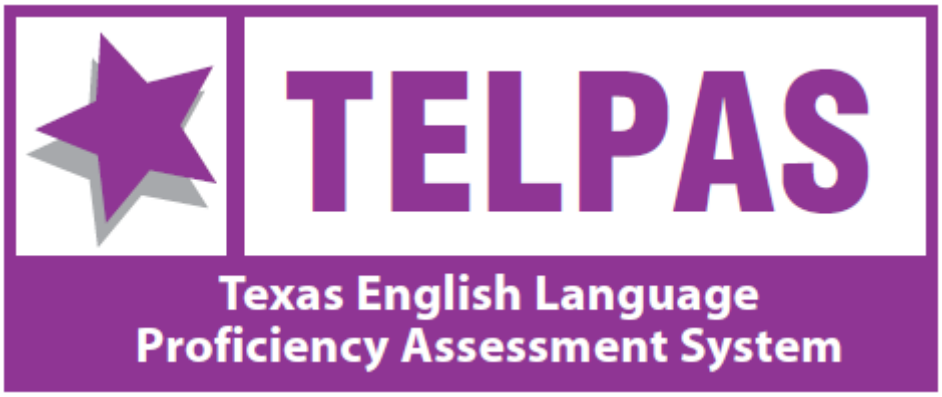


Texas English Language Proficiency Assessment System (TELPAS)

Standards Review Technical Report

May 22, 2014

Table of Contents

Chapter 1: Texas English Language Proficiency Assessment System (TELPAS) Standards Review

This chapter provides an overview of the TELPAS program and the TELPAS standards review process. It includes the following sections:

- Purpose of the TELPAS Program
- TELPAS Curriculum Standards
- TELPAS Proficiency Level Standards
- Purpose of TELPAS Reading Standards Review
- TELPAS Standards Review Approach
- TELPAS Standards Review Process

Purpose of the TELPAS Program

English language proficiency assessments are federally required to evaluate the progress that English language learners (ELLs) make in becoming proficient in the use of academic English. Title III, Part A of the Elementary and Secondary Education Act (ESEA) requires states to conduct annual statewide English language proficiency assessments for ELLs in grades K–12. TELPAS assesses the English language proficiency of K–12 ELLs in four language domains: listening, speaking, reading, and writing. The TELPAS assessments are performance-based and holistically rated, with the exception of the reading assessments for grades 2–12, which are multiple-choice tests.

- Grades K–1: TELPAS includes holistically rated listening, speaking, reading, and writing assessments based on ongoing classroom observations and student interactions.
- Grades 2–12: TELPAS includes multiple-choice reading tests, holistically rated student writing collections, and holistically rated listening and speaking assessments. The listening and speaking assessments are based on ongoing classroom observations and student interactions.

TELPAS is used, in conjunction with the State of Texas Assessments of Academic Readiness (STAAR) to show the extent to which districts and the state meet federal Annual Measureable Achievement Objective (AMAO) accountability indicators that are specific to English language proficiency and academic achievement of ELLs. Composite performance (which combines performance on listening, speaking, reading, and writing), rather than individual language domain performance, is used in TELPAS AMAO indicators.

TELPAS results are also used at the student level to help teachers design instruction and plan interventions that appropriately address the student's linguistic and academic needs.

TELPAS Curriculum Standards

TELPAS measures acquisition of the English language in alignment with the Texas English Language Proficiency Standards (ELPS) that are part of the STAAR state-mandated curriculum, the Texas Essential Knowledge and Skills (TEKS). The ELPS are second-language-acquisition curriculum standards that support the ability of ELLs to acquire academic English, while at the same time allowing them to engage meaningfully in regular, all-English academic instruction at their grade level. Approved by the State Board of Education in 2007–2008, the ELPS are set forth in Title 18, Chapter 74.4 of the Texas Administrative Code (TAC).

Districts are required to implement the ELPS as an integral part of each foundation and enrichment subject contained in the state-required curriculum standards. The ELPS outline the instruction that ELLs must receive to support their ability to develop academic English language proficiency and acquire challenging academic knowledge and skills. In addition, the ELPS include proficiency level descriptors (PLDs), characterizing the four English language proficiency levels reported in Texas. TELPAS is designed to directly support the state's educational goals for meeting the language and content needs of ELLs.

TELPAS Proficiency Level Standards

The TELPAS proficiency level standards are the three cut scores on each TELPAS test that divide students into the four English language proficiency levels used in Texas.

PROFICIENCY LEVEL LABELS AND GLOBAL DEFINITIONS

For each language domain, TELPAS measures four levels, or stages, of increasing English language proficiency:

- beginning
- intermediate
- advanced
- advanced high

Global definitions provide a common definition of the characteristics specific to each proficiency level across language domains. Global definitions explain what it means for a student to be classified as beginning, intermediate, advanced, or advanced high across language domains. These definitions apply to all TELPAS assessments. Understanding the global definitions and features provides the foundation for understanding the types of language acquisition skills students possess at each proficiency level. The global definitions and key features of each proficiency level are displayed in Table 1.1.

Table 1.1: Global Definitions of the TELPAS Proficiency Levels

Proficiency Level	Global Definitions	Key Features
Beginning	Beginning students have little or no ability to understand and use English. They may know a little English but not enough to function meaningfully in social or academic settings.	Little or no English ability
Intermediate	Intermediate students do have some ability to understand and use English. They can function in social and academic settings as long as the tasks require them to understand and use simple language structures and high-frequency vocabulary in routine contexts.	Limited ability, simple language structures, high-frequency vocabulary, routine contexts
Advanced	Advanced students are able to engage in grade-appropriate academic instruction in English, although ongoing second language acquisition support is needed to help them understand and use grade-appropriate language. These students function beyond the level of simple, routinely used English.	Ability to engage in grade-appropriate academic instruction with second language acquisition support
Advanced High	Advanced high students have attained the command of English that enables them, with minimal second language acquisition support, to engage in regular, all-English academic instruction at their grade level.	Ability to engage in grade-appropriate academic instruction with minimal second language acquisition support

ELPS PLDs

While the global definitions apply across language domains, the ELPS PLDs present the major characteristics of each language proficiency level in each language domain. The PLDs are domain specific and define how well ELLs at the four proficiency levels are able to understand and use English in grade-level academic settings. The descriptors show the progression of second language acquisition from one proficiency level to the next and serve as a road map to help teachers tailor instruction to the linguistic needs of ELLs. PLDs are also a critical part of the process used to set and review the TELPAS standards. They provide a common framework for understanding the language acquisition skills needed to be classified within each proficiency level.

The TELPAS reading PLDs for grades 2–12 are shown in Table 1.2. PLDs for the other TELPAS domains can be found at:
http://www.tea.state.tx.us/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=2147489015&libID=2147489014

Table 1.2: ELPS-TELPAS PLDs Grades 2–12 Reading

Beginning	Intermediate	Advanced	Advanced High
Beginning English language learners (ELLs) have little or no ability to read and understand English used in academic and social contexts.	Intermediate ELLs have the ability to read and understand simple, high-frequency English used in routine academic and social contexts.	Advanced ELLs have the ability to read and understand, with second language acquisition support, grade-appropriate English used in academic and social contexts.	Advanced high ELLs have the ability to read and understand, with minimal second language acquisition support, grade appropriate English used in academic and social contexts.
These students: • read and understand the very limited recently practiced, memorized, or highly familiar English they have learned; vocabulary predominantly includes – environmental print – some very high-frequency words – concrete words that can be represented by pictures • read slowly, word by word • have a very limited sense of English language structures • comprehend predominantly isolated familiar words and phrases; comprehend some sentences in highly routine contexts or recently practiced, highly familiar text • are highly dependent on visuals and prior knowledge to derive meaning from text in English • are able to apply reading comprehension skills in English only when reading texts written for this level	**These students:** • read and understand English vocabulary on a somewhat wider range of topics and with increased depth; vocabulary predominantly includes – everyday oral language – literal meanings of common words – routine academic language and terms – commonly used abstract language such as terms used to describe basic feelings • often read slowly and in short phrases; may re-read to clarify meaning • have a growing understanding of basic, routinely used English language structures • understand simple sentences in short, connected texts, but are dependent on visual cues, topic familiarity, prior knowledge, pretaught topic-related vocabulary, story predictability, and teacher/peer assistance to sustain comprehension • struggle to independently read and understand grade-level texts • are able to apply basic and some higher-order comprehension skills when reading texts that are linguistically accommodated and/or simplified for this level	**These students:** • read and understand, with second language acquisition support, a variety of grade-appropriate English vocabulary used in social and academic contexts: – with second language acquisition support, read and understand grade-appropriate concrete and abstract vocabulary, but have difficulty with less commonly encountered words – demonstrate an emerging ability to understand words and phrases beyond their literal meaning – understand multiple meanings of commonly used words • read longer phrases and simple sentences from familiar text with appropriate rate and speed • are developing skill in using their growing familiarity with English language structures to construct meaning of grade-appropriate text • are able to apply basic and higher-order comprehension skills when reading grade-appropriate text, but are still occasionally dependent on visuals, teacher/peer assistance, and other linguistically accommodated text features to determine or clarify meaning, particularly with unfamiliar topics	**These students:** • read and understand vocabulary at a level nearly comparable to that of their native English-speaking peers, with some exceptions when low-frequency or specialized vocabulary is used • generally read grade-appropriate, familiar text with appropriate rate, speed, intonation, and expression • are able to, at a level nearly comparable to native English-speaking peers, used their familiarity with English language structures to construct meaning of grade-appropriate text • are able to apply, with minimal second language acquisition support and at a level nearly comparable to native English-speaking peers, basic and higher-order comprehension skills when reading grade-appropriate text

Purpose of TELPAS Reading Standards Review

Once a set of standards has been adopted for an assessment, the standards may apply as long as they are judged to be appropriate for expressing the students' proficiency levels on the assessment. An assessment program may consider reviewing the standards either when changes occur in the assessment program or the assessed curriculum, or as a periodic check to evaluate the continued appropriateness of the standards.

The original TELPAS reading proficiency level standards were established in 2008 when the Texas Assessment of Knowledge and Skills (TAKS) was the academic assessment in Texas. The move from TAKS to STAAR in 2011–2012 made it necessary to evaluate whether the TELPAS reading standards needed to be reviewed. The STAAR assessment program differs from TAKS in a number of ways. STAAR has a stronger emphasis on academic rigor in terms of the cognitive demands and the level of skill needed to pass each assessment. Additionally, the STAAR program was designed to be a comprehensive system, with high school curriculum and performance standards aligning and linking back to elementary and middle school and projecting forward to postsecondary readiness. Given these changes, the PLDs for TELPAS were reviewed and deemed to be as relevant and applicable to English language acquisition as they were in 2008. Since the PLDs are the rubrics for the holistically rated assessments, proficiency level standards for listening, speaking, writing, and K–1 reading were not included as part of the standards review.

However, a standards review was conducted to evaluate the multiple-choice components of TELPAS, grades 2–12 reading. While the reading PLDs were still relevant, it was necessary to review the original TELPAS reading proficiency level standards so that performance on TELPAS reading could be considered to be a meaningful indicator of the level of English language proficiency required to be successful on STAAR reading. A standards review committee was convened to make recommendations about how to adjust the TELPAS reading proficiency level standards to align with STAAR so that TELPAS reading performance would reflect students' ability to engage in all-English academic instruction at their grade level. The focus of this report is on the process used to review the TELPAS reading proficiency level standards in grades 2–12.

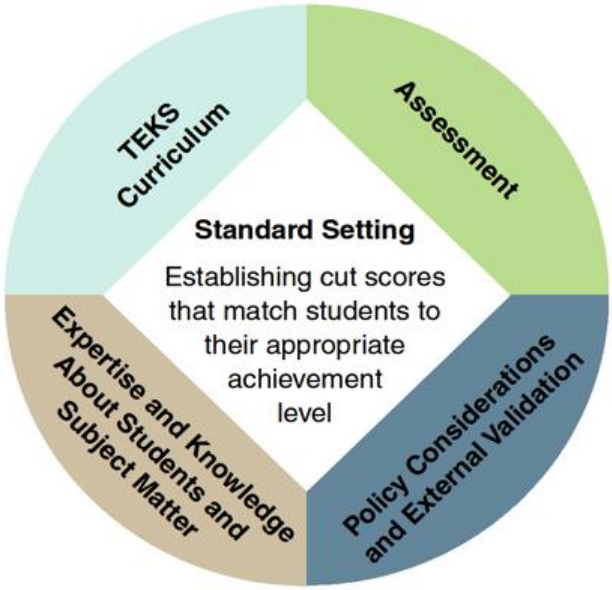
TELPAS Standards Review Approach

The Texas Education Agency (TEA) used an evidence-based standard-setting approach (O'Malley, Keng, & Miles, 2012) to review the cut scores for the four proficiency levels (beginning, intermediate, advanced, and advanced high) on the TELPAS reading assessments in grades 2–12. This is similar to the approach that was used to establish the performance standards on the STAAR assessments.

The standards review approach for TELPAS reading involved a process of combining considerations regarding policy, the ELPS global definitions and PLDs, the TEKS content standards, educator knowledge about what ELLs should know and be able to do, and information about how ELL performance on TELPAS reading aligns with performance on the

STAAR reading assessments. Standards review advisory committees, made up of groups of classroom teachers, bilingual specialists, and English language acquisition experts, considered the interaction of all these elements for each TELPAS reading assessment. Figure 1.1 illustrates the critical elements of the evidence-based standard-setting approach that was used to review the TELPAS reading proficiency level standards.

Figure 1.1: Critical Elements of the Evidence-Based Standard-Setting Approach



Each element of the evidence-based standard-setting approach as it relates to the TELPAS reading assessments is described below.

- **TEKS Curriculum:** The ELPS, which are part of the TEKS curriculum standards, support the ability of ELLs to acquire academic English while allowing them to engage meaningfully in regular, all-English, on-grade level academic instruction. They provide the underlying basis for several key components of the TELPAS reading standards review process, including the proficiency level labels, global definitions, and PLDs.
- **Assessment:** Each TELPAS assessment is developed to measure the knowledge and skills described in the ELPS. Each TELPAS reading assessment is based on the student expectations and reporting categories specified in the TELPAS assessed curriculum document and the TELPAS reading test blueprint.
- **Policy Considerations and External Validation:** Empirical studies that correlated performance on the TELPAS reading assessments with scores on the STAAR reading assessments were conducted. Comparisons of the text complexity between selected STAAR passages and advanced high TELPAS reading passages were also performed. Results of the empirical studies and text complexity evaluations were used to inform the standards review process.

- **Expertise and Knowledge about Students and Subject Matter:** Texas educators, including classroom teachers, bilingual specialists, and English language acquisition experts, brought content knowledge and experience with ELLs to the standards review process. They played an integral role in reviewing and recommending new proficiency level standards.
- **Standard Setting:** Within the framework of evidence-based standard-setting, an established standard-setting method, known as the item-mapping with external data method (Ferrara, Lewis, Mercado, D'Brot, Barth, & Egan, 2011; Phillips, 2012), was used to recommend the TELPAS reading proficiency level standards.

TELPAS Standards Review Process

The TELPAS standards review process included the following steps:

1. Conduct empirical studies
2. Develop proficiency level labels and global definitions
3. Develop proficiency level descriptors
4. Develop neighborhoods
5. Convene standards review committees
6. Review proficiency level standards for reasonableness
7. Approve proficiency level standards
8. Implement proficiency level standards
9. Review proficiency level standards

A description of each step in the TELPAS standards review process is provided below.

STEP 1: CONDUCT EMPIRICAL STUDIES

TEA conducted extensive research to support the TELPAS standards review process. The empirical research studies

- evaluated the relationship between TELPAS reading and STAAR reading assessments;
- evaluated the relationship between TELPAS reading and TELPAS writing;
- evaluated the relationship between TELPAS reading and TAKS reading;
- compared the performance of ELLs and non-ELLs on STAAR reading; and
- compared the text complexity between advanced high TELPAS reading and STAAR reading passages.

To support reliable and meaningful score interpretations, evaluation of the empirical relationships between the two assessments was based on the same students taking the TELPAS reading assessment and the STAAR reading assessments of interest. Chapter 2 provides more details about each of the empirical studies.

STEP 2: DEVELOP PROFICIENCY LEVEL LABELS AND GLOBAL DEFINITIONS

The proficiency level labels and global definitions for TELPAS were developed before the initial TELPAS standards were set in 2008 and provide a foundation for identifying the characteristics of ELLs in each proficiency level. They apply across all TELPAS language domains (reading, writing, listening, and speaking) and grade levels (K–12). Additional information about the proficiency level labels and global definitions can be found in the *TELPAS Proficiency Level Standards* section of this chapter.

STEP 3: DEVELOP PROFICIENCY LEVEL DESCRIPTORS

The PLDs translate the global definitions into domain-specific descriptions of student performance at each proficiency level. The PLDs were also developed before the initial TELPAS standards were set in 2008. They define how well ELLs at the four proficiency levels are able to understand and use English in grade-level academic settings, show the progression of second language acquisition from one proficiency level to the next, serve as a road map to help teachers tailor instruction to the linguistic needs of ELLs, and are a critical part of reviewing the proficiency level standards. More details about the TELPAS reading PLDs can be found in the *TELPAS Proficiency Level Standards* section of this chapter.

STEP 4: DEVELOP NEIGHBORHOODS

A neighborhood is a range within which it would be reasonable to set each TELPAS reading proficiency level standard. Using the results from the empirical studies conducted, TEA constructed neighborhoods for each proficiency level standard on all TELPAS reading assessments in grades 2–12. The neighborhoods were also evaluated across TELPAS reading assessments to confirm that they reflected an appropriate increase in English language acquisition from grade to grade. Additional information about the TELPAS reading neighborhoods is provided in Chapter 2.

STEP 5: CONVENE STANDARDS REVIEW COMMITTEES

Committees comprised of Texas educators, including classroom teachers, bilingual specialists, and English language acquisition experts, used the TELPAS reading proficiency level labels and global definitions, the TELPAS reading PLDs, the TELPAS reading test questions, the TELPAS and STAAR reading passages, and the neighborhoods for each proficiency level standard to review and recommend cut scores for the TELPAS reading assessments. More details about the TELPAS standards review committees, including the committee composition, meeting proceedings and outcomes, are provided in Chapter 3.

STEP 6: REVIEW PROFICIENCY LEVEL STANDARDS FOR REASONABLENESS

TEA reviewed the cut scores recommended by the standards review committees across grade levels to evaluate the reasonableness of the standards as a system and made adjustments as appropriate. Refer to Chapter 3 for more information about the reasonableness review process for TELPAS reading.

STEP 7: APPROVE PROFICIENCY LEVEL STANDARDS

In August 2013, the commissioner of education approved the TELPAS reading proficiency level standards based on the recommendations from the standards review committees and TEA.

STEP 8: IMPLEMENT PROFICIENCY LEVEL STANDARDS

The approved proficiency level standards are posted on the TEA website and will be applied to grades 2–12 TELPAS reading assessments beginning with the spring 2014 administration.

STEP 9: REVIEW PROFICIENCY LEVEL STANDARDS

TEA will continue to monitor the performance of ELLs on TELPAS in relationship to their performance on the STAAR assessments, especially in light of the fact that the STAAR performance standards will increase incrementally over time (a phase-in of standards) and new STAAR English assessments will be implemented, beginning in spring 2014. If necessary, the TELPAS proficiency level standards may be reviewed again and adjusted as appropriate.

Chapter 2: Validity Studies and Neighborhood Development

This chapter provides details about conducting validity studies and developing proficiency level standard ranges ("neighborhoods") to inform the standards review process. The sections in this chapter include:

- Use of Validity Studies in Standard Setting
- Types of Validity Studies
- Using and Displaying Validity Study Results
- TELPAS Neighborhood Development

Use of Validity Studies in Standard Setting

TELPAS is designed to measure the level of English language proficiency obtained by ELLs in Texas, and to track their progress toward English proficiency. Though TELPAS is not used as an exit criterion from ELL programs, there has been an expectation that ELLs who score advanced high on TELPAS will have attained a sufficient level of English proficiency to be successful in their content-area assessments with an additional year of instruction. When TAKS was the academic assessment, this expectation was usually accurate. However, with the new STAAR assessments and higher content standard expectations overall, many advanced high students are no longer successful with an additional year of instruction. Therefore, a review of the TELPAS reading proficiency level standards in grades 2–12 was conducted so that performance on TELPAS reading could be a meaningful indicator of the level of English language proficiency required to be successful on STAAR assessments. Validity studies evaluating the relationship between TELPAS and STAAR performance were used to provide guidance in reviewing the TELPAS reading standards.

TEA designed and implemented a systematic approach to incorporate validity study information into the TELPAS standards review process. This approach was derived from the evidence-based standard-setting method (Beimers, Way, McClarty, & Miles, 2012; O'Malley, Keng, & Miles, 2011). The approach involved making use of the combined expertise of content specialists and measurement experts to select appropriate validity studies, develop data collection plans, and execute appropriate analysis methods. By blending components of several traditional standard-setting methods, an evidence-based standard-setting method was developed that was uniquely suited to fulfill the requirements of reviewing the proficiency level standards for the TELPAS reading assessments.

Types of Validity Studies

Seven validity studies were conducted to inform the TELPAS reading standards review. These studies were designed to compare the text complexity of TELPAS advanced high passages to STAAR passages, to compare TELPAS and STAAR performance, and to compare the performance of ELLs and non-ELLs on STAAR. Brief descriptions of each study are provided below. Refer to Appendix 1 for more detailed information about analysis methods, implementation, and results.

Text Complexity Analyses

STAAR reading passages differ from TAKS passages in the level of text complexity, the level of skill needed to read and understand the passages, and the genres included. STAAR reading passages are generally more difficult than TAKS passages at the same grade level. Although TELPAS advanced high passages provided an adequate preparation for reading TAKS passages, it was unclear how well they prepared ELLs for STAAR passages. TEA and Pearson created a rubric to compare the text complexity of TELPAS advanced high and STAAR passages at the same grade level.

STAAR Pass Rate Comparison

Although a large percentage of students classified as attaining advanced high on TELPAS passed TAKS the following year, this was not true for STAAR. Pass rates at each of the TELPAS proficiency levels were compared to pass rates of former ELLs in their first two years of monitoring and non-ELLs matched on socioeconomic status. Additionally, STAAR pass rates at the Level II standard (phase-in and recommended) were calculated for a variety of possible TELPAS advanced high cut scores.

Average STAAR Score Comparison

To get a sense of how different ELL and non-ELL performance is on STAAR, average STAAR scale scores were compared for each of the TELPAS proficiency levels, for former ELLs, and for non-ELLs matched on socioeconomic status. Additionally, average STAAR scale scores were calculated for a variety of possible TELPAS advanced high cut scores.

TELPAS and STAAR Concordance Tables

A concordance table was developed for each grade, indicating the predicted STAAR score for each TELPAS score. ELL performance data on TELPAS and STAAR were used to develop a regression equation where STAAR scores were regressed on TELPAS scores. The concordance table was used to identify the TELPAS score associated with the following predicted STAAR scale scores:

- the STAAR scale score achievable by guessing
- the STAAR scale score associated with the TAKS bridge cut
- the STAAR scale scores associated with the Level II phase-in 1, phase-in 2, and final recommended standards.

The TAKS bridge cut is the location on the STAAR scale that best represents the TAKS passing standard. For more information about the TAKS bridge study and results, refer to the *STAAR Standard Setting Technical Report* and the *TAKS Equivalent Information (Bridge Study) Tables*, available on the TEA website (http://www.tea.state.tx.us/student.assessment/staar/performance-standards/).

TELPAS and STAAR Decision Consistency

In this study, decision consistency was defined as the percent of students who passed STAAR at the Level II standard (phase-in and recommended) and scored advanced high on TELPAS, or

who failed STAAR and scored below advanced high on TELPAS. Decision consistency was calculated for a variety of possible advanced high cut scores. The point along the TELPAS scale where decision consistency was maximized was used to inform the identification of reasonable ranges for the TELPAS standards review.

TELPAS IMPACT DATA

TELPAS reading impact data, or the percent of students within each of the proficiency levels, were compared based on a variety of cut score options and used to inform reasonable cut score ranges. The reading impact data were also compared to impact data observed for the three holistically-rated language domains: writing, listening, and speaking. Though the impact data for each of the TELPAS domains are not expected to be the same, it is also not expected that the percentages of students classified into the four proficiency levels would be substantially different.

P-VALUE BY PROFICIENCY LEVEL ANALYSIS

The proportion of students who answer an item correctly is called a p-value. For TELPAS reading, p-values are calculated for students in each of the four proficiency levels. P-values by proficiency level are used to classify items that measure beginning, intermediate, advanced, and advanced high levels of reading proficiency. Because p-values by proficiency level change when the proficiency level cut scores change, these values were recalculated for a variety of cut score options. Content experts considered how changes in the standards would impact item p-values and proficiency level classifications. This content analysis provided validation of neighborhood ranges that resulted in interpretations of item content that were consistent with the PLDs.

Results from these seven studies were used to inform reasonable ranges for the TELPAS proficiency level cut scores. Additional sources of information, such as the TELPAS reading vertical scale, were also considered.

