.50
	Personal Narrative	61	.70	.58	.57
	Expository_TS	164	.74	.61	.52
	1000029	108	.72	.60	.54
	1000031	42	.76	.65	.55
	Total	304	.69	.48	.52

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; The numbers are the prompt IDs of timed samples.

Table L3. Correlations between Rating Scores and STAAR Scale Scores: English I

Score	Sample[a]	N	Teacher	ESC	TR1
Organization	TS1	666	.50	.47	.47
	PS1	282	.60	.49	.50
	PS2	584	.56	.47	.49
	TS2	666	.61	.51	.42
	Analytic	109	.31	.37	.43
	Expository	301	.60	.52	.54
	Other	58	.60	.61	.47
	Personal Narrative	314	.59	.40	.43
	Persuasive	84	.39	.51	.42
	Expository_TS	1332	.54	.49	.45

Score	Sample[a]	N	Teacher	ESC	TR1
	1000032	388	.43	.44	.43
	1000033	408	.61	.49	.42
	21000001	536	.56	.52	.48
	Total	2198	.54	.48	.46
Content	TS1	666	.51	.47	.48
	PS1	282	.59	.48	.52
	PS2	584	.50	.49	.49
	TS2	666	.51	.45	.41
	Analytic	109	.30	.37	.46
	Expository	301	.60	.53	.56
	Other	58	.60	.70	.41
	Personal Narrative	314	.53	.41	.41
	Persuasive	84	.38	.52	.46
	Expository_TS	1332	.50	.46	.45
	1000032	388	.49	.43	.45
	1000033	408	.57	.46	.42
	21000001	536	.48	.49	.46
	Total	2198	.50	.47	.47
Language	TS1	666	.51	.51	.50
	PS1	282	.61	.49	.50
	PS2	584	.56	.48	.51
	TS2	666	.61	.50	.45
	Analytic	109	.39	.34	.45
	Expository	301	.60	.54	.54
	Other	58	.63	.52	.41
	Personal Narrative	314	.56	.43	.45
	Persuasive	84	.48	.45	.44
	Expository_TS	1332	.55	.51	.48
	1000032	388	.49	.47	.47
	1000033	408	.57	.50	.45
	21000001	536	.58	.53	.50
	Total	2198	.55	.50	.49
Conventions	TS1	666	.60	.49	.51
	PS1	282	.55	.54	.53

Score	Sample[a]	N	Teacher	ESC	TR1
	PS2	584	.59	.44	.51
	TS2	666	.62	.51	.48
	Analytic	109	.41	.33	.48
	Expository	301	.54	.62	.56
	Other	58	.69	.33	.38
	Personal Narrative	314	.60	.40	.41
	Persuasive	84	.41	.39	.44
	Expository_TS	1332	.59	.49	.49
	1000032	388	.52	.40	.46
	1000033	408	.62	.47	.51
	21000001	536	.62	.56	.49
	Total	2198	.59	.49	.51

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; The numbers are the prompt IDs of timed samples.

Table L4. Correlations between Rating Scores and STAAR Scale Scores: English II

Score	Sample[a]	N	Teacher	ESC	TR1
Organization	TS1	463	.43	.48	.48
	PS1	426	.38	.35	.35
	PS2	423	.45	.34	.47
	TS2	383	.53	.52	.43
	Analytic	160	.54	.36	.50
	Expository	182	.51	.42	.50
	Other	156	.22	.15	.22
	Personal Narrative	98	.42	.40	.25
	Persuasive	253	.46	.31	.30
	Persuasive_TS	846	.47	.49	.45
	1000035	281	.50	.48	.50
	1000036	302	.44	.44	.41
	1000037	263	.46	.55	.43
	Total	1695	.44	.42	.42
Content	TS1	463	.41	.47	.49
	PS1	426	.42	.37	.35
	PS2	423	.43	.44	.50

Score	Sample[a]	N	Teacher	ESC	TR1
	TS2	383	.42	.49	.44
	Analytic	160	.68	.50	.46
	Expository	182	.57	.46	.48
	Other	156	.24	.18	.21
	Personal Narrative	98	.46	.42	.35
	Persuasive	253	.41	.34	.33
	Persuasive_TS	846	.42	.47	.46
	1000035	281	.43	.43	.49
	1000036	302	.39	.43	.41
	1000037	263	.41	.55	.48
	Total	1695	.41	.43	.43
Language	TS1	463	.50	.51	.50
	PS1	426	.49	.37	.41
	PS2	423	.56	.43	.54
	TS2	383	.50	.55	.51
	Analytic	160	.71	.54	.52
	Expository	182	.64	.48	.56
	Other	156	.32	.17	.28
	Personal Narrative	98	.55	.43	.36
	Persuasive	253	.46	.31	.39
	Persuasive_TS	846	.50	.52	.49
	1000035	281	.49	.49	.56
	1000036	302	.49	.49	.42
	1000037	263	.50	.59	.49
	Total	1695	.50	.46	.47
Conventions	TS1	463	.57	.56	.52
	PS1	426	.54	.28	.39
	PS2	423	.67	.42	.48
	TS2	383	.66	.55	.51
	Analytic	160	.75	.40	.40
	Expository	182	.72	.36	.54
	Other	156	.21	.20	.35
	Personal Narrative	98	.52	.51	.40
	Persuasive	253	.51	.29	.32

Score	Sample[a]	N	Teacher	ESC	TR1
	Persuasive_TS	846	.61	.55	.51
	1000035	281	.61	.54	.55
	1000036	302	.54	.50	.45
	1000037	263	.64	.61	.52
	Total	1695	.61	.45	.45

[a] TS1=Timed Sample 1; PS1=Process Sample 1; PS2=Process Sample 2; TS2=Timed Sample 2; A genre of timed sample ends with a suffix "_TS", while a genre of process sample does not have such a suffix; The numbers are the prompt IDs of timed samples.