Using and Displaying Validity Study Results

TEA developed a way to summarize and display the large amount of information obtained from the validity studies so that the resulting TELPAS standards would be well-aligned across grade levels, and appropriate for preparing ELLs for the level of language proficiency required for success on STAAR. During the development and implementation of the TELPAS validity studies, a plan was created to incorporate the study results into the standards review process in three ways (see Appendix 1, Table A1.6):

1. **Neighborhood Development.** Study results were used to inform reasonable ranges, or neighborhoods, for the proficiency level standards **before** the committee meetings. More information about neighborhood development will be provided in the next section.
2. **Panelist Information.** Study results were provided to panelists as background information and feedback data **during** the committee meetings.
3. **Reasonableness Review.** Study results were used to inform the reasonableness review of the recommended standards **after** the committee meetings.

Some studies were used for only one purpose; others were used for all three. For all three purposes, an important consideration was how best to display the results and communicate them clearly to a variety of audiences. After carefully considering how to present the validity study results, TEA selected three types of displays to communicate the validity study results during the TELPAS standards review process. Data displays included TELPAS reading scale charts, impact data, and vertical scale graphs. These are listed, along with the three uses described above, in Table 2.1.

Table 2.1: Displaying the TELPAS Validity Study Results

Type of Display	Display Uses
TELPAS Reading Scale Charts	Neighborhood Development
Impact Data	Neighborhood Development Panelist Information Reasonableness Review
Vertical Scale Graphs	

TEA also provided information about the text complexity analysis to the standards review panelists as background information to help inform their judgments. This included a summary of results of the text complexity study, as well as a sample of STAAR passages and TELPAS advanced high passages for the panelists to read and compare.

TELPAS Reading Scale Charts

TELPAS reading scale charts were used to show how various validity study results relate to TELPAS performance. The basis for the charts was the 2012 TELPAS raw score to scale score tables (http://www.tea.state.tx.us/student.assessment/ell/telpas/convtables/). These tables were used because most of the studies used 2012 TELPAS data and 2013 STAAR reading data so that STAAR performance would reflect an additional year of instruction for advanced high ELLs. The charts included six columns containing the following information:

- 2012 TELPAS reading raw scores
- 2012 TELPAS reading scale scores
- Proficiency levels based on the 2008 TELPAS reading standards
- TELPAS reading impact data
- TELPAS reading percent correct
- Validity study results

Linking the validity study results to a particular raw score on a TELPAS test form was useful because the audience for this information was content and psychometric staff who were very familiar with the TELPAS forms. Therefore, seeing student performance data in relation to a specific TELPAS form provided a helpful frame of reference for those involved in using and interpreting the information contained in the charts.

The process for mapping the validity study results to the TELPAS reading scale charts is detailed in Appendix 3. Different studies provided information about different cut scores. The following color scheme was used to help identify which study results informed which cuts:

- Green = Intermediate Cut
- Blue = Advanced Cut
- Red = Advanced High Cut

Using the process described in Appendix 3 and color coding of studies based on the cuts they were intended to inform, TELPAS reading scale charts were developed for all six TELPAS grade clusters. These charts are provided in Appendix 4. The charts also include neighborhood ranges that will be described in the *Developing the TELPAS Neighborhoods* section.

Although there is a considerable amount of data included in the TELPAS reading scale charts, information was not provided in this format to the committees charged with reviewing the TELPAS standards. Instead, the charts were used by TEA to set reasonable cut score ranges to help guide the standards review committee. These cut score ranges were applied to the ordered item booklet (OIB), which is a set of TELPAS items ordered from easiest to most difficult within which panelists make judgments about where the standards should be set. More information about the OIBs is provided in Chapter 3. A subset of these data was provided to panelists during the committee meetings using the impact data and vertical scale displays described next.

IMPACT DATA

Impact data were used before, during, and after the TELPAS standards review meeting. TEA reviewed impact data when determining reasonable ranges for the TELPAS reading proficiency level cut scores. More information will be provided about this in the *Developing the TELPAS Neighborhoods* section. Additionally, panelists were provided with impact data during the standards review committee meeting. Finally, TEA reviewed impact data after the committee meeting to evaluate the reasonableness of the standards across all grade clusters.

During the standards review committee meeting, impact data were provided as feedback to panelists in the following ways:

- The percentage of students within each TELPAS reading proficiency level was provided based on cut score recommendations for a specific grade cluster.
- The percentage of advanced high ELLs who would be expected to pass STAAR reading the next year based on the advanced high cut score recommendation for a particular grade. STAAR reading pass rates for non-ELLs matched on socioeconomic status were also provided as a point of comparison.
- The percentage of students within each TELPAS reading proficiency level based on the committee's recommendations across all grade clusters.

An example of each of these types of impact data displays is provided in Appendix 5.

VERTICAL SCALE GRAPHS

TELPAS reading scale scores are on a vertical scale. Because student performance on a vertical scale can be compared from grade to grade in order to gauge progress in English language development across time, the vertical scale was used to evaluate the alignment of proficiency level standards across grade clusters. As with the impact data, vertical scale information was used before, during, and after the standards review meeting:

- The alignment of neighborhoods was evaluated using the vertical scale.
- Panelists considered the reasonableness of cut score recommendations across grade clusters using the vertical scale.
- TEA evaluated the recommended TELPAS cut scores during the reasonableness review using the vertical scale.

An example of these types of vertical scale displays is provided in Appendix 5.

TELPAS Neighborhood Development

PURPOSE OF NEIGHBORHOODS

Neighborhoods are an essential aspect of the evidence-based standard-setting method used with TELPAS. Creating neighborhoods provides a way to synthesize important policy considerations, content considerations, and validity study results into a set of reasonable ranges within which standards review committee members can make judgments. The groundwork involved in the development of the neighborhoods requires careful consideration prior to the standard setting meeting. However, the neighborhoods allow the committee to focus on key content and data considerations, while still placing cut scores within regions that have been developed using all relevant considerations.

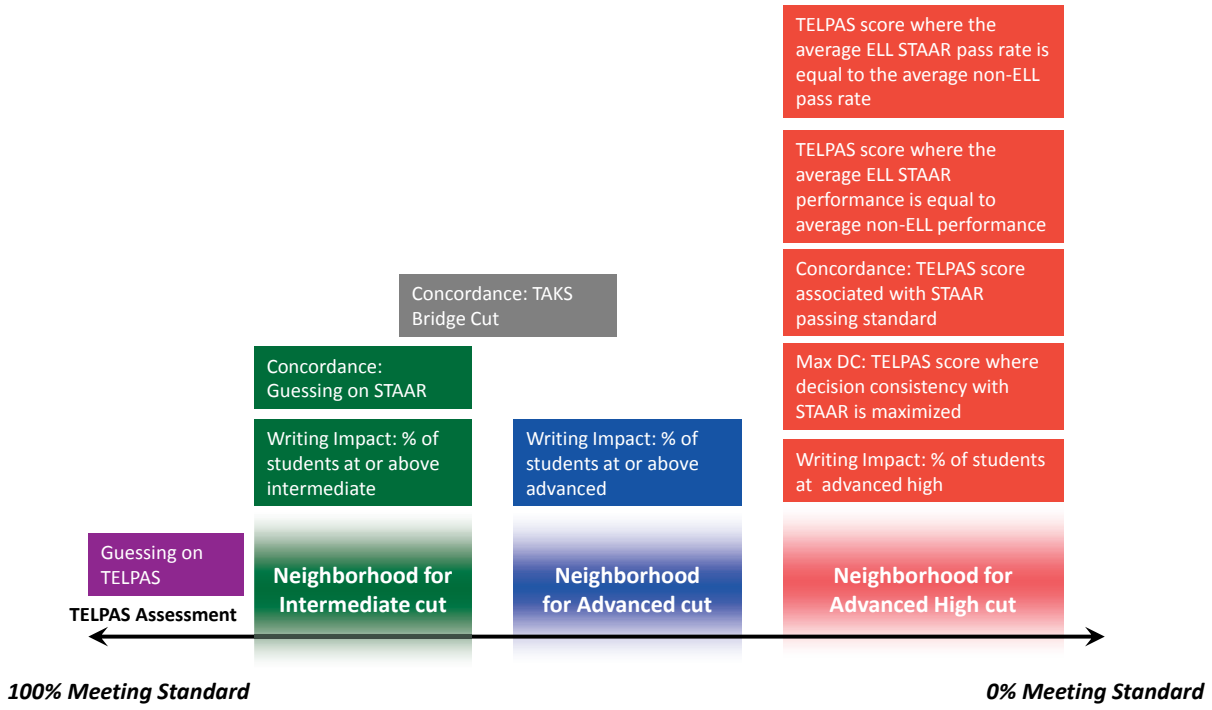
Using neighborhoods for the TELPAS standards review also was preferred in light of the fact that this approach was used successfully with the STAAR assessments. Although the purpose of neighborhood development in STAAR was somewhat different, use of a similar neighborhood development approach with TELPAS was consistent with the desire to better align TELPAS with the STAAR assessments.

DEVELOPING NEIGHBORHOOD GUIDING PRINCIPLES

In the previous section (*Using and Displaying Validity Study Results*), the first type of data display described was the TELPAS reading scale charts. These charts, included in Appendix 4, include information from the 2012 TELPAS raw score to scale score tables and the results of the validity studies. In Appendix 3, the method for mapping the validity studies to the TELPAS scale score is provided. After the studies are mapped to the scale, a process is needed for using the TELPAS reading scale charts to define neighborhoods.

Figure 2.1 provides a graphical representation of how the empirical study information was intended to inform the neighborhoods. As can be seen, most of the study results pertained to the advanced high cut score. An evaluation of the study results provided in the TELPAS reading scale charts in Appendix 4 indicates that the study results often spread over a large region of the TELPAS scale. This is especially true of the validity studies used to inform the advanced high cut. Moreover, the studies map to different regions of the score scale at different grade clusters.

Figure 2.1: Graphical Illustration of TELPAS Neighborhood Development using Validity Study Results
NOTE: DC = Decision Consistency.

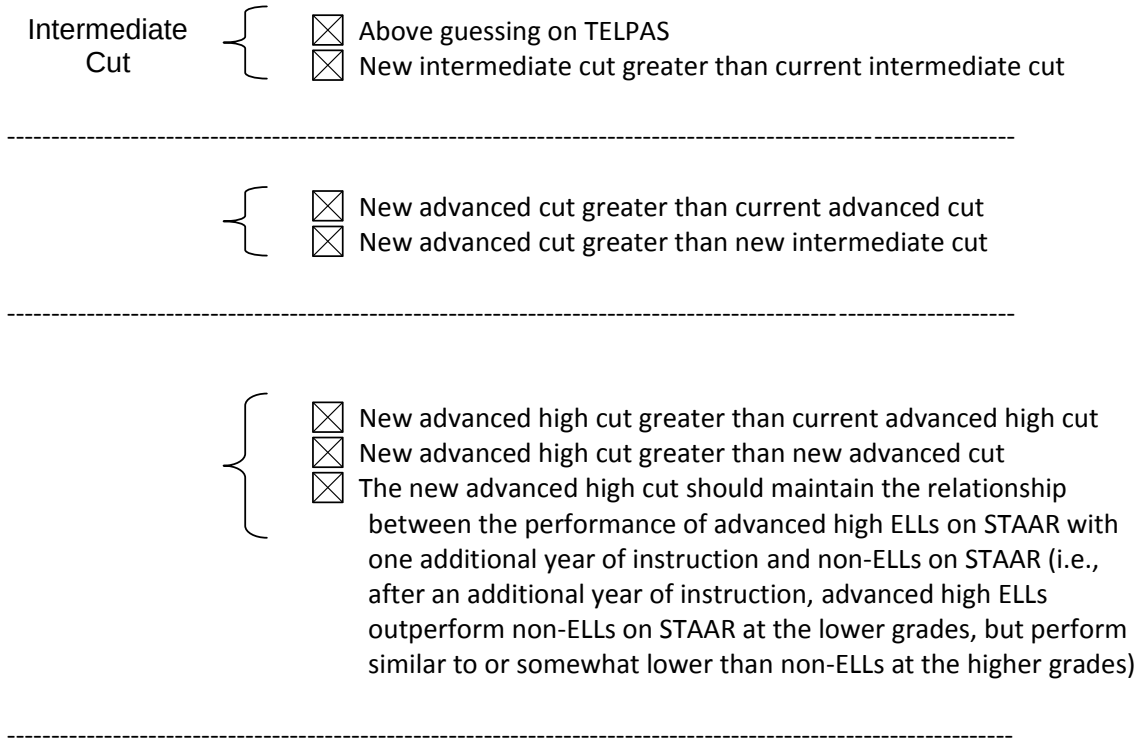


TEA developed a set of guiding principles to help to narrow and obtain more reasonable neighborhoods. The general guiding principles for the TELPAS neighborhoods included:

- Rigor of TELPAS reading cuts will not be lowered
- Aligned proficiency level standards across grade clusters
- Proficiency level standards informed by validity study results
- Impact data for holistic domains considered
- Item classification changes make sense in light of the PLDs
- Measurement precision where the cut scores are set
- Reasonable reading raw score cuts
- Reasonable reading impact data

These guiding principles incorporate the studies from Figure 2.1 as well as information from the p-value by proficiency level analysis, vertical scale information, impact data, and the current TELPAS proficiency levels to inform neighborhoods for all three proficiency level cut scores. Guiding principles were also developed for each proficiency level (see Figure 2.2).

Figure 2.2: TELPAS Neighborhood Development Guidelines



DEVELOPING AND EVALUATING NEIGHBORHOODS

Using the guiding principles, neighborhoods were developed and evaluated using the following steps.

Step 1. Using the TELPAS reading scale charts and the guiding principles, preliminary neighborhoods were created. Intermediate, advanced, and advanced high neighborhoods were color coded (see Appendix 4) so that it was easy to evaluate the following criteria:

- TELPAS reading impact data
- TELPAS percent correct values
- Location of neighborhoods in relation to the 2008 TELPAS reading standards
- Raw scores and scale scores included in the neighborhoods
- Validity study information that fell into the neighborhood regions

The position of neighborhoods was adjusted as needed to make sure that the ranges aligned with the guiding principles.

Step 2. TELPAS reading impact data were compared across grade clusters. Impact data were compared both for the upper and lower bounds of the neighborhoods. This information was used to estimate the percent of students within each of the TELPAS proficiency levels based on the new TELPAS standards. Additionally, this information was used to make sure that the neighborhoods were well aligned across grade clusters. Appendix 6 provides the impact data used to compare neighborhoods across the TELPAS grade clusters.

Step 3. TELPAS vertical scale information was also used to evaluate the neighborhoods across grade clusters. The upper and lower boundaries of the neighborhoods were plotted for each grade cluster to make sure the neighborhoods appeared reasonable across grades. The vertical scale chart for the final neighborhoods is provided in Appendix 5. There were a couple of places where the vertical scale suggested reconsideration of the neighborhood ranges. Specifically, the advanced neighborhood for grade 3, and the intermediate neighborhood for grade 8-9 appeared narrow compared to the neighborhoods for the other grade clusters. However, based on the content analysis described next, the neighborhoods were not adjusted.

Step 4. One of the most important components of the neighborhood development process was the content analysis that was used to adjust and validate the initial neighborhood ranges. The content analysis was used to satisfy the guiding principle of "item classification changes make sense in light of the PLDs". During the content analysis, p-value by proficiency level information was calculated based on the lower-bound neighborhood cut scores and the upper-bound neighborhood cut scores. Changes in p-value information and item classifications were compared to the item content. Neighborhoods were adjusted as necessary so that the classification of an item into a particular proficiency level accurately reflected the proficiency level the item was intended to measure, based on the PLDs. In other words, if the content of an item would classify the item as intermediate based on the PLDs, the neighborhood was set so that the item would be classified as intermediate.

Using the four steps described above, neighborhoods were obtained. These neighborhoods balanced the in-depth content analysis of TELPAS items with study information from the STAAR pass rate comparison, the average STAAR score comparison, concordance tables, decision consistency analyses, and impact data.

Chapter 3: Standards Review

This chapter provides details about the standards review meetings, the reasonableness review process, and the approval and implementation of the new TELPAS reading proficiency level standards. The sections in this chapter include:

- Purpose of Standards Review Committee Meetings
- Committee Composition and Attendees
- Meeting Proceedings
- Recommended TELPAS Reading Cut Scores
- Reasonableness Review
- Approval and Implementation of TELPAS Reading Cut Scores

Purpose of Standards Review Committee Meetings

All standard settings and standards reviews activities are based to a large degree on educator judgment. Educators that are part of the standards review committees use their experience and knowledge to make expert recommendations. These judgments help establish the criteria for interpreting test scores using a specific standard-setting method. The purpose of holding TELPAS standards review meetings was to gather expert recommendations for the proficiency level standards on each TELPAS reading assessment in grades 2–12.

Each committee was asked to recommend three cut scores for a given TELPAS reading assessment. The *Intermediate* cut distinguishes between ELLs who are in the beginning and intermediate proficiency levels; the *Advanced* cut distinguishes between the intermediate and advanced proficiency levels; and the *Advanced High* cut distinguishes between ELLs classified as advanced and advanced high in TELPAS reading.

The committees used the following types of information to make their judgments:

- TELPAS reading proficiency levels and global definitions
- TELPAS reading proficiency level descriptors
- TELPAS reading test questions
- TELPAS and STAAR reading passages
- reasonable ranges (or neighborhoods) within which each cut score should fall
- student performance data on TELPAS reading and STAAR reading

Committee Composition and Attendees

Three standards review committees were convened to recommend cut scores on the TELPAS reading assessments. Each committee focused on two of the TELPAS grades or grade clusters:

- grades 2 and 3
- grades 4–5 and 6–7

- grades 8–9 and 10–12

When selecting panelists for the standards review committee, TEA placed an emphasis on content knowledge, language acquisition expertise, and classroom experience, particularly with ELLs. However, the judgments and cut score recommendations made by the committees were also guided by empirical studies, both through the neighborhoods and as feedback provided after each round of judgment.

The tables in Appendix 7 summarize the characteristics and experience of the panelists on each TELPAS reading standards review committee. These tables provide demographic information about the committee members, as well as information about the members' current positions in education, the number of years they have been in their positions, their experience working with the various types of student populations, and the types of districts they represent.

Meeting Proceedings

On August 5–7, 2013, three standards review committees were convened to review the proficiency level standards on the TELPAS reading assessments. The committees recommended a total of 18 cut scores—three cut scores (Intermediate, Advanced, and Advanced High) for each of the six TELPAS reading assessments. Table 3.1 shows the agenda for the standards review committee meetings.

Table 3.1: TELPAS Reading Standards Review Committee Meeting Agenda

General Session	• Welcome and Introductions • Background Information • Overview of Standards Review • ELPS and PLDs • Borderline Students
Breakout Sessions	• Standards Review Training • Round 1: Judgment and Feedback* • Round 2: Judgment and Feedback* • Round 3: Judgment*
Closing Session	• Vertical Articulation • Process Evaluation

*These tasks were repeated for each assessment for which the committee was recommending standards.

A description of each topic in the agenda is provided next.

GENERAL SESSION

The general session was attended by panelists from all three standards review committees. The purpose of the general session was to welcome the standards review committees; to give background information about TELPAS, the transition from TAKS to STAAR, and standard setting; and to describe the standard-setting committees' responsibilities. The panelists also reviewed the English Language Proficiency Standards (ELPS) global definitions and PLDs, examined the text complexity of STAAR and TELPAS advanced high passages, and generated descriptors for borderline students.

Welcome and Introductions

TEA welcomed the panelists. The facilitators of the standards review meetings were introduced and general housekeeping tasks were covered, including the non-disclosure agreement, security protocols, and reimbursement forms. Committee members were introduced once panelists had moved to their breakout sessions.

Background Information

Panelists were provided with the history and purpose of the TELPAS program, an overview of the difference in curricula and rigor between TAKS and STAAR, and a description of the text complexity evaluations conducted between STAAR and TELPAS passages.

Overview of Standards Review

To help panelists understand what a standards review is and the reason they were asked to be part of a standards review committee, facilitators discussed the purpose of reviewing standards and provided panelists with information about the specific approach being used during the meetings.

The panelists also had an opportunity to take a sample test form for each TELPAS reading assessment on which they would be reviewing standards. The goal was for each committee member to see what a test form looks like and get a feel for the types of items and content, as well as the depth of knowledge required by the TELPAS reading assessments. After taking a sample assessment, panelists checked their responses and discussed the test-taking experience.

ELPS and PLDs

To help inform discussions, facilitators directed panelists to review the global definitions and PLDs for the reading domain in the ELPS. The PLDs are a framework for a common understanding of the knowledge, skills, and abilities possessed by an ELL at each proficiency level (beginning, intermediate, advanced, and advanced high) to understand and use English in grade-level academic settings. The PLDs gave the panelists guidance about what ELLs should know and be able to do within each proficiency level for the TELPAS reading assessments. When reviewing the PLDs, panelists were asked to think about what most differentiates advanced high ELLs from advanced ELLs, advanced ELLs from intermediate ELLs, and intermediate ELLs from beginning ELLs.

The panelists were also provided with example TELPAS advanced high and STAAR reading passages. They then discussed in groups the difference in text complexity they observed between the passages and the implications of these differences.

Borderline Students

After reviewing the PLDs, panelists were asked to think about the group of ELLs who would just barely reach the intermediate proficiency level, the group of ELLs who would just barely reach the advanced proficiency level, and the group of ELLs who would just barely reach the advanced high proficiency level. These are known as "borderline" students—defined as those students who have the minimum amount of English language knowledge necessary to be considered intermediate, advanced, or advanced high. Panelists were asked to work in their table groups to draft descriptors that characterize what a borderline student for intermediate, advanced, and advanced high should know and be able to do. Whereas the PLDs described the student in the middle of a proficiency level, the borderline descriptors focused on students with just enough knowledge to get them into a proficiency level. Table groups shared their borderline student descriptors. The committee as a whole then discussed each group's contribution to develop a master set of borderline descriptors for intermediate, advanced, and advanced high that applied to all TELPAS reading assessments. These descriptors were used by panelists while making their judgments throughout the remainder of the meeting.

BREAKOUT SESSIONS

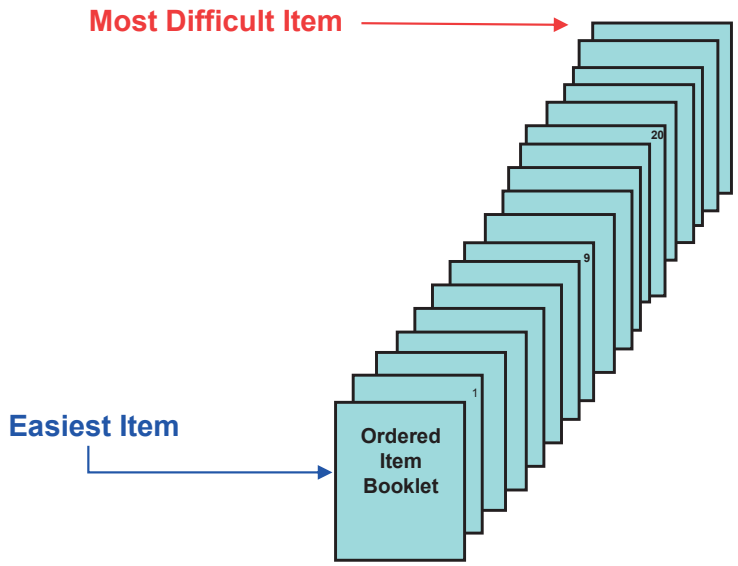
After the general session, panelists moved into their grade level specific breakout sessions. Within each committee, panelists were divided into three table groups. Each table group consisted of committee members representing various areas of expertise so that there was a blend of perspectives at each table. Table leaders were identified to facilitate the discussions and assist in meeting logistics (for example, by collecting judgment forms) at each table.

Standards Review Training

The committee members received training on the item-mapping procedure (Lewis, Mitzel, Green & Patz, 1999) that they would use to review and recommend proficiency level standards.

The item-mapping procedure, also known as the bookmark procedure, required panelists to review a set of test questions, or items, and decide which of them were likely to be answered correctly by students just barely within a given proficiency level. The items were compiled in a test booklet known as the ordered item booklet (OIB). Each OIB's test items were ordered from easiest to most difficult (see Figure 3.1). As items became progressively more difficult, panelists reviewed each item individually and decided whether a student just barely within a proficiency level (that is, a borderline student) would be likely to respond correctly.

Figure 3.1: Arrangement of Items in an Ordered Item Booklet



Given the importance of the OIB to the standard-setting process, each booklet was carefully constructed to give panelists the most information about the types of items falling within the neighborhoods, which were formed based on empirical studies. After being administered to students, test items were calibrated using the Rasch model (Rasch, 1980) to obtain Rasch item difficulty values. These values were used to order the items from easiest to most difficult in the OIBs. A sample test form was used as the starting point for each OIB.

Since the neighborhoods represented the reasonable range within which the cut scores should fall, items not part of the original test blueprint were added to the OIB in order to increase the number of items within the neighborhood bounds. This allowed panelists to make finer distinctions between items within the area where the cuts could be set.

Once the neighborhoods for each TELPAS reading assessment were formed, each OIB was evaluated to make sure that the full scale range of TELPAS reading was represented by the items in the OIB. Areas of the OIB that did not have item representation along the TELPAS reading scale were identified as gaps. This information, as well as the item's Rasch item difficulty, was used to select additional items to fill in gaps in the OIB. Areas of the OIB with an overrepresentation of items along the TELPAS reading scale were identified as clusters. For clusters, the number of items appearing in that section of the OIB was reduced.

The item-mapping procedure with a response probability (RP) value of 0.67 was used to create the OIBs and facilitate panelist judgments for meetings. That is, items were mapped to the difficulty scale at the point at which students had roughly a two thirds probability of answering the item correctly.

After the panelists were trained on the standards review method, they practiced evaluating items and making cut-score recommendations using an abbreviated "practice" OIB in order to try out the item-mapping procedure. Before making judgments, panelists were asked to read each item, identify the knowledge and skills needed for a correct response, and review the PLDs. As they made their judgments, panelists were asked to think about the borderline student and the descriptors they had previously developed. They were asked to look through the "practice" OIB, identify the last item that a borderline student would have a two thirds probability of answering correctly, and place a marker, or bookmark accordingly. After the practice session, the group discussed any questions or difficulties related to the mechanics of the item-mapping procedure.

Judgment and Feedback Rounds

For each TELPAS reading assessment, panelists took part in three rounds of judgments, in which they gave three cut-score recommendations (intermediate, advanced, and advanced high). The panelists were also provided with feedback data based on their cut-score recommendations.

Prior to making their first round of judgments, panelists were given information about the purpose of the neighborhoods and how the neighborhoods were determined using the empirical studies. The actual OIB that the panelists received was "pre-marked" with the boundaries for the intermediate, advanced and advanced high neighborhoods. There were items in each OIB that did not fall in any of the three neighborhoods. Panelists were instructed to review such items, but to focus on and to place their bookmarks only on items that were within the neighborhood for the proficiency level standard they were considering.

For each round, panelists were asked to consider the items in the OIB, starting with the easiest item. Each panelist made a cut-score recommendation for the intermediate proficiency level first, followed by a recommendation for the advanced cut score, and then the cut-score recommendation for the advanced high proficiency level.

Round 1: Judgment and Feedback

During the first round of judgments, committee members made their cut-score recommendations primarily based on the content of the OIB and the neighborhood ranges identified within the OIB. After the Round 1 judgments, the following types of feedback were presented:

- the panelist's individual Round 1 cut-score recommendations (bookmarked pages) for intermediate, advanced, and advanced high
- table-level Round 1 cut-score recommendations—the minimum, maximum, mean, and median bookmarked pages for intermediate, advanced, and advanced high
- committee-level Round 1 cut-score recommendations—the minimum, maximum, mean, and median bookmarked pages for intermediate, advanced, and advanced high
- the percentage of ELLs answering each item in the OIB correctly (p-values) by the original TELPAS proficiency level and overall

An example of committee-level Round 1 feedback can be found in Appendix 8.

Round 2: Judgment and Feedback

For the second round of judgments, committee members made their cut-score recommendations based on the first-round feedback, discussion with their table groups, and the content of the items in the OIB. After completing their Round 2 judgments, panelists were provided with the following second-round feedback:

- the panelist's individual Round 2 cut-score recommendations (bookmarked pages) for intermediate, advanced, and advanced high
- table-level Round 2 cut-score recommendations—the minimum, maximum, mean, and median bookmarked pages for intermediate, advanced, and advanced high
- committee-level Round 2 cut-score recommendations—the minimum, maximum, mean, and median bookmarked pages for intermediate, advanced, and advanced high
- TELPAS impact data, which is the percentage of ELLs in each TELPAS proficiency level, based on the committee's Round 2 cut-score recommendations
- STAAR impact data, which is the percentage of advanced high ELLs that would be expected to pass the STAAR reading assessments in the following year, based on the committee's Round 2 cut-score recommendations for advanced high

The TELPAS impact data were based on the performance of ELLs on TELPAS reading during the spring 2013 administration. The STAAR impact data were based on the performance of the group of ELLs who took TELPAS reading in spring 2012 and the next-grade level STAAR reading assessments in spring 2013[1]. An example of committee-level Round 2 feedback can be found in Appendix 8.

Round 3: Judgment

During the third round of judgments, committee members made their final individual cut-score recommendations based on all the feedback they received in the first two rounds. Panelists were not provided Round 3 feedback in their breakout sessions. It was instead given as part of the vertical articulation process (described below).

CLOSING SESSION

After each committee had completed three rounds of judgments for two TELPAS reading assessments, all three committees reconvened as a group for the closing session. The panelists were arranged in the same groups that they were in for the general session and participated in vertical articulation and process evaluation activities.

Vertical Articulation

The purpose of vertical articulation was to look at the cut-score recommendations (presented as page numbers in the OIB) that were made across all six TELPAS reading assessments and

[1] For the 10-12 TELPAS grade cluster, STAAR impact data were based on the performance of the group of ELLs who took TELPAS reading and STAAR English II reading in spring 2013. For more information, see Appendix 1.

evaluate the reasonableness of these cuts. Panelists were shown the impact data resulting from their Round 3 cut-score recommendations across all TELPAS reading assessments. They were also presented with the vertical scale score information for their recommended cut scores. Recommendations for cut-score adjustments could be made by the committee as a group after reviewing the Round 3 feedback and group discussion. Any recommended changes made during the vertical articulation had to be supported by a review of the OIB for that assessment.

Process-Evaluation Survey

At the end of the standards review meeting, panelists were asked to complete a process evaluation survey. The purpose of the survey was to collect information about each panelist's experience in recommending cut scores for the TELPAS reading assessments.

The seven-part survey asked committee members to provide feedback on

1. The level of success of the various components of the meeting
2. The usefulness of the activities conducted during the meeting
3. The adequacy of the various components of the meeting
4. How confident committee members were that the PLDs accurately reflected student performance at each proficiency level
5. How confident committee members were about the final cut-score recommendations
6. Whether committee members thought that they had been given adequate opportunities to express their professional opinions, ask questions, and interact with others
7. Whether committee members thought that their judgments and opinions had been respected by their fellow panelists and by the facilitators

Panelists were asked not to include any identifying information on the survey so that their responses would be anonymous.

Overall, virtually all committee members thought that the various components of the meeting were "successful" or "very successful." Virtually all panelists thought that the activities conducted during the meeting were either "useful" or "very useful." They also reported that the time spent on training, table discussions, and judgment tasks was "adequate" to "more than adequate." When asked about their confidence in the PLDs and the cut scores, all panelists felt "confident" or "very confident." All committee members thought that they were given adequate opportunity to express their opinions, ask questions, and interact with other committee members. Finally, all panelists felt that their opinions and judgments were respected by others. A summary of the responses to the standards review committee process evaluation is provided in Appendix 9.

Recommended TELPAS Reading Cut Scores

Recommendations Resulting from Judgment Rounds

The cut-score recommendations resulting from each round of judgment are presented (in terms of OIB page number) in Appendix 10. Descriptive statistics (including the minimum, maximum, standard deviation, mean, and median cut-score recommendations) for each round of judgment can be found in Appendix 11. Graphical representations of data regarding panelist agreement across rounds can be found in Appendix 12. In general, variation across panelist judgments decreased across rounds.

Recommendations Resulting from Vertical Articulation

The Round 3 impact data shown during the vertical articulation and impact data resulting from changes made during the articulation can be found in Appendix 13. The actual page-number recommendations resulting from the articulation can be found in Appendix 10. Table 3.2 shows the changes that were recommended during the articulation of each content area and the rationale that the committee used to support the change.

Table 3.2: Vertical Articulation Recommendations

TELPAS Reading Assessment	Articulation Result
Grade 2	No change was made.
Grade 3	No change was made.
Grades 4–5	No change was made.
Grades 6–7	The advanced high cut was lowered to align better with the advanced high cuts of the other TELPAS reading assessments.
Grades 8–9	No change was made.
Grades 10–12	The advanced high cut was lowered to align better with the advanced high cuts of the other TELPAS reading assessments.

Reasonableness Review

Following the standards review meeting, TEA conducted a reasonableness review of the cut-score recommendations across TELPAS reading grade levels and made adjustments as appropriate.

Of the eighteen cut scores recommended by the standards review committees, an adjustment was made to only one of the cut scores. The adjustment was recommended to the advanced high cut score in grades 10–12 to better align with the advanced high cut-score recommendations in the other five assessments. Table 3.3 summarizes the recommended cut scores for intermediate, advanced, and advanced high on the vertical scale score system for TELPAS reading after reasonableness review.

Table 3.3: Proficiency Level Cut Scores for TELPAS Reading After Reasonableness Review

TELPAS Reading	Intermediate	Advanced	Advanced High
Grade 2	579	645	701
Grade 3	620	674	732
Grades 4–5	643	698	766
Grades 6–7	652	712	783
Grades 8–9	661	720	796
Grades 10–12	680	737	815

A visual representation of the recommended cut scores on the TELPAS vertical scale score system is provided in Figure 3.2.

Figure 3.2: Proficiency Level Cut Scores for TELPAS Reading After Reasonableness Review

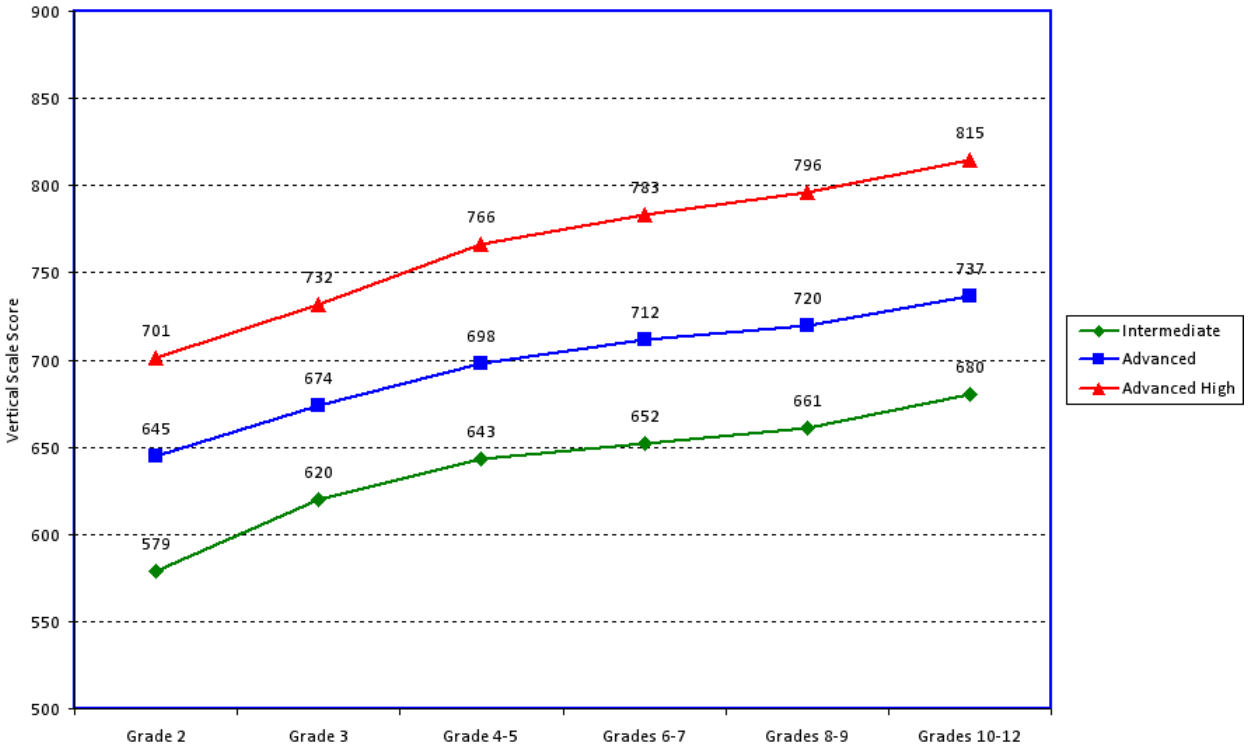


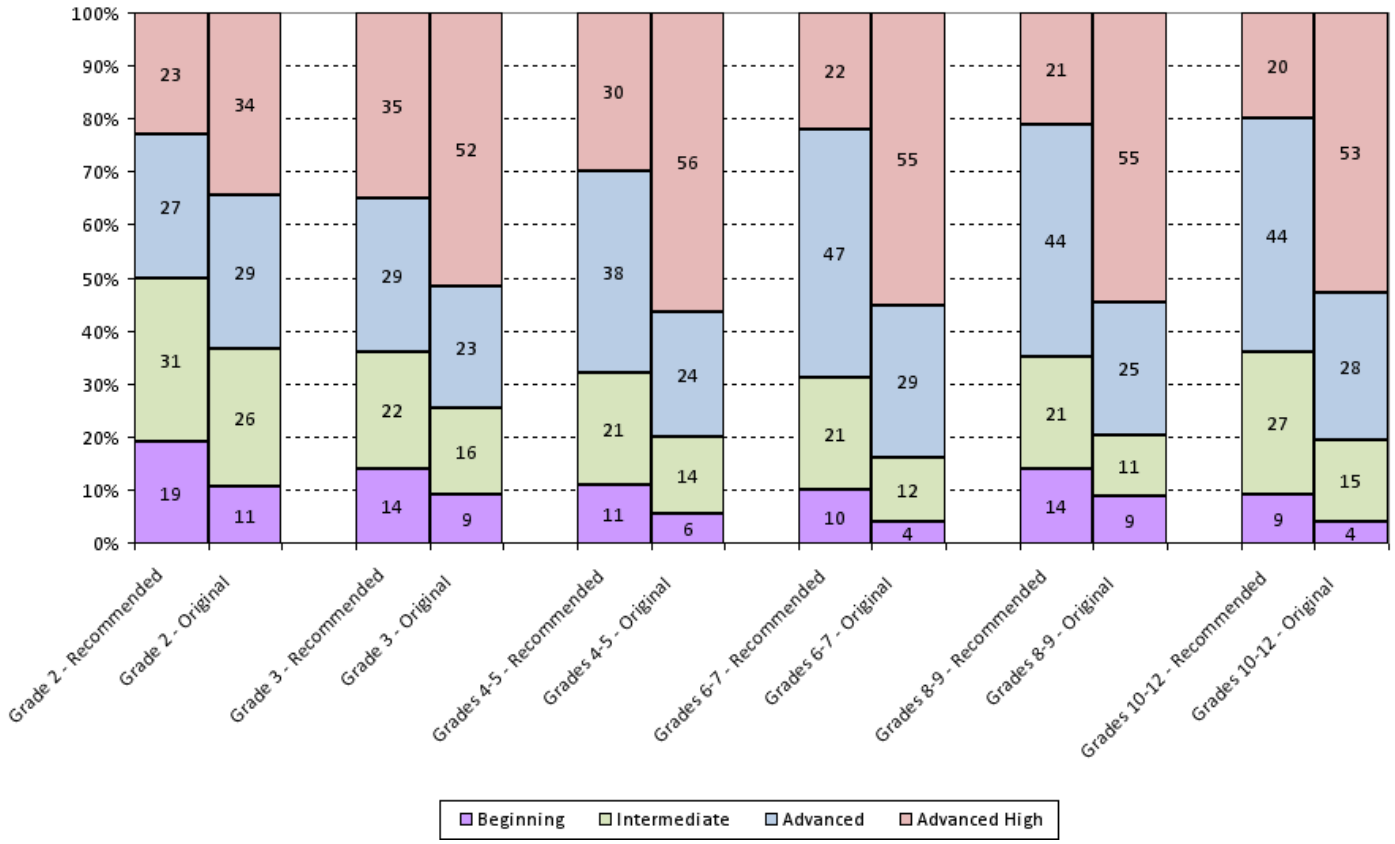
Figure 3.3 compares the impact data for the original and recommended proficiency level cut scores based on how ELLs actually performed on the TELPAS reading assessments during the spring 2013 administration.

Approval and Implementation of TELPAS Reading Cut Scores

After reasonableness review, the recommended cut scores were approved by the commissioner of education in August 2013 without any further adjustments. The approved proficiency level standards are posted on the TEA website and will be applied to TELPAS reading assessments beginning with the spring 2014 administration.

Figure 3.3: Impact Data for Recommended and Original Proficiency Level Cut Scores

NOTE: Impact data based on actual performance of ELLs on TELPAS reading during the spring 2013 administration.



Appendix 1: Validity Studies Methodological Notes

This appendix provides additional information about each of the seven validity studies conducted to inform the TELPAS standards review.

For the studies that used STAAR and TELPAS data, with one exception, the TELPAS data came from the 2012 operational administration, and the STAAR data came from the 2013 operational administration. Data from these two administrations were used so that STAAR performance could be evaluated based on an additional year of instruction after achieving advanced high on TELPAS. Student scores were matched across the two years, using a single group design, such that the TELPAS and STAAR scores were from the same group of students.

Validity studies using STAAR data for the TELPAS 10-12 grade cluster were the exception to the matching process described above. Because STAAR English III will be administered on a voluntary basis beginning in spring 2016, data were not available in 2013 to inform the 10-12 TELPAS grade cluster. To obtain some information for the 10-12 TELPAS grade cluster, a single group design was used with 2013 TELPAS grade 10 data and 2013 English II reading data. Although use of data from 2013 only does not allow for an additional year of instruction for advanced high students, these validity studies were used to provide some indication of how well aligned the 10-12 TELPAS advanced high cut was with English II reading performance.

Additionally, legislation passed in 2013 eliminates future development of the STAAR English I reading and English II reading assessments. In 2014 and beyond, the reading and writing components of the STAAR English assessments will be merged into a single assessment, rather than being administered separately as they had been previously. However, students will take many of the same types of reading items, and ELLs will need to be prepared for the same level of rigor as seen in the STAAR English reading assessments used for the validity studies conducted for the TELPAS standards review. Although the STAAR English assessments are changing in 2014, the validity study results are still believed to provide meaningful information about the alignment of TELPAS standards with the rigor of the STAAR assessments.

Finally, the results of many of the validity studies suggested that the 2008 TELPAS standards were adequate for the lower grade clusters. However, TEA made the decision to review the TELPAS standards across all grade clusters in 2013. There are several reasons for this decision.

- The TELPAS standards were set in 2008 and have not been reviewed since. In 2014, these standards will be six years old. Therefore, including all grade clusters in the standards review process is important even if the outcome in some cases is no change.
- By including all of the grade clusters in the standards review, committees can evaluate the alignment of standards as a complete set.
- The content review (p-value by proficiency level) results validated neighborhoods that were higher than original standards, even for lower grades.

In the next sections, a detailed summary of each study is provided, followed by information about how each study was used during the TELPAS standards review process.

TEXT COMPLEXITY ANALYSES

The STAAR assessment program differs from TAKS in a number of ways. STAAR has a stronger emphasis on academic rigor in terms of the cognitive demands and the level of skill needed to pass each assessment. The STAAR program was designed to be a comprehensive system, with high school curriculum and performance standards aligning and linking back to elementary and middle school and projecting forward to postsecondary readiness. STAAR reading passages are not only more rigorous compared to TAKS passages, but they also cover new genres (e.g., poetry).

Because of these changes, TELPAS advanced high passages were compared to typical passages included on STAAR. This comparison was made by content experts from both TEA and Pearson to evaluate whether the difficulty, or text complexity, of advanced high TELPAS passages would adequately prepare ELLs for the types of passages they would be expected to read and comprehend on STAAR. Although proficiency level standards are set based on items, not passages, the evaluation of text complexity provided a content link between TELPAS and STAAR that could be used to inform future development of TELPAS items and passages, as well as provide an indication of whether the 2008 definition of advanced high was too low to prepare ELLs for the rigor of STAAR.

For the text complexity comparison, TEA selected STAAR reading passages and advanced high level TELPAS passages at grades 3, 4, 6, 8 and 10/English II. TELPAS passages were chosen to represent the various genres developed at advanced high and to include topics and language that might bridge well to STAAR. Selected STAAR passages were included on operational test forms and were representative of the range of complexity and genres on STAAR.

Next, TEA developed a rubric[2] (see Appendix 2) that considered four aspects of text complexity: purpose and meaning, organization and structure, language, and knowledge demands. The rubric provided four rating categories: not complex, slightly complex, moderately complex, and very complex. The rubric was designed to be applied to both TELPAS and STAAR passages, and to result in a rating for each of the four text aspects as well as an overall rating.

Two meetings were held in April 2013 to compare the text complexity of TELPAS and STAAR reading passages. During the first meeting, TEA and Pearson ELL and English language arts content experts met to discuss differences between an academic achievement test and a language proficiency test. They also discussed text complexity differences between TAKS and STAAR. During the next meeting, TEA provided training to recalibrate participants to the rubric.

[2] The rubric was largely based on two published rubrics for evaluating the text complexity of classroom reading materials: Hess, K., & Hervey, S. (2010, updated 2011). Tools for examining text complexity. Retrieved from nciea.org/publication_PDFs/Updated toolkit-text complexity_KH12.pdf; Student Achievement Partners. (2013). Text complexity: Qualitative measures rubric. Retrieved from http://www.achievethecore.org/ela-literacy-common-core/text-complexity/qualitative-measures

Questions about language in the rubric were answered and examples were provided as needed. TEA discussed the profile of an ELL so that participants had similar frame of reference. TEA and Pearson content staff were split into three groups to evaluate elementary, middle, and high school texts. The groups completed a rating of one passage together as a practice exercise, before beginning their grade-specific evaluations.

Results of the text complexity meeting indicated that students who can read and understand advanced high TELPAS passages at the elementary levels can be reasonably expected to have enough English to access the language of the STAAR passages after one more year of instruction. At the middle school grade clusters, there was a larger gap between grade-level passages for STAAR and advanced high passages of TELPAS. Though the gap between TELPAS and STAAR at grade 8 is wider than at the lower grades, content experts agreed that students who score advanced high on the grades 8–9 TELPAS test would, with one more year of instruction, likely have enough English to access the language of STAAR passages corresponding to their grade level. However, the difference in text complexity between TELPAS advanced high passages at grades 10–12 and English II passages was found to be significant. STAAR English II passages were found to be much more complex than TELPAS advanced high passages. This included both passages that were developed specifically for STAAR (commissioned passages), and passages that were excerpted from existing literature for use on STAAR (permissioned passages). Content experts agreed that students who scored advanced high on TELPAS in grades 10–12 would need more than one additional year of instruction to have enough English to access the language of the STAAR English II passages.

These results suggest that adjustments to TELPAS passage development might be needed—especially at the 10-12 grade cluster—to adequately prepare ELLs for the rigorous STAAR English II passages. Additionally, the text complexity evaluation pointed to a need for more substantial increases in proficiency level standards at the higher grades than at the lower grades.

STAAR PASS RATE COMPARISON

To help inform the TELPAS standards review, STAAR pass rates were compared for ELLs, for former ELLs who were being monitored in their first two years of exited status (1st and 2nd year monitored students), and for non-ELLs matched on socioeconomic status. Non-ELLs matched on socio-economic status were those classified as economically disadvantaged (Econ Dis.). Comparing the STAAR pass rates of economically disadvantaged non-ELLs to the performance of ELLs provides a more realistic picture of performance gaps because the majority of ELLs in Texas are classified as economically disadvantaged, and low socio-economic status has historically been associated with lower academic achievement scores (e.g., White, 1982).

Table A1.1 provides an example of a pass rate comparison for ELLs who took TELPAS in grade 8 and STAAR English I reading in grade 9. STAAR pass rates were computed for a variety of hypothetical TELPAS advanced high cut scores, beginning with the original 2008 cut score that corresponded to a raw score of 49 on the 2012 TELPAS 8-9 test, and ending at the maximum attainable TELPAS score of 63 (100% correct). At the original TELPAS advanced high standard,

34% of grade 8 ELLs classified as advanced high in 2012 passed English I reading in 2013 at the phase-in 1 Level II STAAR standard. If the phase-in 2 standard had been the passing standard in 2013, 20% would have passed; if the final standard had been the passing standard in 2013, 12% would have passed. In comparison, approximately 60% of non-ELLs passed STAAR English I reading at the phase-in 1 standard, 46% at the phase-in 2 standard, and 35% at the final recommended standard. Therefore, advanced high ELLs with an additional year of instruction are not performing nearly as well as economically disadvantaged non-ELLs. These data suggest that an increase to the advanced high standard should be considered. As an example, if the TELPAS advanced high cut score was raised so that students had to obtain a raw score of 55 on the 2012 TELPAS test, 49% of advanced high ELLs would have passed English I reading in 2013 at phase-in 1, 31% at phase-in 2, and 20% at the final standard.

Table A1.1: STAAR English I Reading Pass Rate Comparison (Grade 8-9 TELPAS)

2012 TELPAS Raw Score	2013 STAAR Pass Rate Phase-In 1	2013 STAAR Pass Rate Phase-In 2	2013 STAAR Pass Rate Final
49	34.4	19.9	11.7
50	35.9	21.0	12.4
51	37.7	22.2	13.3
52	39.9	23.7	14.3
53	42.1	25.5	15.6
54	45.2	27.8	17.3
55	48.9	30.8	19.5
56	53.0	34.7	22.3
57	57.8	39.3	25.8
58	63.7	45.3	30.9
59	69.2	51.8	36.3
60	77.2	59.7	44.5
61	83.6	69.7	53.3
62	91.1	78.3	62.2
63	96.7	82.4	67.0
1st Year Monitored	32.4	21.3	12.7
2nd Year Monitored	50.1	33.2	21.5
Non-ELL (Econ Dis.)	60.1	46.3	35.4

NOTE: This table begins at a raw score of 49 because that was the advanced high score cut in 2012.

Tables like A1.1 were used to identify reasonable ranges for the advanced high cut score by comparing pass rates for advanced high ELLs, 1st and 2nd year monitored former ELLs, and non-ELLs matched on socioeconomic status. Differences were observed in STAAR pass rates across grade clusters. Pass rate comparisons indicated that TELPAS proficiency level standards for 2, 3, and to some extent 4-5 appeared fairly reasonable. Increases in the advanced high proficiency level standard were supported by these data for 6-7, 8-9, and 10-12, with larger changes needed for the higher grade clusters.

AVERAGE STAAR SCORE COMPARISON

Similar to the pass rate comparison, average STAAR scores were compared for ELLs, former ELLs in their first two years of monitoring, and non-ELLs matched on socioeconomic status. These average scores were compared to each other and to the Level II STAAR cut scores for phase-in 1, phase-in 2, and final recommended standards. Comparisons were used to inform reasonable ranges for the TELPAS advanced high cut score.

Table A1.2 provides average STAAR English I reading scale scores for ELLs based on hypothetical TELPAS advanced high cut scores, as well as the average STAAR English I reading scores for 1st and 2nd year monitored students and economically disadvantaged non-ELLs. These average scale scores do not change based on the phase-in standards. Instead, the scale scores associated with the phase-in standards can be used as reference points. For example, the phase-in 1 Level II standard for English I reading is 1875. However, a student who obtained the lowest score in the advanced high proficiency level on the 2012 TELPAS 8-9 reading assessment (a raw score of 49) would on average obtain a STAAR English I reading scale score of 1801 after an additional year of instruction. These data indicate that the average advanced high student would fail STAAR English I reading at the phase-in 1 standard. By comparison, the average non-ELL STAAR English I reading scale score was 1920, indicating that the average non-ELL would pass STAAR English I reading at the phase-in 1 standard.

Table A1.2: Average STAAR English I Reading Scale Scores (Grade 8-9 TELPAS Reading)

2012 TELPAS Raw Score	Average 2013 STAAR Scale Score
49	1801
50	1808
51	1816
52	1825
53	1835
54	1847
55	1862
56	1879
57	1898
58	1923
59	1949
60	1986
61	2023
62	2071
63	2109
1st Year Monitored	1770
2nd Year Monitored	1861
Non-ELL (Econ Dis.)	1920

NOTE: The STAAR English I reading scale score for the phase-in 1 Level II passing standard is 1875; the phase-in 2 Level II passing standard is 1950; the final Level II passing standard is 2000. This table begins at a raw score of 49 because that was the advanced high score cut in 2012.

As in the pass rate comparison, the average STAAR score comparison suggests that an increase in the advanced high cut score should be considered. For example, the advanced high cut would need to be increased to 56 or higher on the 2012 TELPAS grade 8-9 assessment in order for the average advanced high ELL to pass STAAR English I reading the next year. Increasing the advanced high cut to a raw score of 58 would result in the average STAAR performance of advanced high ELLs exceeding the average STAAR scale score of non-ELLs matched on socioeconomic status. This suggests that the TELPAS scale score associated with a raw score of 58 on the 2012 form might provide a reasonable upper bound to the advanced high neighborhood for the 8-9 grade cluster.

Tables like A1.2 were created for all grade clusters and used to identify reasonable ranges for the advanced high cut score. Average STAAR scale score comparisons indicated that TELPAS proficiency level standards for 2, 3, and to some extent 4-5, appeared fairly reasonable. Increases in the advanced high proficiency level standard were supported by these data for 6-7, 8-9, and 10-12, with larger changes needed for the higher grade clusters.

TELPAS AND STAAR CONCORDANCE TABLES

In cases where an empirical link between two assessments is needed but no assumptions about score equivalency can be made, regression-based approaches can be applied. Ordinary least square (OLS) regression was used to calculate a regression equation:

$$Predicted\_STAAR\_Score = a*TELPAS\_Score + b$$

where a is the slope coefficient and b is the intercept coefficient. By plugging each TELPAS score into the regression equation, a concordance table can be constructed, providing the predicted STAAR score for each TELPAS score.

Concordance tables were constructed for TELPAS reading and STAAR reading scores. However, the two assessments measure fundamentally different things. The concordance results are not meant to imply that TELPAS scores could be used with a concordance table to provide students with a STAAR score. Students must take STAAR to get an accurate measurement of their academic reading ability. Their STAAR score may or may not be similar to the score they were predicted to get using TELPAS because of the estimation error involved in developing the concordance table.

However, concordance tables can be used to inform reasonable ranges for TELPAS cut scores by indicating how students are likely to perform on STAAR reading given how they performed on TELPAS reading. For example, predicted STAAR scores from the concordance table can be compared to the STAAR passing standard (phase-in and recommended) to determine how high ELLs would have to score on TELPAS to be predicted to pass STAAR.

Table A1.3 provides an example concordance table between grade 8 TELPAS reading and STAAR English I reading for the advanced high score range on TELPAS. As indicated in Table A1.3, students who receive a score on the grade 8-9 TELPAS reading test of 62 in grade 8 are predicted to pass STAAR reading at the phase-in 1 Level II standard. A score of 62 on TELPAS is one raw score away from a perfect score. Even an 8th grade student who received a perfect score (100% correct) on the 2012 TELPAS reading test would not be predicted to pass STAAR English I reading the next year at the phase-in 2 standard. The advanced high standard for the 8-9 TELPAS reading test was at a raw score of 49 in 2012. However, students who obtained a TELPAS score of 49 in 8th grade were predicted to obtain a STAAR English I reading scale score of 1738 the next year, which is well below the phase-in 1 Level II passing standard. These data indicate that the advanced high standard established for TELPAS reading in 2008 is not high enough to be confident that advanced high students will be successful on STAAR after an additional year of instruction. This suggests that an increase in the advanced high standard should be considered. These data also suggest that the TELPAS test forms will need to be built with more difficult items so that a higher standard can be set without placing it at a perfect score.

Table A1.3: Concordance Table for Grade 8 TELPAS Reading and STAAR English I Reading

2012 TELPAS Raw Score	2013 Predicted English I Reading Scale Score
49	1738
50	1749
51	1760
52	1771
53	1782
54	1793
55	1804
56	1815
57	1826
58	1837
59	1848
60	1859
61	1870
62	1880
63	1891

NOTE: The STAAR English I reading scale score for the phase-in 1 Level II passing standard is 1875; the phase-in 2 Level II passing standard is 1950; the final Level II passing standard is 2000.

Similar results were obtained for the grade 9 TELPAS and STAAR English II concordance. At the elementary grade clusters, the results are much different. For the earliest grade clusters, the STAAR passing standards fall within the advanced or early stages of the advanced high regions of the TELPAS test, indicating that increasing the advanced high proficiency level cut score may not be needed at these grade clusters. These results are consistent with what was seen in the STAAR pass rate and average score comparisons.

TELPAS AND STAAR DECISION CONSISTENCY

This study included only ELL data and was based on a method recommended in a U.S. Department of Education report called "National Evaluation of Title III Implementation Supplemental Report: Exploring Approaches to Setting English Language Proficiency Performance Criteria and Monitoring English Learner Progress" (http://www2.ed.gov/rschstat/eval/title-iii/implementation-supplemental-report.html). This method compares "consistent decisions" that are defined as "passing" both the academic and English language proficiency assessments or "failing" both assessments. Although scores on English language proficiency assessments are not usually classified into pass/fail categories, the score used to exit students from ELL services can be used as a "passing" cut score. Holding the cut score on the academic assessment constant, while increasing the cut score for the English language proficiency assessment often results in an increase in decision consistency up to a certain point, after which decision consistency decreases. The scores close to the point where decision consistency is maximized should be considered for the cut score on the English language proficiency assessment, especially in cases where the cut score is used to make decisions about whether a student is ready to be exited from ELL services.

The Level II passing standard on STAAR reading was used to determine whether or not ELLs had passed or failed the academic assessment. Due to the phase-in of performance standards for STAAR, there are actually three passing standards. The decision consistency analysis was conducted for all three. TELPAS does not have a passing standard, and is not used for exiting decisions. However, for the purposes of this study, scoring at or above advanced high was considered "passing" TELPAS; scoring below advanced high was considered "failing" TELPAS. Decision consistency was evaluated for each TELPAS raw score point to identify the location where the advanced high cut would maximize consistent decisions. Unlike the data in the U.S. Department of Education report, the rigor of the STAAR assessments often resulted in patterns like that shown in Figure A1.1, where decision consistency could only be maximized by setting the advanced high cut at the very top of the TELPAS scale. This was especially true for the higher grade clusters and the final recommended STAAR Level II passing standard.

Figure A1.1: Decision Consistency results for 8th graders on TELPAS Reading (2012) and 9th graders on STAAR English I Reading (2013) at the final recommended STAAR Level II performance standard

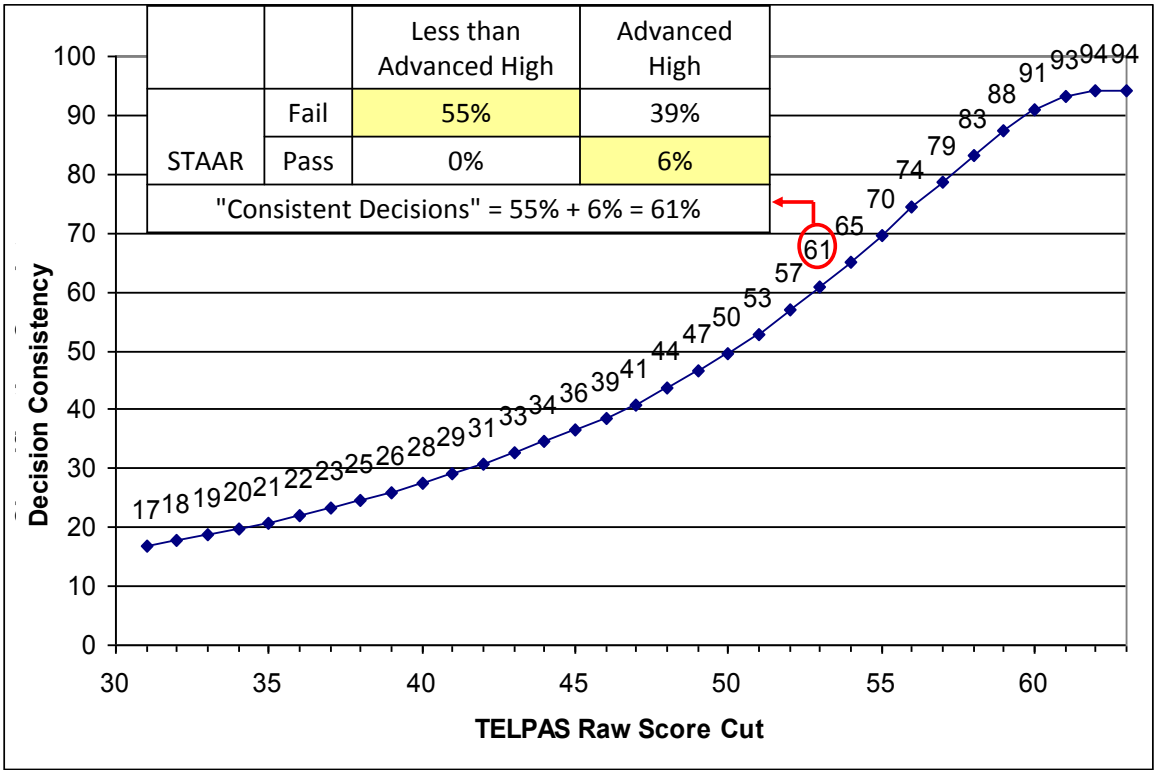


Figure A1.1 provides, as an example, the decision consistency results for grade 8 TELPAS reading performance and the performance of the same group of students in 9th grade on STAAR English I reading, using the STAAR final recommended Level II performance standard. By moving the advanced high cut up to the top of the scale, 94% of students would "pass" both STAAR and TELPAS, or "fail" both STAAR and TELPAS. These results corroborated those seen in the concordance table results, indicating that the advanced high standard likely needs to be increased, and that future tests need to be built with more difficult items, especially at the higher grade clusters.

TELPAS Impact Data

So far, all of the validity studies have involved comparing TELPAS reading performance to STAAR performance. However, TELPAS reading performance can also be compared to performance on the other TELPAS domains: writing, listening, and speaking. These domains are holistically rated by teachers, using the proficiency level descriptors (PLDs). Although the holistic domains are administered and scored very differently than the online multiple-choice reading assessment, and they provide scores on different language domains, they all assess facets of English language proficiency. One way to evaluate whether the proficiency level standards for TELPAS reading appear reasonable is to compare the percentage of students that are classified into each of the four proficiency levels for reading, to the percentages for the

other domains. An example of these percentages, called impact data, are provide in Table A1.4 for the 8th grade ELLs who took TELPAS in 2012.

Table A1.4: Percent of Grade 8 ELLs Classified into Each TELPAS Proficiency Level by Domain (2012 Impact Data)

	Beginning	Intermediate	Advanced	Advanced High
Reading	7%	10%	25%	58%
Writing	6%	21%	37%	37%
Listening	4%	12%	29%	54%
Speaking	6%	16%	31%	47%

NOTE: These data are taken from the statewide summary files on the TEA website (http://www.tea.state.tx.us/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=2147507548&libID=2147507539).

Table A1.4 shows that the highest percentage of students is classified as advanced high on reading. This may indicate that the advanced high cut score is not high enough for the 8-9 grade cluster. Although the reading impact is not expected to be the same as impact data for the other domains, there is some expectation that the impact would be similar.

The percent of students scoring advanced high on Writing is the lowest. This is consistent with expectations about second-language acquisition. Reading and listening are considered "receptive" language skills that students tend to acquire first. Speaking and writing are "productive" language skills that take longer for students to master.

TELPAS impact data for all grades can be found in the statewide reports on the TEA website (http://www.tea.state.tx.us/student.assessment/ell/telpas/rpt/sum/).

In addition to comparing impact data across the four domains, TELPAS reading impact data were considered for a variety of potential cut scores. In other words, the percent of students in each of the four proficiency levels was evaluated based on different options for making the proficiency level standards more rigorous. More information about this is included in the *Using and Displaying Validity Study Results* section of Chapter 2.

P-VALUE BY PROFICIENCY LEVEL ANALYSIS

TELPAS items are developed for students of a particular proficiency level. The items are subsequently field-tested, and p-values, or the percent of students who obtained the correct answer, are calculated for the total group of test takers and for each of the four proficiency levels. Item-level data are used to evaluate whether the proficiency level the item was designed for is actually the level for which the item provides the best measurement. In some cases, the data suggest that the item measures a different proficiency level from the one for which it was developed. For example, an item may have been developed as a beginning level item. However, p-values might indicate that the item is very difficult for beginners, of moderate difficult for intermediate students, and not difficult for advanced and advanced high students. This pattern would suggest that the item is more accurate at measuring intermediate students and should be reclassified.

However, if the proficiency level standards change, the students who are classified into each of the proficiency levels also changes. This change would result in different p-values by proficiency level. If the standards are shifted to be substantially more difficult, the average performance of students in each proficiency level will increase. In other words, the lowest performing intermediate students would be classified into the beginning category, thereby increasing the average performance of students in the beginning category. The lowest performing advanced students would be classified as intermediate, thereby increasing the average performance of students in the intermediate category. Finally, the lowest performing advanced high students would be classified as advanced, increasing the average performance of both the advanced and the advanced high categories. Therefore, the p-values by proficiency level are likely to increase. An item that previously measured language proficiency best at the intermediate level might now appear too easy for students classified as intermediate; the p-value pattern might suggest the item is actually better classified as a beginning level item.

Because of the relationship between the TELPAS proficiency level standards and how students and items are classified into proficiency levels, a p-value analysis was conducted to show how p-values by proficiency level shift for various cut scores. Table A1.5 provides a set of p-values for a hypothetical item. In this example, the p-values for the item using the TELPAS standards that were set in 2008 shows that the item is difficult for beginning ELLs. Only 25% of beginning level ELLs correctly answered the item, or about the percentage that would get the item correct by guessing. Intermediate ELLs correctly answer the item 60% of the time, indicating that more than half of the intermediate ELLs can answer the item correctly. Nearly all advanced and advanced high ELLs can answer the hypothetical question correctly (98% and 99% respectively).

Table A1.5: P-Values by Proficiency Level for a Hypothetical Item

TELPAS Standard	Beginning Level P-Value	Intermediate Level P-Value	Advanced Level P-Value	Advanced High Level P-Value
2008	25%	60%	98%	99%
Moderate Increase	40%	70%	99%	100%
Large Increase	60%	95%	100%	100%

A moderate increase to each of the proficiency level cut scores might result in the pattern of p-values shown in Table A1.5 in the “Moderate Increase” row. Now 40% of beginning level ELLs can answer the item correctly, 70% of intermediate ELLs can answer it correctly, and, nearly all the advanced and advanced high ELLs can correctly answer the question. Finally, in the “Large Increase” row, a third set of p-values by proficiency level are provided, this time based on a hypothetical large shift in all of the proficiency level cut scores. Again, the p-values increase for each proficiency level. Based on the “Large Increase,” the item appears to fit the classification criteria for a beginning level item, rather than an intermediate level item.

As part of the standards review process, a set of options for possible cut score changes was considered. P-values by proficiency level were calculated for each option. The classifications of items into proficiency levels were compared across different cut score options. Finally, content

experts compared the item classifications across options, and identified options that resulted in item classifications consistent with the PLDs. For example, if, in the hypothetical example above, the content of the item suggested that the item was a beginning-level item, the "Large Increase" in standards option, but not the "Moderate Increase" option, would provide proficiency level standards that would result in an accurate classification of the item. This combination of data and content analysis was performed on hundreds of TELPAS items to identify reasonable TELPAS standards.

USING THE VALIDITY STUDIES IN THE STANDARDS REVIEW PROCESS

The validity studies played an important role in the evidence-based standards review process. Results were used **before** the standards review meeting to help inform the development of neighborhoods (neighborhood development). Results were used **during** the standards review meeting as panelist background information and feedback data (panelist information). And results were used **after** the standards review meeting to inform the reasonableness review. Table A1.6 provides a description of how each study served some or all of these purposes.

For specific information about how the studies were used to inform neighborhoods, refer to the *Developing and Evaluating Neighborhoods* section in Chapter 2 and Appendix 3.

Table A1.6: Use of the TELPAS Validity Studies in the Standards Review Process

TELPAS Validity Study		Use in Standards Review Process	Study Description and Usage
1	Text Complexity	Panelist Information	Text complexity was compared for TELPAS advanced high level passages and STAAR passages for corresponding grades. Passages and text complexity information were provided to panelists during the standard setting meetings to help them understand the level of language complexity in STAAR reading passages.
2	STAAR Pass Rate Comparison	Neighborhood Development Panelist Information Reasonableness Review	Pass rates for STAAR were calculated for each TELPAS score, for 1st and 2nd year monitored students, and for non-ELLs matched on socio-economic status. Pass rates were used to inform reasonable ranges for the new TELPAS standards. Panelists were given pass rates on STAAR for non-ELLs and for advanced high ELLs based on their recommended cut scores. For the reasonableness review meeting, STAAR pass rates based on panel recommended advanced high cut scores were reviewed across grades and compared to pass rates for non-ELLs.
3	Average STAAR Score Comparison	Neighborhood Development	The average STAAR score was computed for each TELPAS score, for 1st and 2nd year monitored students, and for non-ELLs. Comparisons of average STAAR scores for ELLs and non-ELLs were used to inform reasonable ranges for the new TELPAS standards.
4	Concordance Tables	Neighborhood Development	A regression study was conducted using TELPAS reading performance to predict STAAR reading performance. Results were used to inform reasonable ranges for the new TELPAS standards.
5	Decision Consistency Analysis	Neighborhood Development	Students classified as passing STAAR and advanced high on TELPAS or failing STAAR and scoring less than advanced high on TELPAS were considered "consistent decisions." Decision consistency was calculated for a variety of possible advanced high cut scores to see where it was maximized. The analysis was repeated for each of the STAAR phase-in standards. Results were used to inform reasonable ranges for the new TELPAS standards.

TELPAS Validity Study		Use in Standards Review Process	Study Description and Usage
6	TELPAS Impact Data	Neighborhood Development Panelist Information Reasonableness Review	The percentage of students in each TELPAS proficiency level was calculated for each domain—reading, writing, listening, and speaking. Additionally, impact data were calculated and compared for a variety of TELPAS reading proficiency level cut scores. This information was used to inform reasonable ranges for the new TELPAS standards. During the standards review meetings, panelists were provided with reading impact data so they could compare the percentage of students in each proficiency level with their knowledge of Texas ELLs. The panelists also reviewed reading impact data across grade clusters during articulation. Finally, reading impact data were considered during the reasonableness review meeting to compare the panel recommended cut scores across grade clusters.
7	P-Value By Proficiency Level	Neighborhood Development	Item p-values were calculated for each of the TELPAS proficiency levels using a variety of potential proficiency level cut-score options. P-value patterns were compared across options to determine which ones lead to item classifications consistent with the TELPAS PLDs. This process of combining item and content analyses was used to validate reasonable ranges for the new TELPAS standards.

Appendix 2: Qualitative Analysis of Text Complexity Worksheet

Passage Title _____ Passage Type _____ Grade _____

Assessment Program (circle one): STAAR / TELPAS Overall Rating _____

	Not Complex	Slightly Complex	Moderately Complex	Very Complex
Purpose and Meaning	• Purpose or theme is clear and stated explicitly in the title or early in the text • Literary texts have one clear theme or level of meaning • Informational texts have a concrete, narrow focus of presenting information	• Purpose or theme may be implicit but is usually easy to identify early in the text • Literary texts may have more than one theme or level of meaning, with levels clearly distinguishable • Informational texts have a broader focus, presenting a range of ideas with detailed examples	• Purpose or theme is implicit and revealed over the entirety of the text • Literary texts have multiple themes or levels of meaning that may be difficult to identify or separate • Informational texts may include explanation or interpretation of concepts	• Purpose or theme is implicit and subtle, revealed over the entirety of the text • Literary texts have multiple and sophisticated themes or layers of meaning that are difficult to separate and interpret • Informational texts may include analysis or evaluation of concepts
Notes and Comments:				

	Not Complex	Slightly Complex	Moderately Complex	Very Complex
Organization and Structure	• Organization is clear, chronological, and/or easy to follow • Informational texts are divided into short sections, with subheadings stating the main idea • Connections between ideas or events are clear and explicit • Simple graphics directly support and help readers interpret whole text	• Literary texts may be occasionally difficult to predict • Informational texts may include a clearly presented argument • Informational texts may be divided into longer sections, with subheadings that may imply the main idea • Connections among some ideas or events are implicit or subtle • Graphics directly support selected parts of text	• Literary texts may include a subplot, time shifts, and/or more complex characters • Informational texts may include multiple arguments and/or a somewhat complex organizational structure • Connections among ideas or events are deeper and often implicit or subtle • Interpreting graphics may be necessary for analyzing text	• Literary texts have a complex plot structure and include other sophisticated elements, such as time shifts and complex characters • Informational texts have a specialized or intricate organizational structure • Connections among ideas or events are deep, intricate, and implicit or subtle • Graphics may provide essential information not included in the text • Interpreting detailed, complex graphics is necessary for analyzing text or synthesizing concepts
Notes and Comments:				

Appendix 2: Qualitative Analysis of Text Complexity Worksheet (cont'd)

	Not Complex	Slightly Complex	Moderately Complex	Very Complex
Language	• Mainly simple sentences • Contemporary, literal, conversational language • Familiar and commonly used vocabulary • Informational texts have a simple language style, sometimes with narrative elements	• Some compound sentences with more complex constructions • Mostly contemporary and conversational language • Some academic, context-dependent, or multiple-meaning words • Context for uncommon vocabulary is usually explicit and located nearby • Informational texts may have a mix of narrative and objective styles, with some use of passive voice	• Many complex sentences, with some containing multiple phrases or clauses • Some figurative or archaic language • Much use of abstract, multiple-meaning, or academic vocabulary • Context for unfamiliar vocabulary may be implied or located far from the word • Informational texts have an objective style and passive voice	• Mainly long, complex sentences, often containing intricate details or multiple concepts • Much figurative or archaic language • Extensive use of uncommon, multiple-meaning, or precise academic vocabulary • Text provides little context for unfamiliar vocabulary; context is implied throughout the text • Informational texts have a formal, specialized disciplinary style
Notes and Comments:				

	Not Complex	Slightly Complex	Moderately Complex	Very Complex
Knowledge Demands	• Minimal assumed personal experience or background knowledge required • Experiences portrayed are everyday and common to most readers • Simple, straightforward ideas	• Some assumed personal experience, content knowledge, or general background knowledge required • Experiences portrayed are common to many readers • Both simple and complex, abstract ideas	• Literary texts assume and require knowledge of cultural, historical, etc. topics explicitly referenced in text • Informational texts assume and require much topic-related knowledge • Experiences portrayed are uncommon to most readers • A range of recognizable ideas and challenging concepts	• Literary texts assume and require extensive knowledge of cultural, historical, etc. topics implied in text • Informational texts assume and require extensive specialized, topic-related knowledge • Experiences portrayed are distinctly different from those of the typical reader • Complex, challenging, abstract, or theoretical concepts
Notes and Comments:				

Appendix 3: Mapping Validity Study Results to TELPAS Reading Scale Charts

Results from five of the seven validity studies were placed onto TELPAS reading scale charts to help inform the standards review process. The five studies were:

- STAAR Pass Rate Comparison
- Average STAAR Score Comparison
- Concordance Tables
- Decision Consistency
- TELPAS Impact

In the first four studies, TELPAS and STAAR performance are compared using a single group design. For most of the studies, student scores from 2012 TELPAS are paired with their 2013 STAAR scores. However, because English III reading was not a graduation requirement in 2013, few ELLs took the assessment. In order to obtain some data to inform the advanced high cut for the TELPAS 10-12 grade cluster assessment, grade 10 TELPAS and English II reading scores were paired in 2013. Table A3.1 shows the data sources for these studies.

Table A3.1: Sources of TELPAS and STAAR Single Group Data for the Validity Studies

TELPAS	STAAR
2012 Grade 2 (Grade 2 Reading)	2013 Grade 3 Reading
2012 Grade 3 (Grade 3 Reading)	2013 Grade 4 Reading
2012 Grade 4 (Grade 4-5 Reading)	2013 Grade 5 Reading
2012 Grade 5 (Grade 4-5 Reading)	2013 Grade 6 Reading
2012 Grade 6 (Grade 6-7 Reading)	2013 Grade 7 Reading
2012 Grade 7 (Grade 6-7 Reading)	2013 Grade 8 Reading
2012 Grade 8 (Grade 8-9 Reading)	2013 English I Reading
2012 Grade 9 (Grade 8-9 Reading)	2013 English II Reading
2013 Grade 10 (Grade 10-12 Reading)	2013 English II Reading

STAAR Pass Rate Comparison

STAAR pass rates for economically disadvantaged non-ELLs were obtained based on the phase-in 1, phase-in 2, and final recommended Level II standards. STAAR pass rates were calculated for each of the phase-in standards for the group of ELLs at or above each of the TELPAS raw scores. As the TELPAS raw score increased, the English proficiency of the ELL group increased, and the STAAR pass rates increased. The lowest TELPAS raw score at which the ELL pass rate was equal to or greater than the non-ELL pass rate was included on the TELPAS reading scale chart for each of the STAAR standards. These points were considered upper bounds for the advanced high cut score neighborhood.

AVERAGE STAAR SCORE COMPARISON

Average STAAR scale scores for economically disadvantaged non-ELLs were calculated based on 2013 performance data. The average 2013 STAAR scale score was calculated for ELLs at or above each of the TELPAS raw scores. As the TELPAS raw score increased, the English proficiency of the ELL group increased and the average STAAR scale score increased. The lowest TELPAS raw score at which the average STAAR scale score for ELLs was equal to or greater than the average STAAR scale score for non-ELLs was included on the TELPAS reading scale chart. This point was considered an upper bound for the advanced high cut score neighborhood.

CONCORDANCE TABLES

A regression equation was developed using TELPAS scores to predict STAAR scores. By plugging each possible TELPAS score into the equation, a predicted STAAR score could be found. Concordance tables include each of the observed TELPAS scores and corresponding predicted STAAR scores. There are a few STAAR scores of particular interest for the TELPAS reading scale chart.

- **The STAAR scale score that can be obtained by guessing.** The following steps were used to map STAAR guessing to the TELPAS reading score charts:

  Step 1. The STAAR raw score associated with guessing was calculated by dividing the total number of multiple-choice items by four and rounding up if needed. English I reading and English II reading include two short answer items with score values 1-3 and a weight of 3. Guessing for these tests was calculated as the number of multiple-choice items divided by four, plus three points for each short answer item.
  Step 2. Next, the STAAR scale score associated with a raw score from **Step 1** was identified. Raw score to scale score tables are available on the TEA website (http://www.tea.state.tx.us/student.assessment/convtables/).
  Step 3. Next, the concordance table was used to find the lowest predicted STAAR scale score that was equal to or greater than the scale score calculated in **Step 2**.
  Step 4. Finally, the TELPAS score associated with the STAAR scale score from **Step 3** was identified and used to map STAAR guessing to the TELPAS reading scale chart to inform the intermediate cut score neighborhood.

  Example: STAAR English I Reading

  Step 1. STAAR English I reading has 38 multiple-choice items and 2 short answer items. Using the directions from Step 1, the raw score associated with guessing would be 38/4 + 6 = 15.5, which rounds up to 16.
  Step 2. Using the 2013 raw-score-to-scale-score table for STAAR English I reading the scale score associated with 16 is found to be 1600.

Step 3. Using the concordance table[3], the lowest predicted STAAR scale score that is greater than or equal to 1600 is 1607.
Step 4. A STAAR scale score of 1607 is predicted by a TELPAS raw score of 37. Therefore, guessing on STAAR English I reading is mapped to the TELPAS scale score chart at a raw score of 37, which is consistent to what is included in Table A4.5 of Appendix 4.

- **TAKS Bridge Cut.** Because of how low the TAKS Bridge falls on the STAAR scale, it is difficult to determine which of the TELPAS proficiency level cut scores this information can be used to inform. It is too low to inform the advanced high cut score but may fall between the intermediate and advanced cuts. The following steps were used to map the TAKS bridge cut to the TELPAS reading scale chart:

  Step 1. The STAAR scale score associated with the TAKS passing standard (TAKS bridge cut) was identified. These scale scores can be found on the TEA website (http://www.tea.state.tx.us/index4.aspx?id=2147507698&menu_id=793).
  Step 2. The lowest predicted STAAR scale score greater than or equal to the scale score obtained in **Step 1** was identified in the concordance table.
  Step 3. The TELPAS raw score associated with the STAAR scale score from **Step 2** was used to put the TAKS bridge cut on the TELPAS reading scale chart.

- **The STAAR scale scores associated with phase-in 1 Level II, phase-in 2 Level II, and final recommended Level II passing standards.** The three STAAR Level II standards were mapped to the TELPAS reading scale charts using the following steps:

  Step 1. The scale scores for the three STAAR Level II standards were identified. The STAAR Level II scale score standards can be found on the TEA website (http://www.tea.state.tx.us/student.assessment/staar/performance-standards/).
  Step 2. The lowest predicted STAAR scale score greater than or equal to the phase-in 1 scale score obtained in Step 1 was identified in the concordance table.
  Step 3. The TELPAS raw score associated with the STAAR scale score from Step 2 was used to put the STAAR phase-in 1 Level II cut score on the TELPAS reading scale chart.
  Step 4. Steps 2 and 3 were repeated for the phase-in 2 and final recommended STAAR scale scores.

DECISION CONSISTENCY

Decision consistency was calculated between TELPAS and STAAR for each of the TELPAS raw scores. The decision consistency value increased as the TELPAS raw score increased for part of the scale, then, for some tests and phase-in standards, began to decrease. The point at which decision consistency was maximized was included on the TELPAS reading scale chart. Decision consistency analyses were conducted for each of the three phase-in STAAR Level II standards.

[3] The full concordance table is not provided, but a section of it can be found in Appendix 1 (see Table A1.3).

Therefore, three decision consistency values were included on the TELPAS reading scale charts. This information was used to help inform the advanced high cut score neighborhood.

TELPAS Impact

TELPAS reading impact data are provided for all of the 2012 TELPAS raw scores included in the TELPAS reading scale chart. Impact data were obtained using TELPAS performance data from the 2012 administration. The TELPAS reading impact data helped provide a reality check for the other study results. Although many of the studies suggest an increase in standards, the impact data show how many ELLs would be able to meet a higher standard.

Impact data for the holistic domains were also compared to the TELPAS reading impact data. Although impact data for all three holistic domains could have been included on the reading scale charts, writing impact data were selected for inclusion. Writing was selected because it is considered an especially important academic language skill, as is reading. Additionally, writing is considered a more challenging language domain to master because it requires producing language, rather than just receiving language. Therefore, writing impact data could be considered an upper bound for reading. Writing impact data were included on the TELPAS reading scale chart by matching the writing and reading impact data. For example, if 80% of students scored intermediate or higher on writing, then the writing impact data were included on the TELPAS reading scale chart at the point where 80% of students scored at that TELPAS reading raw score or higher. Writing impact data were used to inform each of the three reading cut score neighborhoods.

Additional Consideration for the 10-12 TELPAS Grade Cluster

Because the TELPAS 10-12 grade cluster used 2013 TELPAS raw scores and 2013 STAAR scale scores, additional steps were needed to map the results to the 2012 TELPAS reading scale charts:

Step 1. Identify the 2013 TELPAS raw score associated with each validity study (using the steps described previously).

Step 2. Use the 2013 TELPAS raw score to scale score tables to identify the 2013 TELPAS scale scores associated with the raw scores from **Step 1.** The 2013 TELPAS raw-score to scale score table can be found on the TEA website (http://www.tea.state.tx.us/student.assessment/ell/telpas/convtables/).

Step 3. Identify the scale scores on the 2012 TELPAS reading scale chart that are closest to the 2013 TELPAS scale scores identified in **Step 2**.

Step 4. Map validity study results to the TELPAS reading scale charts using the 2012 TELPAS scale scores identified in **Step 3.**

See Appendix 4 for TELPAS reading scale charts for each of the six grade clusters.

Appendix 4: TELPAS Reading Scale Charts

Table A4.1: TELPAS Reading Scale Chart for Grade Cluster 2

TELPAS Raw Score	TELPAS Scale Score	2008 Proficiency Levels	TELPAS Impact	Percent Correct	Validity Study Results
0	271	Beginning	1.000	0	
…	…	…	…	…	
17	535	Beginning	0.946	35	Concordance: Guessing Grade 3 Read
18	541	Beginning	0.933	37	
19	548	Beginning	0.918	39	
20	554	**Intermediate Cut**	0.901	41	
21	560	Intermediate	0.884	43	
22	566	Intermediate	0.864	45	
23	572	Intermediate	0.843	47	Concordance: Grade 3 Read TAKS Bridge
24	578	Intermediate	0.822	49	Writing Impact: Intermediate and Above Non-ELL Pass Rate Phase-In 1 LII Grade 3 Read
25	584	Intermediate	0.800	51	
26	590	Intermediate	0.776	53	Non-ELL Average Grade 3 Read Score
27	596	Intermediate	0.752	55	
28	602	Intermediate	0.725	57	Non-ELL Pass Rate Phase-In 2 LII Grade 3 Read
29	608	Intermediate	0.698	59	
30	614	Intermediate	0.669	61	Non-ELL Pass Rate Final LII Grade 3 Read
31	620	**Advanced Cut**	0.637	63	Concordance: Phase-In 1 LII Grade 3 Read
32	626	Advanced	0.605	65	Max DC: Phase-In 1 LII Grade 3 Read
33	632	Advanced	0.571	67	
34	638	Advanced	0.536	69	
35	644	Advanced	0.500	71	
36	651	Advanced	0.463	73	Writing Impact: Advanced and Above
37	658	Advanced	0.424	76	
38	665	Advanced	0.386	78	Concordance: Phase-In 2 LII Grade 3 Read
39	672	**Advanced High Cut**	0.347	80	
40	680	Advanced High	0.308	82	Max DC: Phase-In 2 LII Grade 3 Read
41	688	Advanced High	0.269	84	
42	698	Advanced High	0.230	86	
43	708	Advanced High	0.192	88	Writing Impact: Advanced High
44	719	Advanced High	0.154	90	Max DC: Final LII Grade 3 Read
45	732	Advanced High	0.118	92	Concordance: Final LII Grade 3 Read
46	748	Advanced High	0.083	94	
47	770	Advanced High	0.052	96	
48	805	Advanced High	0.026	98	
49	865	Advanced High	0.008	100	

NOTE: Read = reading; DC = decision consistency; LII = Level II: Satisfactory Performance. Bold cuts are the 2008 TELPAS cut scores; green cells represent the neighborhood for the intermediate cut; blue cells represent the neighborhood for the advanced cut; red cells represent the neighborhood for the advanced high cut. Study titles are color coded using the same convention if they were intended to inform one of the three cuts.

Table A4.2: TELPAS Reading Scale Chart for Grade Cluster 3

TELPAS Raw Score	TELPAS Scale Score	2008 Proficiency Levels	TELPAS Impact	Percent Correct	Validity Study Results
0	303	Beginning	1.000	0	
…	…	…	…	…	
25	594	Beginning	0.919	43	
26	599	**Intermediate Cut**	0.908	45	Writing Impact: Intermediate and Above Concordance: Guessing Grade 4 Read
27	604	Intermediate	0.896	47	
28	608	Intermediate	0.884	48	
29	613	Intermediate	0.872	50	
30	617	Intermediate	0.859	52	
31	622	Intermediate	0.845	53	
32	626	Intermediate	0.831	55	
33	631`	Intermediate	0.817	57	
34	636	Intermediate	0801	59	Non-ELL Pass Rate Phase-In 1 LII Grade 4 Read
…	…	…	…	…	
37	649	**Advanced Cut**	0.747	64	Non-ELL Average Grade 4 Read Score
38	654	Advanced	0.726	66	Concordance: Grade 4 Read TAKS Bridge
39	659	Advanced	0.705	67	Non-ELL Pass Rate Phase-In 2 LII Grade 4 Read
40	664	Advanced	0.682	69	
41	669	Advanced	0.657	71	
42	674	Advanced	0.632	72	Concordance: Phase-In 1 LII Grade 4 Read
43	680	Advanced	0.604	74	Writing Impact: Advanced and Above Non-ELL Pass Rate Final LII Grade 4 Read
44	685	Advanced	0.574	76	
45	691	Advanced	0.543	78	
46	697	Advanced	0.509	79	Max DC: Phase-In 1 LII Grade 4 Read
47	703	**Advanced High Cut**	0.474	81	
48	710	Advanced High	0.436	83	
49	717	Advanced High	0.396	84	Concordance: Phase-In 2 LII Grade 4 Read
50	725	Advanced High	0.356	86	
51	734	Advanced High	0.313	88	Max DC: Phase-In 2 LII Grade 4 Read
52	743	Advanced High	0.268	90	Writing Impact: Advanced High
53	754	Advanced High	0.221	91	
54	767	Advanced High	0.176	93	
55	783	Advanced High	0.131	95	Max DC: Final LII Grade 4 Read
56	804	Advanced High	0.088	97	Concordance: Final LII Grade 4 Read
57	840	Advanced High	0.049	98	
58	899	Advanced High	0.017	100	

NOTE: Read = reading; DC = decision consistency; LII = Level II: Satisfactory Performance. Bold cuts are the 2008 TELPAS cut scores; green cells represent the neighborhood for the intermediate cut; blue cells represent the neighborhood for the advanced cut; red cells represent the neighborhood for the advanced high cut. Study titles are color coded using the same convention if they were intended to inform one of the three cuts.

Table A4.3: TELPAS Reading Scale Chart for Grade Cluster 4-5

TELPAS Raw Score	TELPAS Scale Score	2008 Proficiency Levels	TELPAS Impact	Percent Correct	Validity Study Results
0	322	Beginning	1.000	0	
…	…	…	…	…	
23	600	Beginning	0.961	38	Concordance: Guessing Grade 5 Read
24	605	Beginning	0.954	39	
25	610	**Intermediate Cut**	0.947	41	
26	615	Intermediate	0.940	43	Writing: Intermediate and Above
27	620	Intermediate	0.932	44	Concordance: Guessing Grade 6 Read
28	625	Intermediate	0.923	46	
29	630	Intermediate	0.915	48	
30	635	Intermediate	0.905	49	
31	640	Intermediate	0.895	51	
32	644	Intermediate	0.883	52	
33	649	Intermediate	0.870	54	
…	…	…	…	…	
37	668	**Advanced Cut**	0.805	61	Concordance: Grade 5 Read TAKS Bridge
38	673	Advanced	0.785	62	
39	678	Advanced	0.764	64	
40	682	Advanced	0.741	66	
41	687	Advanced	0.717	67	Writing Impact: Advanced and Above Non-ELL Pass Rate Phase-In 1 LII Grade 5 Read
42	692	Advanced	0.691	69	
43	697	Advanced	0.664	70	Concordance: Phase-In 1 LII Grade 5 Read
44	703	Advanced	0.635	72	Average Non-ELL Grade 5 Read Score Non-ELL Pass Rate Phase-In 2 LII Grade 5 Read
45	708	Advanced	0.604	74	Concordance: Grade 6 Read TAKS Bridge
46	713	Advanced	0.574	75	Max DC: Phase-In 1 LII Grade 5 Read
47	719	**Advanced High Cut**	0.541	77	
48	725	Advanced High	0.507	79	Non-ELL Pass Rate Final LII Grade 5 Read
49	731	Advanced High	0.472	80	
50	737	Advanced High	0.435	82	
51	744	Advanced High	0.397	84	Concordance: Phase-In 2 LII Grade 5 Read Concordance: Phase-In 1 LII Grade 6 Read
52	751	Advanced High	0.359	85	Writing Impact: Advanced High Non-ELL Pass Rate Phase-In 1 LII Grade 6 Read
53	759	Advanced High	0.319	87	Max DC: Phase-In 2 LII Grade 5 Read
54	767	Advanced High	0.279	89	Max DC: Phase-In 1 LII Grade 6 Read Average Non-ELL Grade 6 Read Score Non-ELL Pass Rate Phase-In 2 LII Grade 6 Read
55	777	Advanced High	0.238	90	
56	787	Advanced High	0.196	92	Non-ELL Pass Rate Final LII Grade 6 Read
57	800	Advanced High	0.154	93	Max DC: Final LII Grade 5 Read
58	816	Advanced High	0.112	95	Max DC: Phase-In 2 LII Grade 6 Read
59	837	Advanced High	0.073	97	Concordance: Final LII Grade 5 Read
60	872	Advanced High	0.038	98	Concordance: Phase-In 2 LII Grade 6 Read Max DC: Final LII Grade 6 Read
61	931	Advanced High	0.012	100	

NOTE: Read = reading; DC = decision consistency; LII = Level II: Satisfactory Performance. Bold cuts are the 2008 TELPAS cut scores; green cells represent the neighborhood for the intermediate cut; blue cells represent the neighborhood for the advanced cut; red cells represent the neighborhood for the advanced high cut. Study titles are color coded using the same convention if they were intended to inform one of the three cuts. Concordance results were not available for grade 6 STAAR reading at the final Level II cut score because the cut score was higher than the scale score predicted by the highest grade 4-5 TELPAS raw score.

Table A4.4: TELPAS Reading Scale Chart for Grade Cluster 6-7

TELPAS Raw Score	TELPAS Scale Score	2008 Proficiency Levels	TELPAS Impact	Percent Correct	Validity Study Results
0	321	Beginning	1.000	0	
…	…	…	…	…	
24	608	Beginning	0.970	38	Concordance: Guessing Grade 8 Read
25	613	**Intermediate Cut**	0.964	40	
26	618	Intermediate	0.958	41	
27	623	Intermediate	0.952	43	Writing Impact: Intermediate and Above
28	628	Intermediate	0.946	44	Concordance: Guessing Grade 7 Read
29	633	Intermediate	0.940	46	
30	638	Intermediate	0.933	48	
…	…	…	…	…	
35	661	Intermediate	0.892	56	
…	…	…	…	…	
38	674	**Advanced Cut**	0.859	60	
…	…	…	…	…	
42	693	Advanced	0.795	67	Concordance: Grade 8 Read TAKS Bridge
43	698	Advanced	0.774	68	
44	702	Advanced	0.752	70	Writing Impact: Advanced and Above
45	707	Advanced	0.728	71	Concordance: Grade 7 Read TAKS Bridge
46	712	Advanced	0.700	73	
47	717	Advanced	0.670	75	
48	723	Advanced	0.636	76	
49	728	Advanced	0.601	78	
50	734	**Advanced High Cut**	0.562	79	Concordance: Phase-In 1 LII Grade 8 Read
51	740	Advanced High	0.521	81	
52	746	Advanced High	0.477	83	Concordance: Phase-In 1 LII Grade 7 Read Max DC: Phase-In 1 LII Grade 8 Read
53	752	Advanced High	0.433	84	Max DC: Phase-In 1 LII Grade 7 Read
54	759	Advanced High	0.386	86	Non-ELL Pass Rate Phase-In 1 LII Grade 7 Read Writing Impact: Advanced High
55	767	Advanced High	0.337	87	
56	775	Advanced High	0.287	89	Non-ELL Average Grade 7 Read Score Non-ELL Pass Rate Phase-In 1 LII Grade 8 Read Non-ELL Pass Rate Phase-In 2 LII Grade 7 Read
57	784	Advanced High	0.237	90	
58	795	Advanced High	0.188	92	Non-ELL Pass Rate Final Grade 7 Read Non-ELL Pass Rate Phase-In 2 LII Grade 8 Read
59	807	Advanced High	0.139	94	Non-ELL Average Grade 8 Read Score Max DC: Phase-In 2 LII Grade 7 Read Max DC: Phase-In 2 Grade 8 Read
60	823	Advanced High	0.095	95	Non-ELL Pass Rate Final Grade 8 Read Concordance: Phase-In 2 LII Grade 8 Read
61	844	Advanced High	0.057	97	Concordance: Phase-In 2 LII Grade 7 Read
62	879	Advanced High	0.027	98	Max DC: Final LII Grade 7 Read Max DC: Final LII Grade 8 Read
63	938	Advanced High	0.007	100	

NOTE: Read = reading; DC = decision consistency; LII = Level II: Satisfactory Performance. Bold cuts are the 2008 TELPAS cut scores; green cells represent the neighborhood for the intermediate cut; blue cells represent the neighborhood for the advanced cut; red cells represent the neighborhood for the advanced high cut. Study titles are color coded using the same convention if they were intended to inform one of the three cuts. Concordance results were not available for grade 7 or grade 8 STAAR reading at the final Level II cut scores because the cut scores were higher than the scale scores predicted by the highest grade 6-7 TELPAS raw score.

Table A4.5: TELPAS Reading Scale Chart for Grade Cluster 8-9

TELPAS Raw Score	TELPAS Scale Score	2008 Proficiency Levels	TELPAS Impact	Percent Correct	Validity Study Results
0	338	Beginning	1.000	0	
…	…	…	…	…	
26	631	Beginning	0.921	41	Writing Impact: Intermediate and Above
27	636	**Intermediate Cut**	0.913	43	
28	640	Intermediate	0.903	44	
29	645	Intermediate	0.894	46	
30	649	Intermediate	0.884	48	
31	654	Intermediate	0.874	49	
32	658	Intermediate	0.863	51	Concordance: Guessing on STAAR English II Read
33	663	Intermediate	0.853	52	
34	667	Intermediate	0.840	54	
35	672	Intermediate	0.827	56	
36	676	Intermediate	0.814	57	
37	681	**Advanced Cut**	0.799	59	Concordance: Guessing on STAAR English I Read
38	685	Advanced	0.784	60	Concordance: English I Read TAKS Bridge
…	…	…	…	…	
43	709	Advanced	0.695	68	Writing Impact Data: Advanced and Above
44	714	Advanced	0.675	70	
45	719	Advanced	0.653	71	
46	724	Advanced	0.629	73	
47	729	Advanced	0.605	75	
48	734	Advanced	0.577	76	
49	740	**Advanced High Cut**	0.547	78	
…	…	…	…	…	
54	772	Advanced High	0.355	86	Writing Impact: Advanced High
55	780	Advanced High	0.306	87	
56	788	Advanced High	0.257	89	
57	797	Advanced High	0.211	90	Concordance: Phase-In 1 LII English II Read Max DC: Phase-In 1 LII English II Read
58	808	Advanced High	0.163	92	Non-ELL Average English I Read Score Max DC: Phase-In 1 LII English I Read Non-ELL Pass Rate Phase-In 1 LII English I Read
59	821	Advanced High	0.118	94	Non-ELL Pass Rate Phase-In 2/Final LII English I Read Non-ELL Pass Rate Phase-In 1 LII English II Read
60	837	Advanced High	0.078	95	Non-ELL Average English II Read Score Max DC: Phase-In 2 LII English I Read Max DC: Phase-In 2/Final LII English II Read Non-ELL Pass Rate Phase-In 2/Final LII English II Read
61	858	Advanced High	0.044	97	
62	893	Advanced High	0.019	98	Concordance: Phase-In 1 LII English I Read Max DC: Final LII English I Read
63	952	Advanced High	0.005	100	Concordance: Phase-In 2 LII English II Read

NOTE: Read = reading; DC = decision consistency; LII = Level II: Satisfactory Performance. Bold cuts are the 2008 TELPAS cut scores; green cells represent the neighborhood for the intermediate cut; blue cells represent the neighborhood for the advanced cut; red cells represent the neighborhood for the advanced high cut. Study titles are color coded using the same convention if they were intended to inform one of the three cuts. Concordance results were not available for STAAR English I reading at the phase-in 2 and final Level II cut scores and for STAAR English II at the final Level II cut score because the cut scores were higher than the scale scores predicted by the highest grade 8-9 TELPAS raw score.

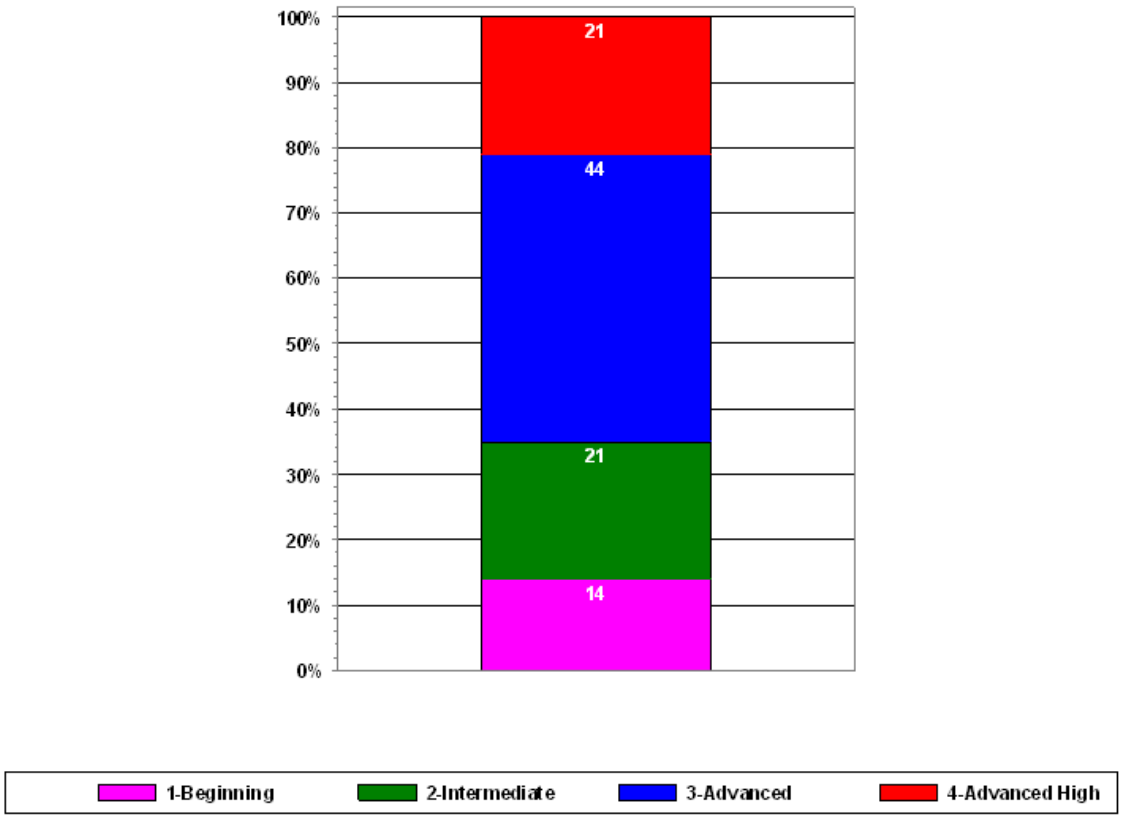
Table A4.6: TELPAS Reading Scale Chart for Grade Cluster 10-12

TELPAS Raw Score	TELPAS Scale Score	2008 Proficiency Levels	TELPAS Impact	Percent Correct	Validity Study Results
0	357	Beginning	1.000	0	
…	…	…	…	…	
22	637	Beginning	0.964	34	Writing Impact: Intermediate and Above
23	642	Beginning	0.958	36	
24	647	Intermediate Cut	0.950	38	
…	…	…	…	…	
28	666	Intermediate	0.916	44	
29	670	Intermediate	0.906	45	
30	675	Intermediate	0.896	47	
31	679	Intermediate	0.884	48	
32	684	Intermediate	0.871	50	
33	688	Intermediate	0.857	52	
34	693	Intermediate	0.843	53	Concordance: Guessing English II Read
…	…	…	…	…	
37	706	Advanced Cut	0.793	58	
38	710	Advanced	0.774	59	
39	715	Advanced	0.754	61	
40	719	Advanced	0.733	63	Writing Impact: Advanced and Above
41	724	Advanced	0.709	64	
42	728	Advanced	0.686	66	
43	733	Advanced	0.662	67	
44	737	Advanced	0.637	69	
45	742	Advanced	0.609	70	
46	747	Advanced	0.581	72	
47	752	Advanced	0.553	73	
48	757	Advanced High Cut	0.522	75	
49	762	Advanced High	0.492	77	
50	767	Advanced High	0.460	78	
51	773	Advanced High	0.427	80	
52	778	Advanced High	0.393	81	
53	785	Advanced High	0.358	83	Writing Impact: Advanced High
54	791	Advanced High	0.322	84	
55	798	Advanced High	0.283	86	Concordance: Phase-In 1 LII English II Read
56	805	Advanced High	0.246	88	Max DC: Phase-In 1 English II Read
57	814	Advanced High	0.207	89	
58	823	Advanced High	0.170	91	Non-ELL Pass Rates Phase-In 1 LII English II Read Max DC: Phase-In 2 English II Read
59	833	Advanced High	0.133	92	Non-ELL Average English II Read Score Non-ELL Pass Rates Phase-In 2/Final LII English II Read
60	846	Advanced High	0.099	94	Concordance: Phase-In 2 LII English II Read Max DC: Final LII English II Read
61	861	Advanced High	0.066	95	
62	883	Advanced High	0.040	97	
63	918	Advanced High	0.019	98	
64	977	Advanced High	0.005	100	

NOTE: Read = reading; DC = decision consistency; LII = Level II: Satisfactory Performance. Bold cuts are the 2008 TELPAS cut scores; green cells represent the neighborhood for the intermediate cut; blue cells represent the neighborhood for the advanced cut; red cells represent the neighborhood for the advanced high cut. Study titles are color coded using the same convention if they were intended to inform one of the three cuts. Concordance results were not available for STAAR English II reading at the final Level II cut score because the cut score was higher than the scale score predicted by the highest grade 10-12 TELPAS raw score.

Appendix 5: Examples of Impact and Vertical Scale Data Displays

Figure A5.1: Total Group Impact Data for TELPAS Reading Grade Cluster 8-9



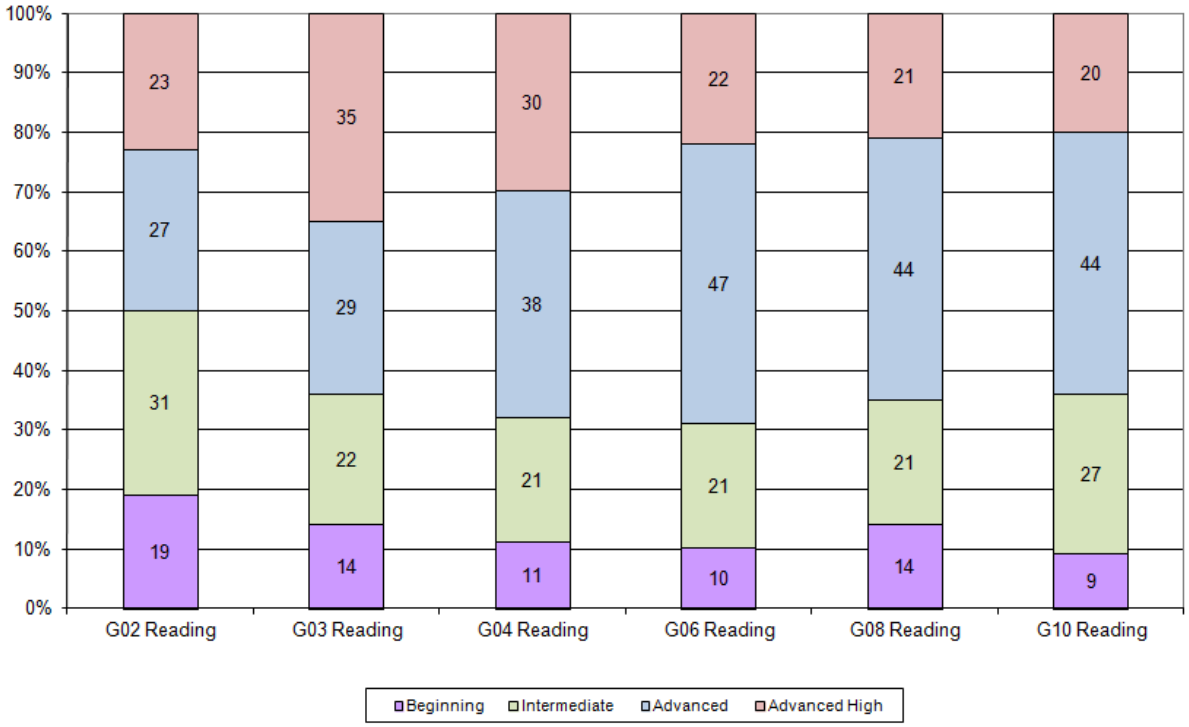
NOTE: Impact data were provided to panelists during the standards review committee meeting after Round 2. For more information about the standards review committee meeting, see Chapter 3.

Table A5.1: STAAR Impact data for Advanced High ELLs and Non-ELLs based on TELPAS Reading Grade 8-9

STAAR English I Reading	Phase-In 1 Standard	Phase-In 2 Standard	Final Standard
Advanced High	58	39	26
Non-ELL	60	46	35
STAAR English II Reading	**Phase-In 1 Standard**	**Phase-In 2 Standard**	**Final Standard**
Advanced High	66	53	43
Non-ELL	73	63	56

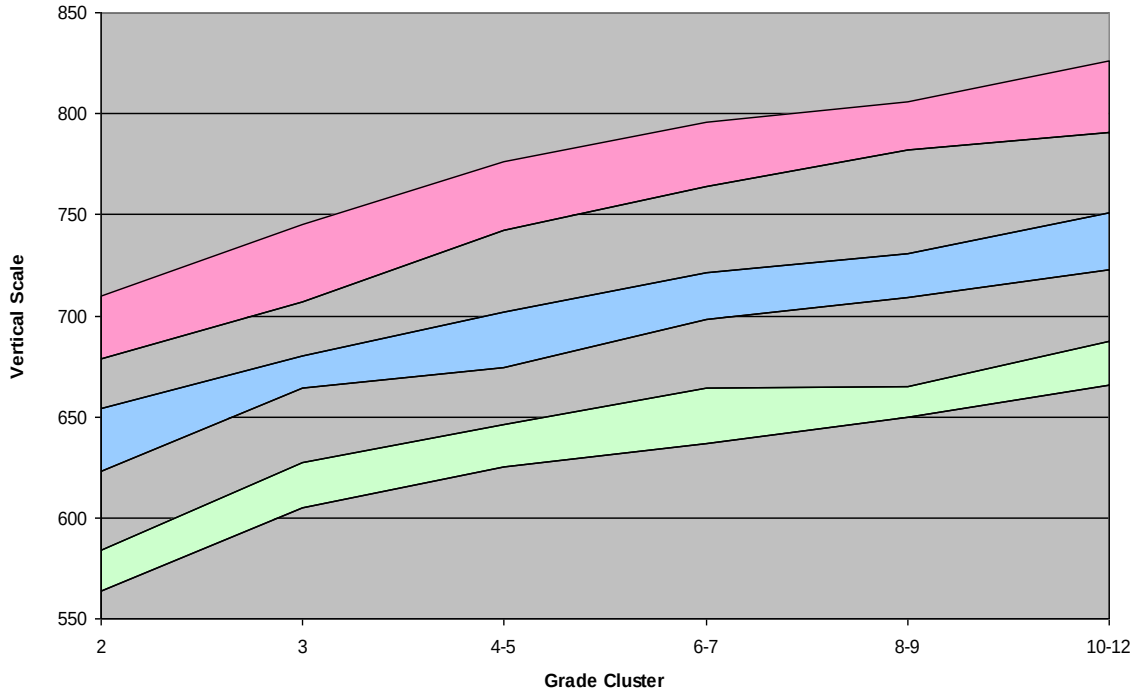
NOTE: Impact data were provided to panelists during the standards review committee meeting after Round 2. For more information about the standards review committee meeting, see Chapter 3.

Figure A5.2: TELPAS Reading Impact Data across Grade Clusters



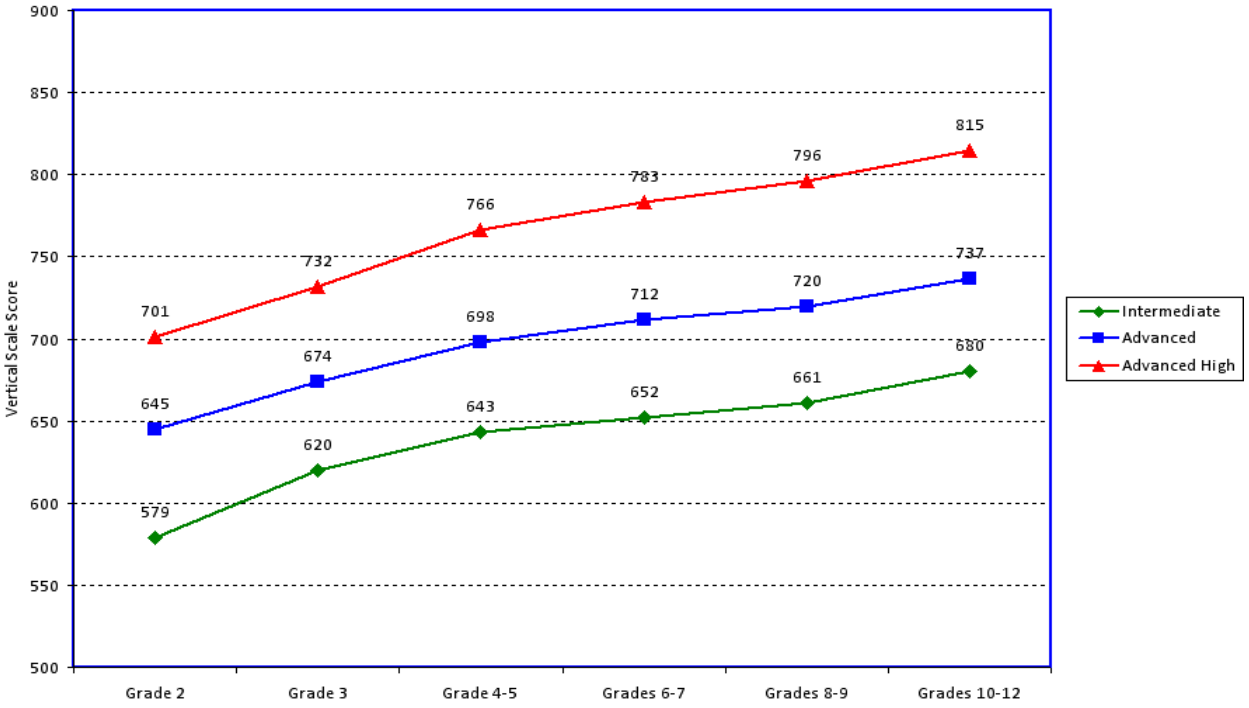
NOTE: Impact data were provided to panelists during the vertical articulation portion of the standards review committee meeting. For more information about the standards review committee meeting, see Chapter 3.

Figure A5.3: TELPAS Reading Neighborhoods on the Vertical Scale



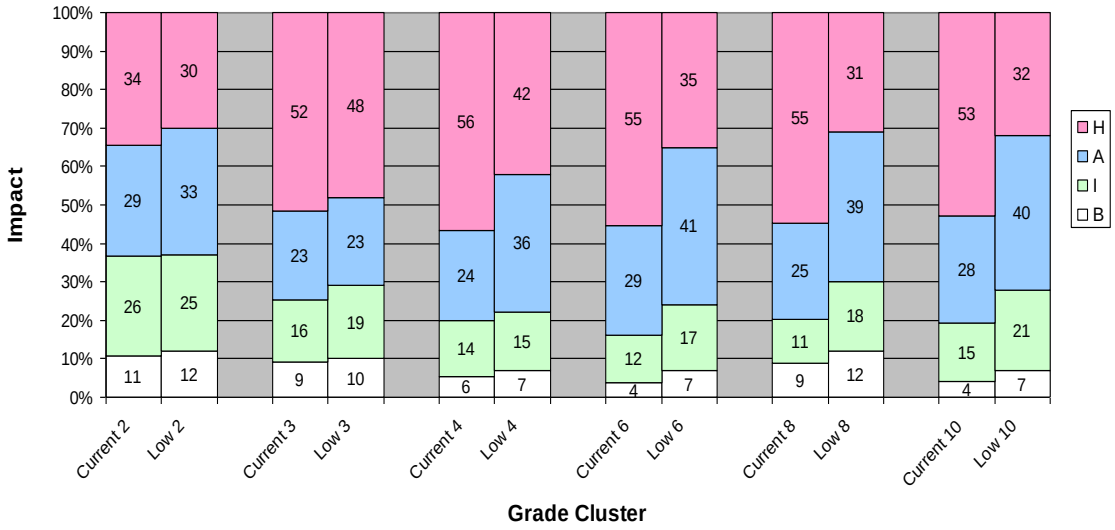
NOTE: The green region represents the neighborhood for the intermediate cut; the blue region represents the neighborhood for the advanced cut; the red region represents the region for the advanced high cut.

Figure A5.4: TELPAS Reading Cut Scores across Grade Clusters on the Vertical Scale



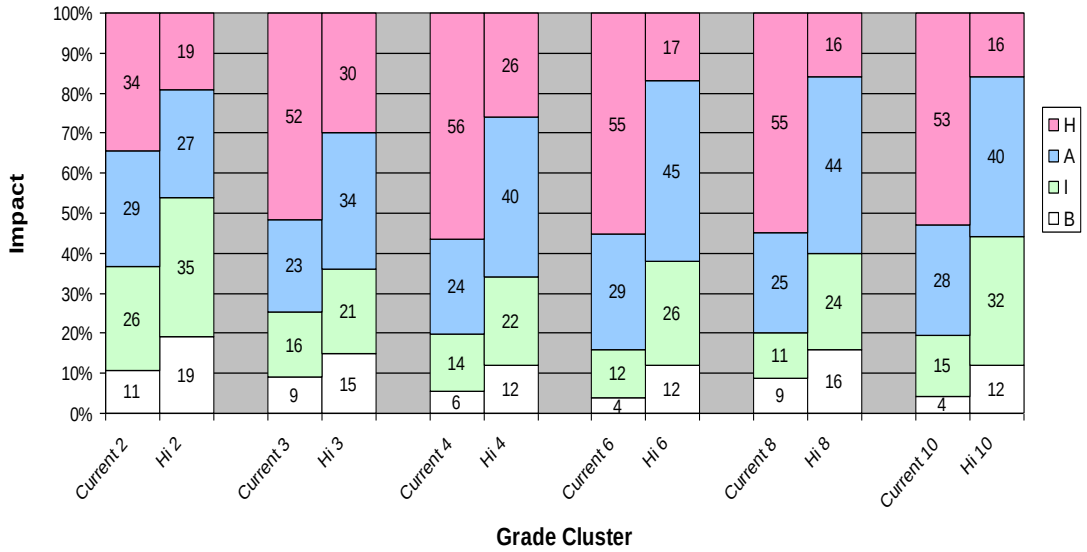
Appendix 6: Impact Data for TELPAS Reading Neighborhoods

Figure A6.1: TELPAS Reading Impact Data: 2008 Standards (Current) and Neighborhood Lower Bounds (Low).



NOTE: B = Beginning, I = Intermediate, A = Advanced, H = Advanced High.

Figure A6.2: TELPAS Reading Impact Data: 2008 Standards (Current) and Neighborhood Upper Bounds (HI).



NOTE: B = Beginning, I = Intermediate, A = Advanced, H = Advanced High.

Appendix 7: TELPAS Standard-Setting Committee Composition

GRADES 2 AND 3 COMMITTEE SUMMARY

Current Position and Years of Experience in Education

		Years of Professional Experience in Education					
		1–5 Years	6–10 Years	11–15 Years	16–20 Years	More Than 20 Years	Total
Current Position	Administrator	0	0	0	1	0	**1**
	Higher Education	0	0	0	0	0	**0**
	Teacher	1	2	0	1	1	**5**
	Other	0	0	0	0	0	**0**
	Total	**1**	**2**	**0**	**2**	**1**	**6**

Gender Distribution

Gender	N-Count
Female	5
Male	1

Ethnicity Distribution

Ethnicity	N-Count
African American	0
Asian or Pacific Islander	0
Hispanic	3
Multi-Racial	0
Native American	0
White	3

Experience with Student Populations

Student Population	N-Count
General Education	6
Special Education	5
English Language Learners	6
Low Socioeconomic Status	5

District Type

Type	N-Count
Metro	2
Suburban	3
Rural	1
Did Not Respond	0

District Size

Type	N-Count
Large	2
Medium	3
Small	1
Did Not Respond	0

District Socioeconomic Status

Type	N-Count
High	1
Moderate	3
Low	2
Did Not Respond	0

District questions apply only to those panelists currently working in a school district.

GRADES 4-5 AND 6-7 COMMITTEE SUMMARY

Current Position and Years of Experience in Education

		Years of Professional Experience in Education					
		1–5 Years	6–10 Years	11–15 Years	16–20 Years	More Than 20 Years	Total
Current Position	Administrator	0	0	1	0	0	**1**
	Higher Education	0	0	0	0	0	**0**
	Teacher	0	2	1	0	0	**3**
	Other*	0	0	1	0	2	**3**
	Total	**0**	**2**	**3**	**0**	**2**	**7**

**Other includes testing coordinators and ESC professionals.*

Gender Distribution

Gender	N-Count
Female	5
Male	2

Ethnicity Distribution

Ethnicity	N-Count
African American	0
Asian or Pacific Islander	0
Hispanic	3
Multi-Racial	0
Native American	0
White	4

Experience with Student Populations

Student Population	N-Count
General Education	7
Special Education	5
English Language Learners	7
Low Socioeconomic Status	7

District Type

Type	N-Count
Metro	0
Suburban	2
Rural	2
Did Not Respond	3

District Size

Type	N-Count
Large	3
Medium	1
Small	0
Did Not Respond	3

District Socioeconomic Status

Type	N-Count
High	0
Moderate	2
Low	2
Did Not Respond	3

District questions apply only to those panelists currently working in a school district.

GRADES 8-9 AND 10-12 COMMITTEE SUMMARY

Current Position and Years of Experience in Education

		Years of Professional Experience in Education					
		1–5 Years	6–10 Years	11–15 Years	16–20 Years	More Than 20 Years	Total
Current Position	Administrator	0	0	0	1	0	**1**
	Higher Education	0	0	0	0	1	**1**
	Teacher	0	4	0	0	1	**5**
	Other*	0	0	1	0	0	**1**
	Total	**0**	**4**	**1**	**1**	**2**	**8**

**Other includes testing coordinators and ESC professionals.*

Gender Distribution

Gender	N-Count
Female	7
Male	1

Ethnicity Distribution

Ethnicity	N-Count
African American	0
Asian or Pacific Islander	0
Hispanic	5
Multi-Racial	0
Native American	0
White	2
Did not respond	1

Experience with Student Populations

Student Population	N-Count
General Education	8
Special Education	6
English Language Learners	8
Low Socioeconomic Status	8

District Type

Type	N-Count
Metro	1
Suburban	1
Rural	4
Did Not Respond	0

District Size

Type	N-Count
Large	3
Medium	1
Small	1
Did Not Respond	1

District Socioeconomic Status

Type	N-Count
High	0
Moderate	2
Low	4
Did Not Respond	0

District questions apply only to those panelists currently working in a school district.

Appendix 8: Example Standards Review Feedback Data

This appendix provides examples of the committee-level feedback data that were presented to the standards review panelists after each round of judgment. The examples given are for TELPAS grades 4-5. Similar types of feedback data were provided for the other TELPAS grades.

For a complete summary of the panelist judgments and standard-setting meeting outcomes for each TELPAS grade, refer to Appendices 10–13.

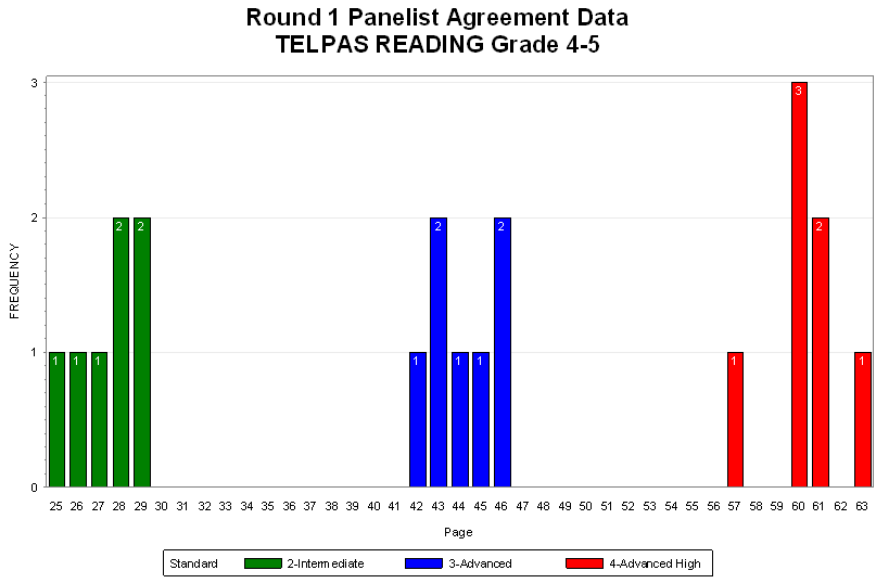
ROUND 1 FEEDBACK DATA

Figure A8.1: Summary of Cut Score Recommendations (Bookmarked Page Numbers)

TELPAS Grade 4-5 Reading —Round 1

Performance Standard	Intermediate	Advanced	Advanced High
Minimum Page Number	25	42	57
Maximum Page Number	29	46	63
Mean Page Number	27.4	44.1	60.3
Median Page Number	28	44	60

Figure A8.2: Cut Score Recommendation (Bookmarked Page Numbers) Distribution



ROUND 2 FEEDBACK DATA

Figure A8.3: Summary of Cut Score Recommendations (Bookmarked Page Numbers)

TELPAS Grade 4-5 Reading —Round 2

Performance Standard	Intermediate	Advanced	Advanced High
Minimum Page Number	27	43	60
Maximum Page Number	29	45	63
Mean Page Number	27.9	44.0	60.7
Median Page Number	28	44	60

Figure A8.4: Cut Score Recommendation (Bookmarked Page Numbers) Distribution

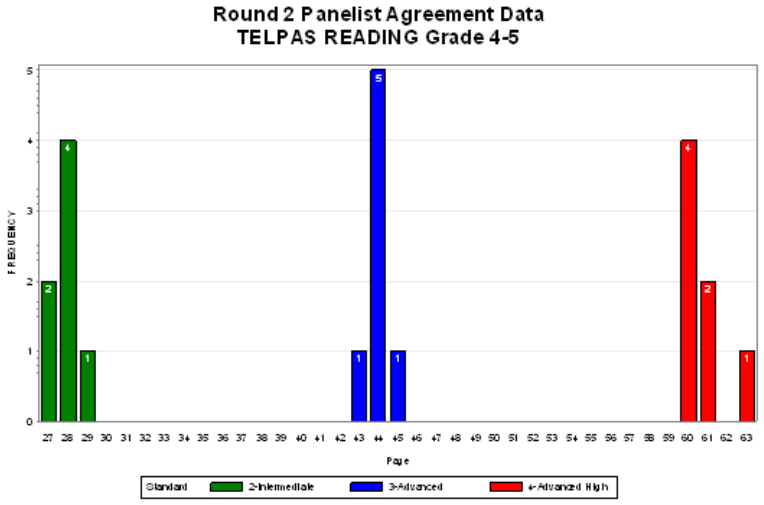


Figure A8.5: Impact Data (Total Group and By Gender) Based on Cut Score Recommendations

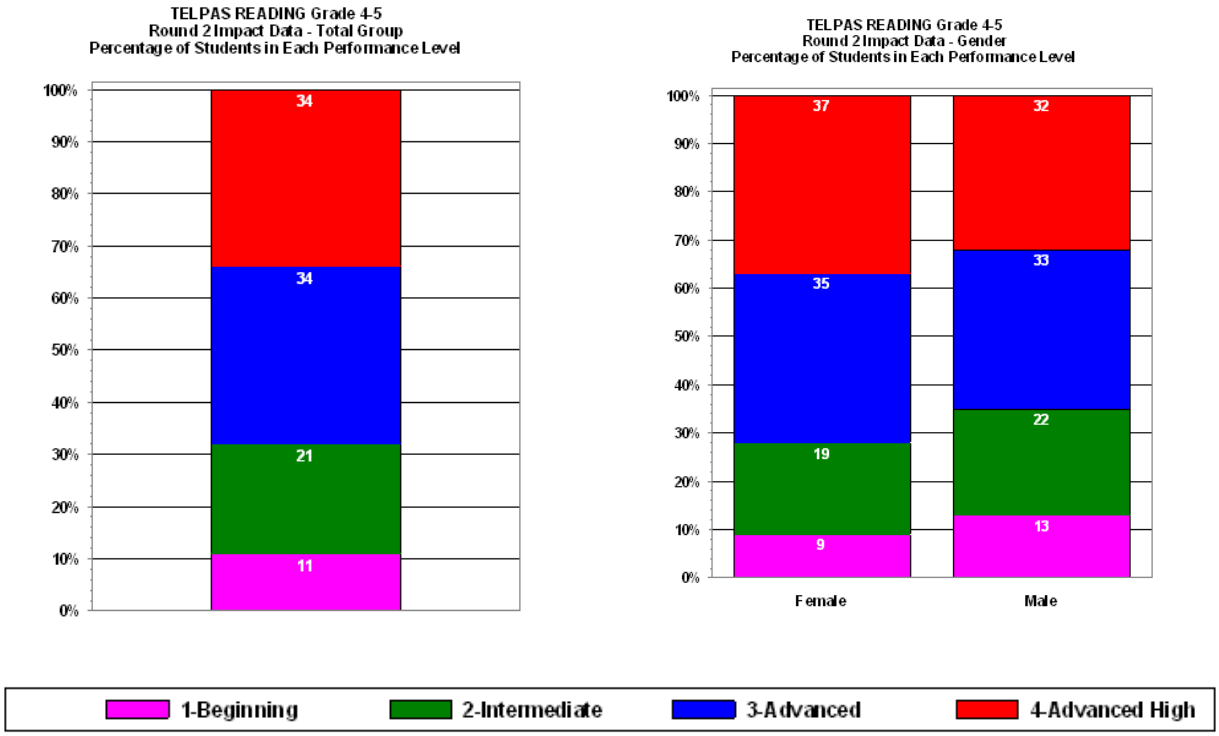


Figure A8.6: Advanced High and Non-ELL Performance on STAAR

TELPAS Grade 4-5 Reading —Round 2

STAAR Grade 5	Phase-In 1 Standard	Phase-In 2 Standard	Final Standard
Advanced High	94	77	45
Non-ELL	72	53	30
STAAR Grade 6	**Phase-In 1 Standard**	**Phase-In 2 Standard**	**Final Standard**
Advanced High	71	48	26
Non-ELL	67	51	32

Appendix 9: Standard-Setting Process Evaluation Summary

SECTION INSTRUCTIONS

The instructions provided for each section of the process evaluation survey for the standard-setting committee meetings are as follows.

- **Section 1** (Meeting Success): Check the column below that best reflects your opinion about the level of success of the various components of the meeting in which you have just participated. The activities were designed to help you both understand the process and be supportive of the recommendations made by the committee.
- **Section 2** (Usefulness of Activities and Information): How useful do you feel the following activities and/or information were in assisting you to make your recommendations?
- **Section 3** (Adequacy of Meeting Elements): How adequate were the following elements of the session?
- **Section 4** (Specific PLDs): In applying the standard-setting method, you were asked to recommend cut scores (separating four proficiency levels) for student performance on TELPAS grades 2-12 reading assessments. How confident do you feel that the specific Proficiency Level Descriptors (PLDs) are reasonable for each student proficiency level?
- **Section 5** (Cut Score Recommendations): How confident do you feel that the final cut score recommendations represent appropriate levels of student proficiency?
- **Section 6** (Opportunities to Express Opinions): Did you have adequate opportunities during the session to do the following?
- **Section 7** (Respect): Do you believe your opinions and judgments were treated with respect by the following?

A summary of responses given by each standard-setting committee is provided in the following sections.

GRADES 2 AND 3

All of the 6 panelists responded to the process evaluation survey.

Section 1: Meeting Success

Meeting Component	Not Successful	Partially Successful	Successful	Very Successful	Omit
Introduction to the process of setting proficiency level standards	0	0	1	5	0
Discussion of proficiency levels	0	1	1	4	0
Taking the actual assessment(s)	0	0	1	5	0
Overview of the item mapping procedure	0	0	1	5	0
Practice exercise for the item-mapping procedure	0	0	1	5	0
Feedback data provided in each round	0	0	0	6	0
Discussion after each round	0	0	0	6	0

Section 2: Usefulness of Activities and Information

Activity or Information	Not Useful	Somewhat Useful	Useful	Very Useful	Omit
Proficiency Level Descriptors (PLDs)	0	1	0	5	0
Training in the bookmark standard setting method	0	0	0	6	0
Feedback data provided after Round 1	0	0	0	6	0
Feedback data provided after Round 2	0	0	0	6	0

Section 3: Adequacy of Meeting Elements

Meeting Element	Not Adequate	Somewhat Adequate	Adequate	More Than Adequate	Omit
Training provided	0	0	1	5	0
Amount of time spent training	0	0	2	4	0
Feedback provided between rounds	0	0	2	4	0
Facilities used for the session	0	0	0	6	0
Total amount of time in breakout groups to make judgments	0	0	1	5	0
Number of rounds for the judgments	0	0	1	5	0

Section 4: Specific PLDs

Proficiency Level	Not Confident	Somewhat Confident	Confident	Very Confident	Omit
Beginning Proficiency Level	0	0	1	5	0
Intermediate Proficiency Level	0	0	1	5	0
Advanced Proficiency Level	0	0	1	5	0
Advanced High Proficiency Level	0	0	1	5	0

Section 5: Cut Score Recommendations

Cut Score	Not Confident	Somewhat Confident	Confident	Very Confident	Omit
Intermediate Proficiency Level	0	0	1	5	0
Advanced Proficiency Level	0	0	1	5	0
Advanced High Proficiency Level	0	0	2	4	0

Section 6: Opportunities to Express Opinions

Category	Not Adequate	Somewhat Adequate	Adequate	More Than Adequate	Omit
Express your opinions about student proficiency levels	0	0	1	5	0
Ask questions about the standards and how they will be used	0	0	0	6	0
Ask questions about the process of making cut score recommendations	0	0	0	6	0
Interact with your fellow panelists	0	0	0	6	0

Section 7: Respect

Party	No	Sometimes	Yes	Omit
Fellow panelists	0	0	6	0
Facilitators	0	0	6	0

GRADES 4-5 AND 6-7

All of the 7 panelists responded to the process evaluation survey.

Section 1: Meeting Success

Meeting Component	Not Successful	Partially Successful	Successful	Very Successful	Omit
Introduction to the process of setting proficiency level standards	0	0	1	6	0
Discussion of proficiency levels	0	0	2	5	0
Taking the actual assessment(s)	0	0	0	7	0
Overview of the item mapping procedure	0	0	1	6	0
Practice exercise for the item-mapping procedure	0	0	1	6	0
Feedback data provided in each round	0	0	0	7	0
Discussion after each round	0	0	2	5	0

Section 2: Usefulness of Activities and Information

Activity or Information	Not Useful	Somewhat Useful	Useful	Very Useful	Omit
Proficiency Level Descriptors (PLDs)	0	0	0	7	0
Training in the bookmark standard setting method	0	0	0	7	0
Feedback data provided after Round 1	0	0	0	7	0
Feedback data provided after Round 2	0	0	0	7	0

Section 3: Adequacy of Meeting Elements

Meeting Element	Not Adequate	Somewhat Adequate	Adequate	More Than Adequate	Omit
Training provided	0	0	3	4	0
Amount of time spent training	0	0	2	5	0
Feedback provided between rounds	0	0	1	6	0
Facilities used for the session	0	1	0	6	0
Total amount of time in breakout groups to make judgments	0	0	0	7	0
Number of rounds for the judgments	0	0	3	4	0

Section 4: Specific PLDs

Proficiency Level Category	Not Confident	Somewhat Confident	Confident	Very Confident	Omit
Beginning Proficiency Level	0	0	0	7	0
Intermediate Proficiency Level	0	0	1	6	0
Advanced Proficiency Level	0	0	1	6	0
Advanced High Proficiency Level	0	0	1	6	0

Section 5: Cut Score Recommendations

Cut Score	Not Confident	Somewhat Confident	Confident	Very Confident	Omit
Intermediate Proficiency Level	0	0	0	7	0
Advanced Proficiency Level	0	0	0	7	0
Advanced High Proficiency Level	0	0	1	6	0

Section 6: Opportunities to Express Opinions

Category	Not Adequate	Somewhat Adequate	Adequate	More Than Adequate	Omit
Express your opinions about student proficiency levels	0	0	1	6	0
Ask questions about the standards and how they will be used	0	0	1	5	1
Ask questions about the process of making cut score recommendations	0	0	0	7	0
Interact with your fellow panelists	0	0	0	7	0

Section 7: Respect

Party	No	Sometimes	Yes	Omit
Fellow panelists	0	0	7	0
Facilitators	0	0	7	0

GRADES 8-9 AND 10-12

7 of 8 panelists responded to the process evaluation survey.

Section 1: Meeting Success

Meeting Component	Not Successful	Partially Successful	Successful	Very Successful	Omit
Introduction to the process of setting proficiency level standards	0	0	1	6	0
Discussion of proficiency levels	0	0	1	6	0
Taking the actual assessment(s)	0	0	0	7	0
Overview of the item mapping procedure	0	0	2	5	0
Practice exercise for the item-mapping procedure	0	0	2	5	0
Feedback data provided in each round	0	0	0	7	0
Discussion after each round	0	0	0	7	0

Section 2: Usefulness of Activities and Information

Activity or Information	Not Useful	Somewhat Useful	Useful	Very Useful	Omit
Proficiency Level Descriptors (PLDs)	0	0	4	3	0
Training in the bookmark standard setting method	0	0	2	5	0
Feedback data provided after Round 1	0	0	1	6	0
Feedback data provided after Round 2	0	0	0	7	0

Section 3: Adequacy of Meeting Elements

Meeting Element	Not Adequate	Somewhat Adequate	Adequate	More Than Adequate	Omit
Training provided	0	0	1	6	0
Amount of time spent training	0	0	1	6	0
Feedback provided between rounds	0	0	0	7	0
Facilities used for the session	0	0	0	7	0
Total amount of time in breakout groups to make judgments	0	0	0	7	0
Number of rounds for the judgments	0	0	1	6	0

Section 4: Specific PLDs

Proficiency Level Category	Not Confident	Somewhat Confident	Confident	Very Confident	Omit
Beginning Proficiency Level	0	0	1	4	2
Intermediate Proficiency Level	0	0	3	4	0
Advanced Proficiency Level	0	0	4	3	0
Advanced High Proficiency Level	0	0	2	5	0

Section 5: Cut Score Recommendations

Cut Score	Not Confident	Somewhat Confident	Confident	Very Confident	Omit
Intermediate Proficiency Level	0	0	1	6	0
Advanced Proficiency Level	0	0	1	6	0
Advanced High Proficiency Level	0	0	1	6	0

Section 6: Opportunities to Express Opinions

Category	Not Adequate	Somewhat Adequate	Adequate	More Than Adequate	Omit
Express your opinions about student proficiency levels	0	0	1	6	0
Ask questions about the standards and how they will be used	0	0	0	7	0
Ask questions about the process of making cut score recommendations	0	0	1	6	0
Interact with your fellow panelists	0	0	0	7	0

Section 7: Respect

Party	No	Sometimes	Yes	Omit
Fellow panelists	0	0	7	0
Facilitators	0	0	7	0

Appendix 10: Summary of Cut-Score Recommendations

This appendix provides a summary of the cut-score recommendations (based on the OIB page number) after each judgment round of the standards review committee meetings, as well as after the vertical articulation and reasonableness review.

	Grade 2			Grade 3			Grades 4-5		
	I	A	H	I	A	H	I	A	H
Round 1	20	36	50	25	43	57	28	44	60
Round 2	21	36	50	26	44	57	28	44	60
Round 3	21	36	50	27	44	57	28	44	61
Articulation	21	36	50	27	44	57	28	44	61
Reasonableness Review	21	36	50	27	44	57	28	44	61

	Grades 6-7			Grades 8-9			Grades 10-12		
	I	A	H	I	A	H	I	A	H
Round 1	33	50	69	31	47	65	29	48	71
Round 2	33	48	69	30	47	66	29	48	70
Round 3	33	50	69	30	47	66	29	48	71
Articulation	32	50	68	30	47	66	29	48	70
Reasonableness Review	32	50	68	**30**	**47**	**66**	**29**	**48**	**69**

Appendix 11: Summary of Standards Review Panelists' Judgments

This appendix provides descriptive statistics — minimum, maximum, mean, standard deviation, and median — of the standards review panelists' cut score recommendations (based on the Ordered Item Booklet (OIB) page number) during each judgment round of the committee meetings. Statistics are given separately for each TELPAS grade or grade cluster.

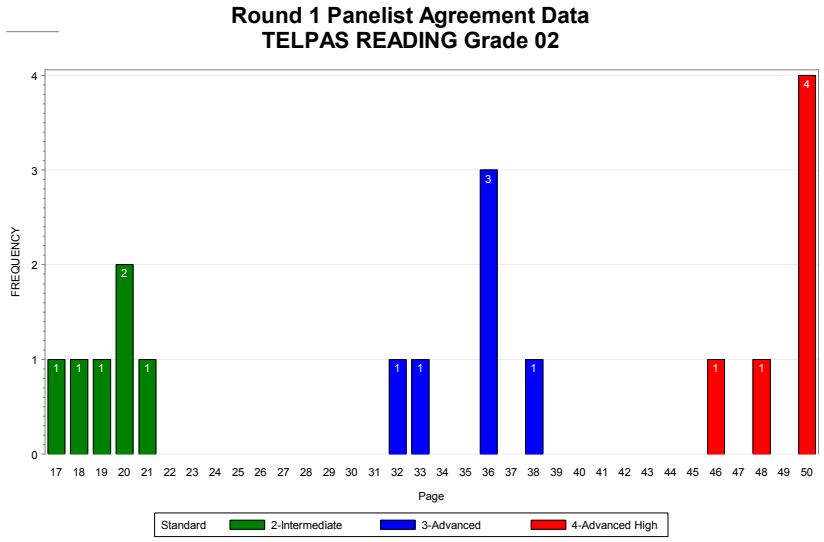
		Round	Minimum	Maximum	Mean	Standard Deviation	Median
Grade 2	I	1	17	21	19.2	1.47	20
		2	19	22	20.3	1.21	21
		3	20	22	20.7	0.82	21
	A	1	32	38	35.2	2.23	36
		2	33	37	35.7	1.37	36
		3	33	37	35.7	1.37	36
	H	1	46	50	49.0	1.67	50
		2	49	51	50.2	0.75	50
		3	50	51	50.3	0.52	50
Grade 3	I	1	24	28	25.6	1.27	25
		2	25	28	26.1	1.21	26
		3	25	27	26.4	0.98	27
	A	1	42	45	43.4	0.98	43
		2	43	45	44.1	0.90	44
		3	43	45	44.0	0.82	44
	H	1	56	59	57.1	0.90	57
		2	56	57	56.9	0.38	57
		3	57	57	57.0	0.00	57
Grades 4-5	I	1	25	29	27.4	1.51	28
		2	27	29	27.9	0.69	28
		3	27	29	28.3	0.76	28
	A	1	42	46	44.1	1.57	44
		2	43	45	44.0	0.58	44
		3	43	45	44.1	0.69	44
	H	1	57	63	60.3	1.80	60
		2	60	63	60.7	1.11	60
		3	60	63	61.3	1.11	61

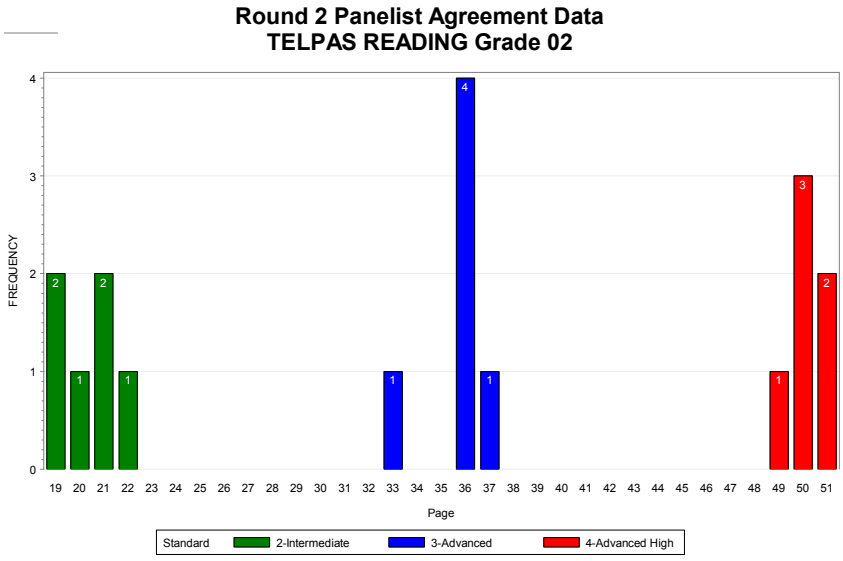
		Round	Minimum	Maximum	Mean	Standard Deviation	Median
Grades 6-7	I	1	31	33	32.4	0.98	33
		2	31	33	32.7	0.76	33
		3	31	33	32.7	0.76	33
	A	1	47	50	48.9	1.46	50
		2	47	50	48.3	0.95	48
		3	48	50	49.1	1.07	50
	H	1	68	71	69.1	0.90	69
		2	68	69	68.9	0.38	69
		3	69	69	69.0	0.00	69
Grades 8-9	I	1	29	32	30.4	1.13	31
		2	30	31	30.4	0.53	30
		3	30	31	30.4	0.53	30
	A	1	46	50	47.7	1.60	47
		2	46	49	47.3	0.95	47
		3	46	49	47.3	0.95	47
	H	1	64	68	65.6	1.27	65
		2	65	68	65.9	1.07	66
		3	65	66	65.7	0.49	66
Grades 10-12	I	1	27	30	28.9	1.07	29
		2	27	29	28.6	0.79	29
		3	27	29	28.6	0.79	29
	A	1	45	51	48.0	2.00	48
		2	48	49	48.1	0.38	48
		3	48	48	48.0	0.00	48
	H	1	69	72	70.7	1.11	71
		2	70	72	70.6	0.79	70
		3	70	72	70.7	0.76	71

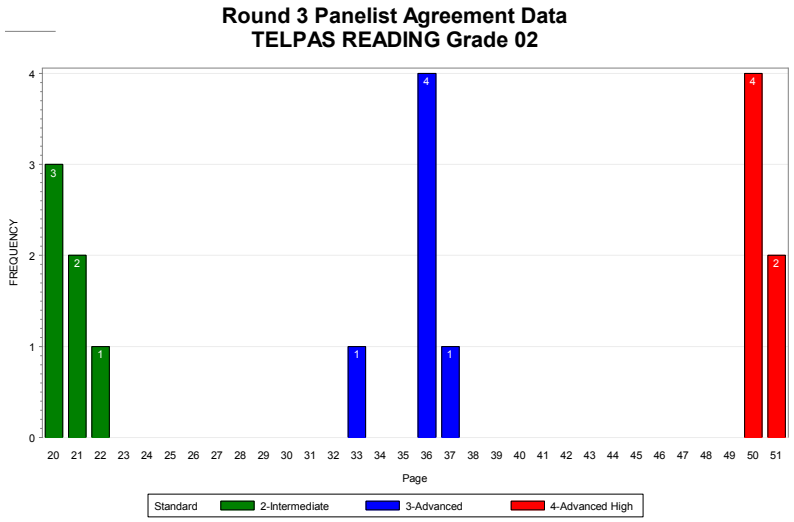
Appendix 12: Standards Review Panelists' Agreement Data

This appendix provides the frequency distribution of the recommended cuts (bookmarked page numbers) after each round of judgments for each TELPAS grade or grade cluster.

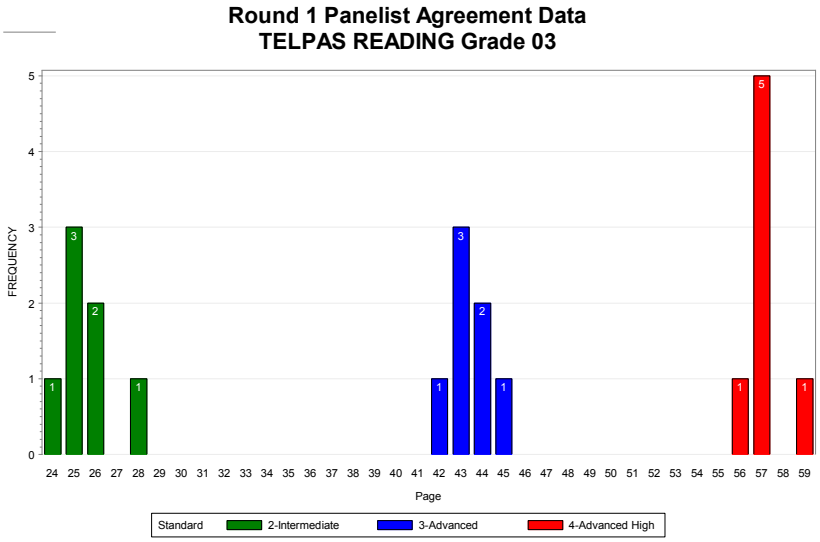
GRADE 2



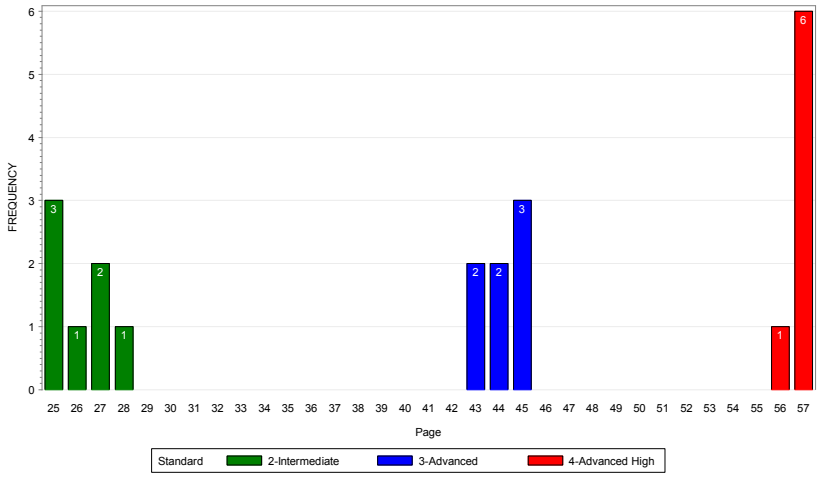



Round 3 Panelist Agreement Data
TELPAS READING Grade 02
FREQUENCY
Page
Standard
2-Intermediate
3-Advanced
4-Advanced High

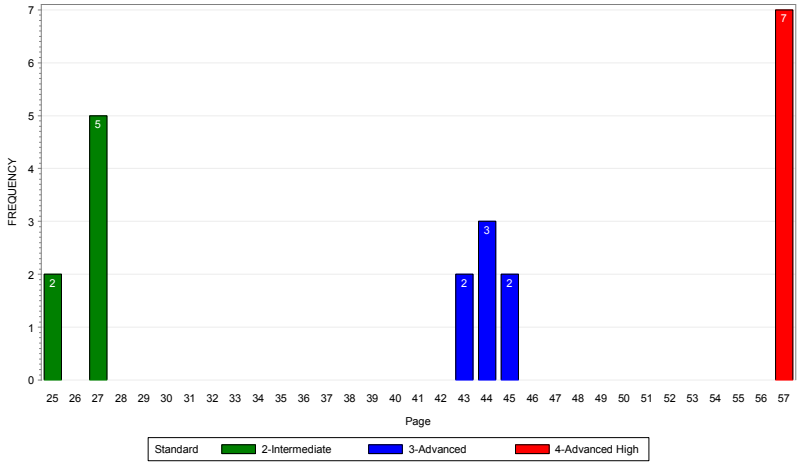
GRADE 3


Round 1 Panelist Agreement Data
TELPAS READING Grade 03
FREQUENCY
Page
Standard
2-Intermediate
3-Advanced
4-Advanced High

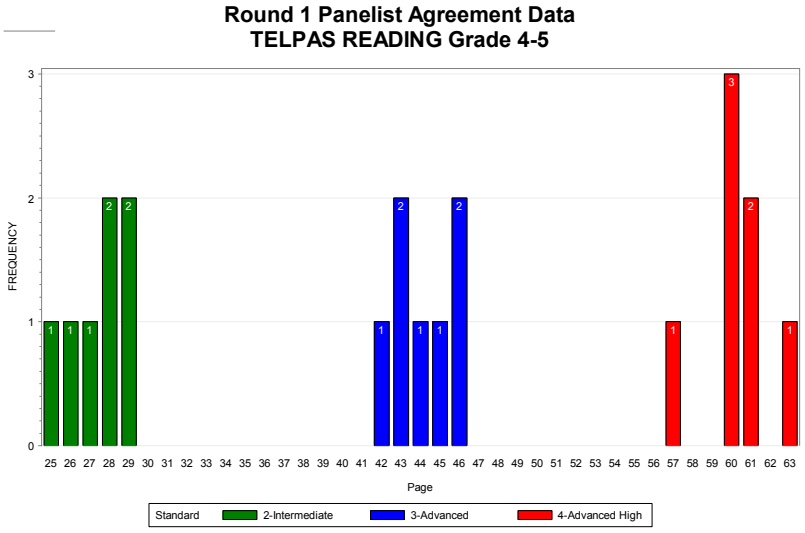
Round 2 Panelist Agreement Data
TELPAS READING Grade 03

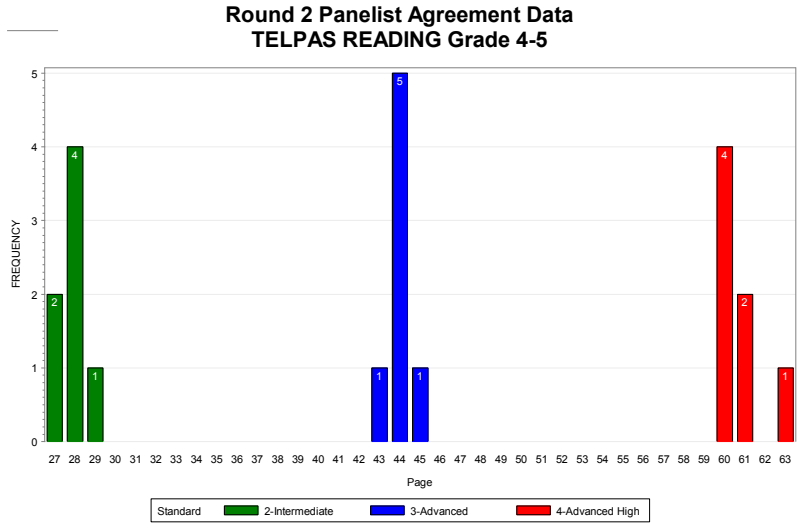


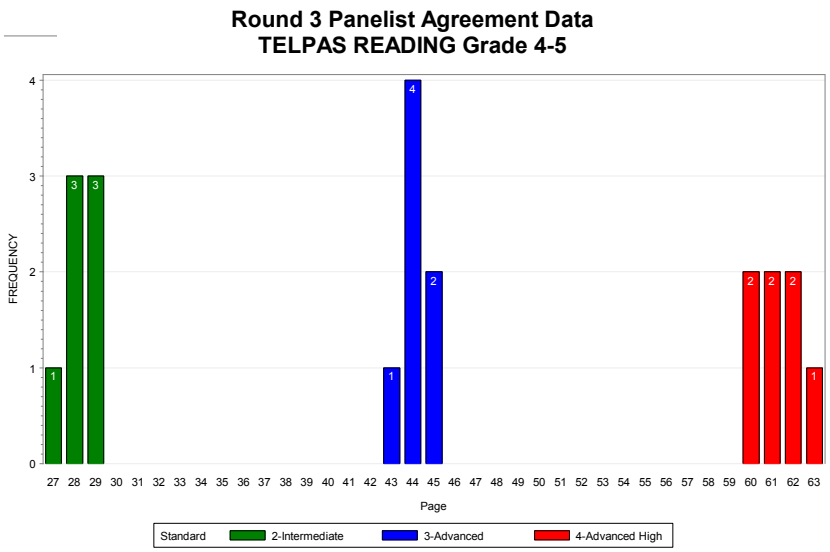
Round 3 Panelist Agreement Data
TELPAS READING Grade 03



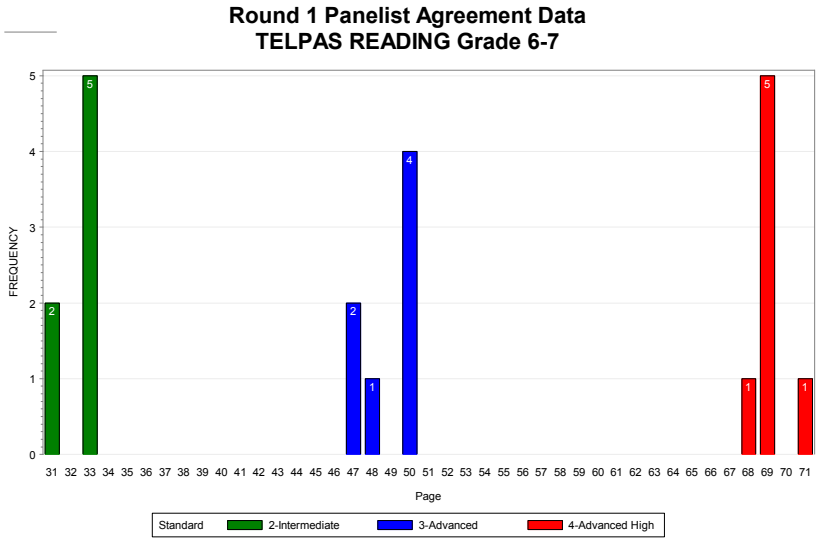
GRADES 4-5


Round 1 Panelist Agreement Data
TELPAS READING Grade 4-5
FREQUENCY
Page
Standard
2-Intermediate
3-Advanced
4-Advanced High

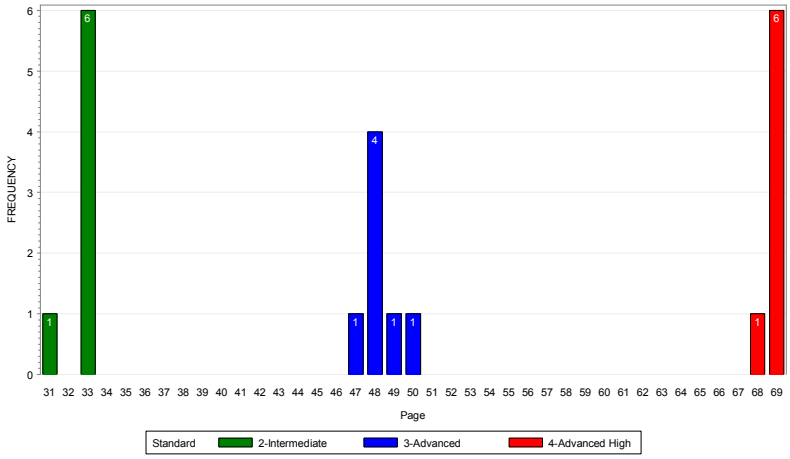

Round 2 Panelist Agreement Data
TELPAS READING Grade 4-5
FREQUENCY
Page
Standard
2-Intermediate
3-Advanced
4-Advanced High


Round 3 Panelist Agreement Data
TELPAS READING Grade 4-5
FREQUENCY
Page
Standard
2-Intermediate
3-Advanced
4-Advanced High

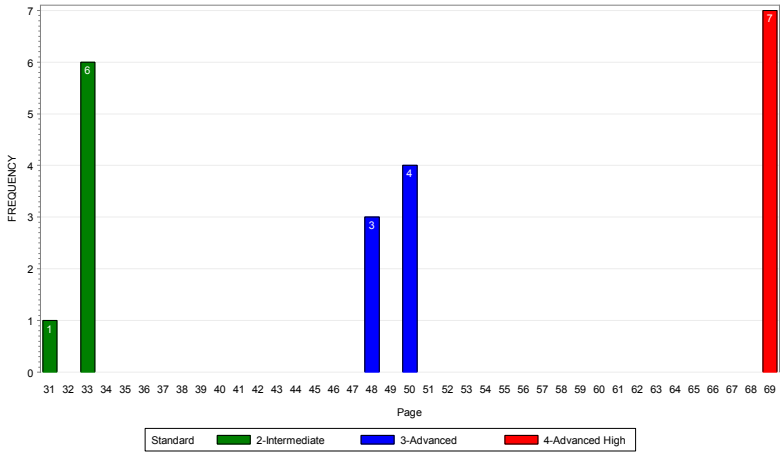
GRADES 6-7


Round 1 Panelist Agreement Data
TELPAS READING Grade 6-7
FREQUENCY
Page
Standard
2-Intermediate
3-Advanced
4-Advanced High

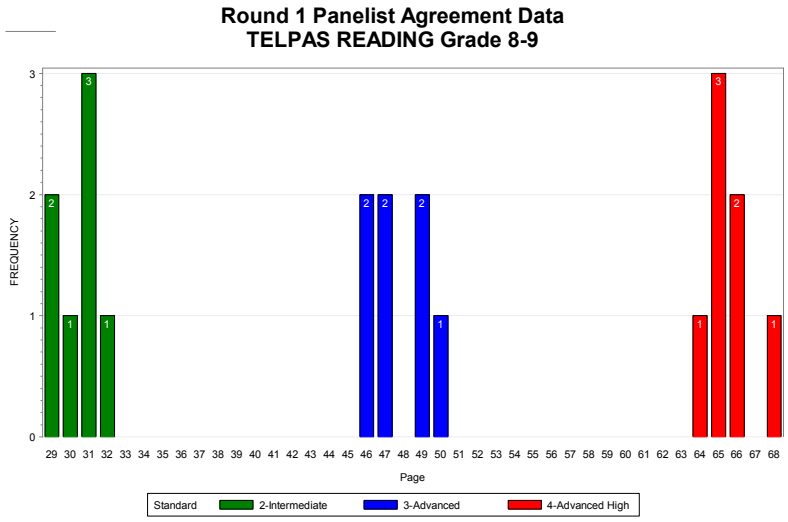
Round 2 Panelist Agreement Data
TELPAS READING Grade 6-7

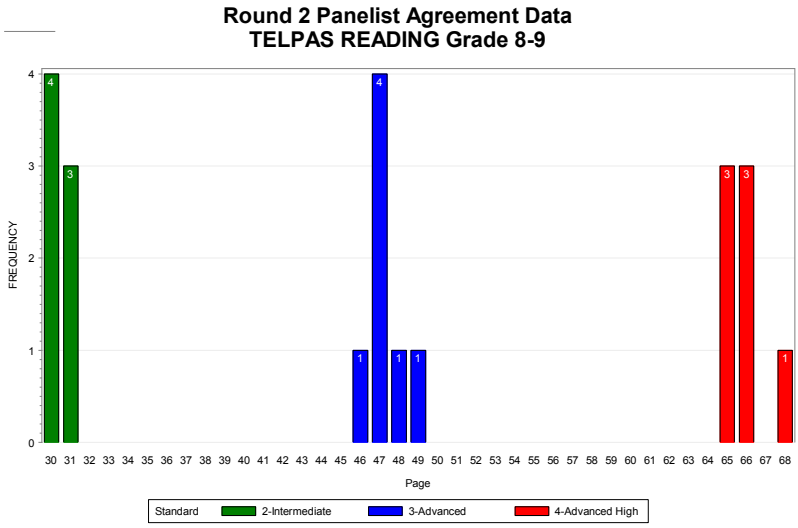


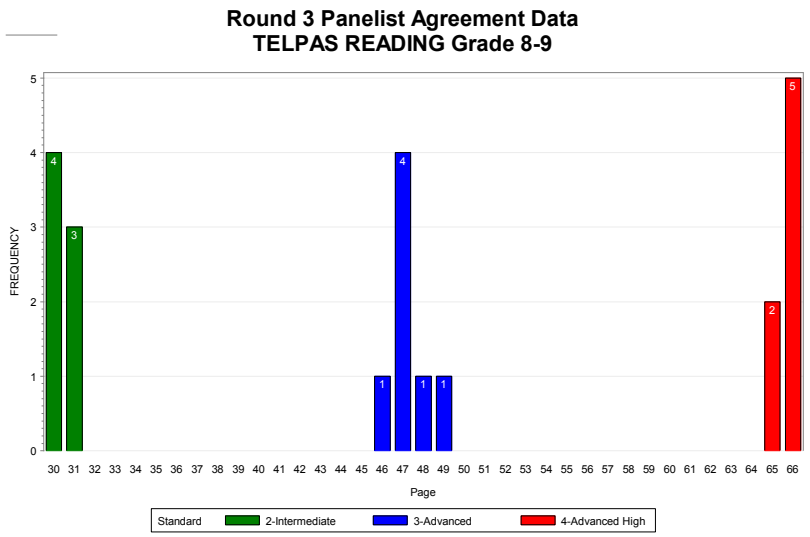
Round 3 Panelist Agreement Data
TELPAS READING Grade 6-7



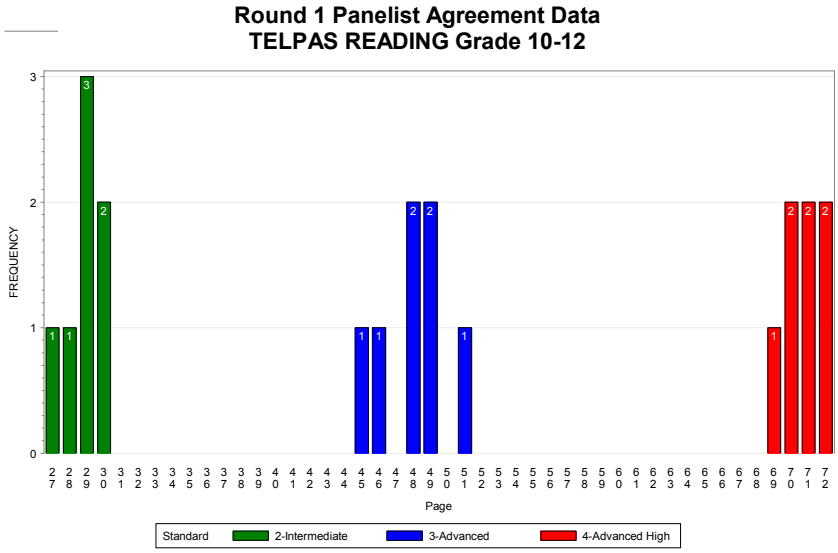
GRADES 8-9


Round 1 Panelist Agreement Data
TELPAS READING Grade 8-9
FREQUENCY
Page
Standard
2-Intermediate
3-Advanced
4-Advanced High

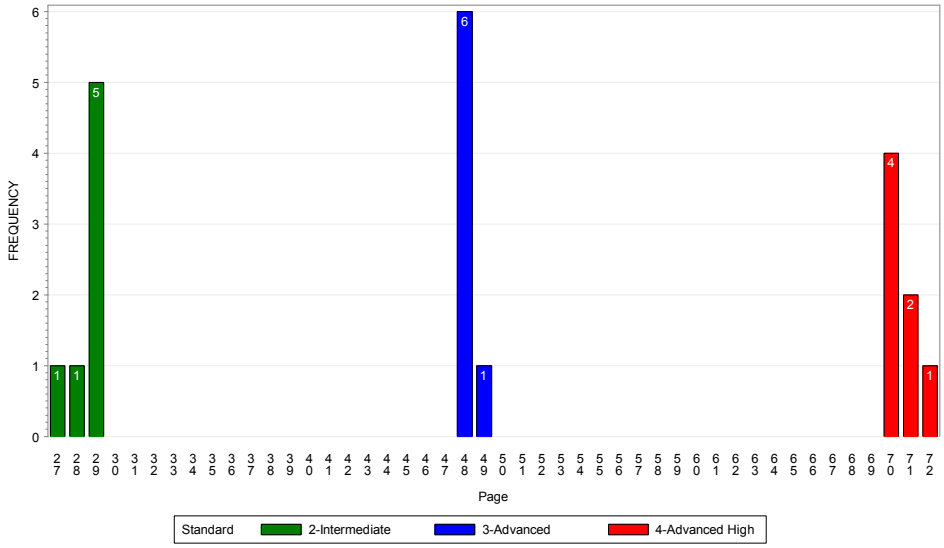

Round 2 Panelist Agreement Data
TELPAS READING Grade 8-9
FREQUENCY
Page
Standard
2-Intermediate
3-Advanced
4-Advanced High


Round 3 Panelist Agreement Data
TELPAS READING Grade 8-9
FREQUENCY
Page
Standard
2-Intermediate
3-Advanced
4-Advanced High

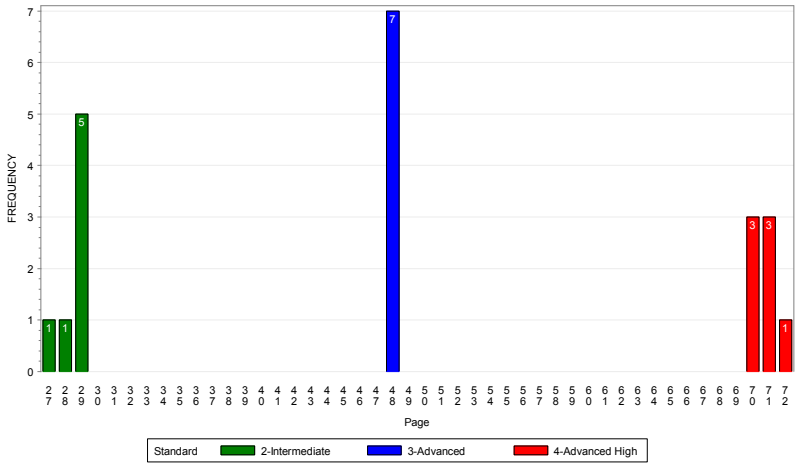
GRADES 10-12


Round 1 Panelist Agreement Data
TELPAS READING Grade 10-12
FREQUENCY
Page
Standard
2-Intermediate
3-Advanced
4-Advanced High

Round 2 Panelist Agreement Data
TELPAS READING Grade 10-12



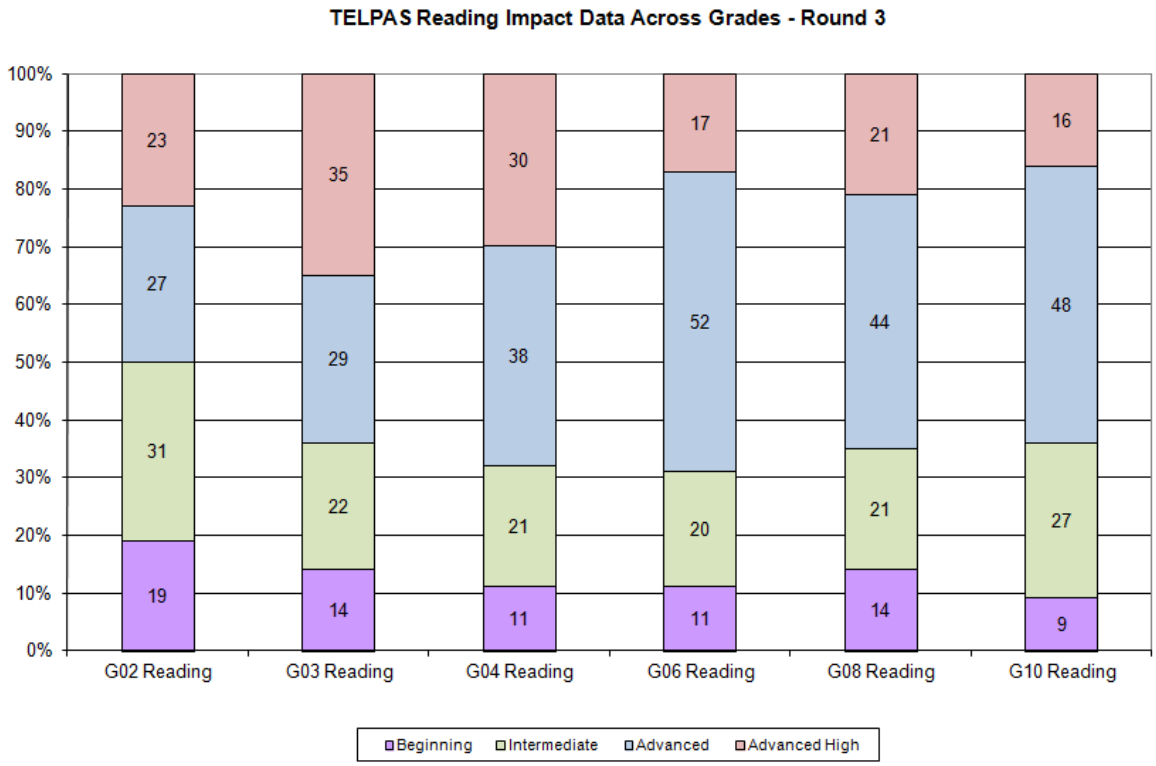
Round 3 Panelist Agreement Data
TELPAS READING Grade 10-12



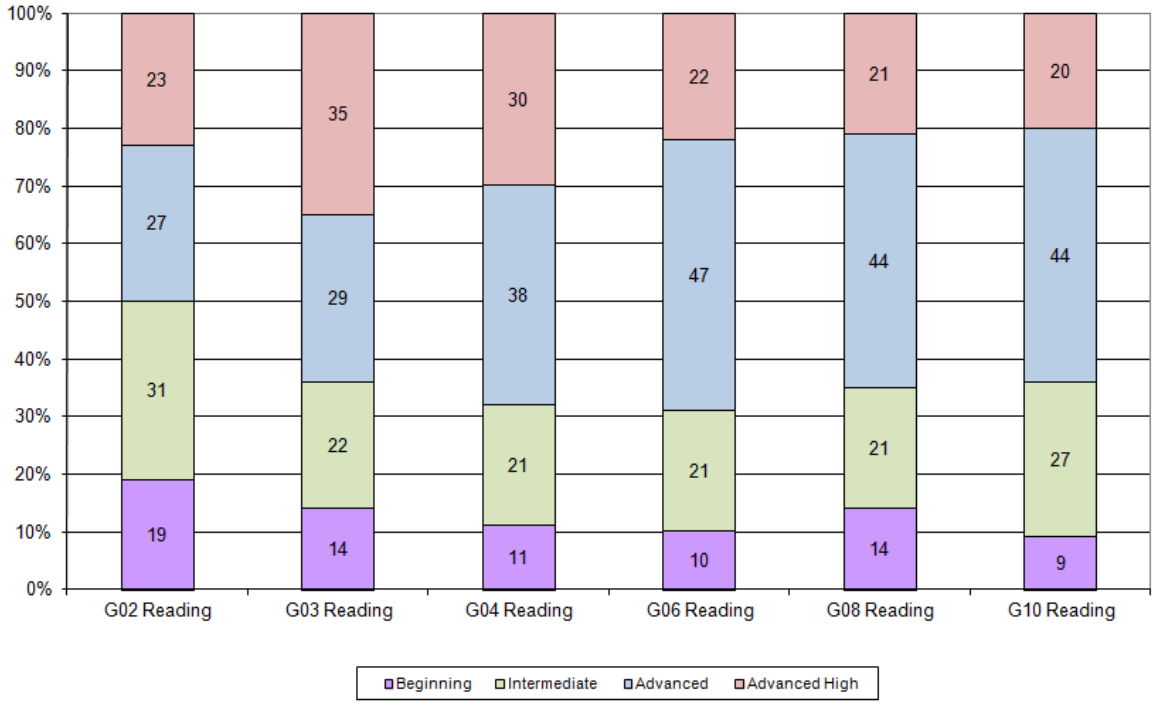
Appendix 13: Estimated Impact Data

This appendix provides the estimated impact data (percentage of students at each proficiency level) based on the cut score recommendations after Round 3 of the standards review committee meetings (Round 3), vertical articulation (Articulation), and the reasonableness review (Reasonableness Review).

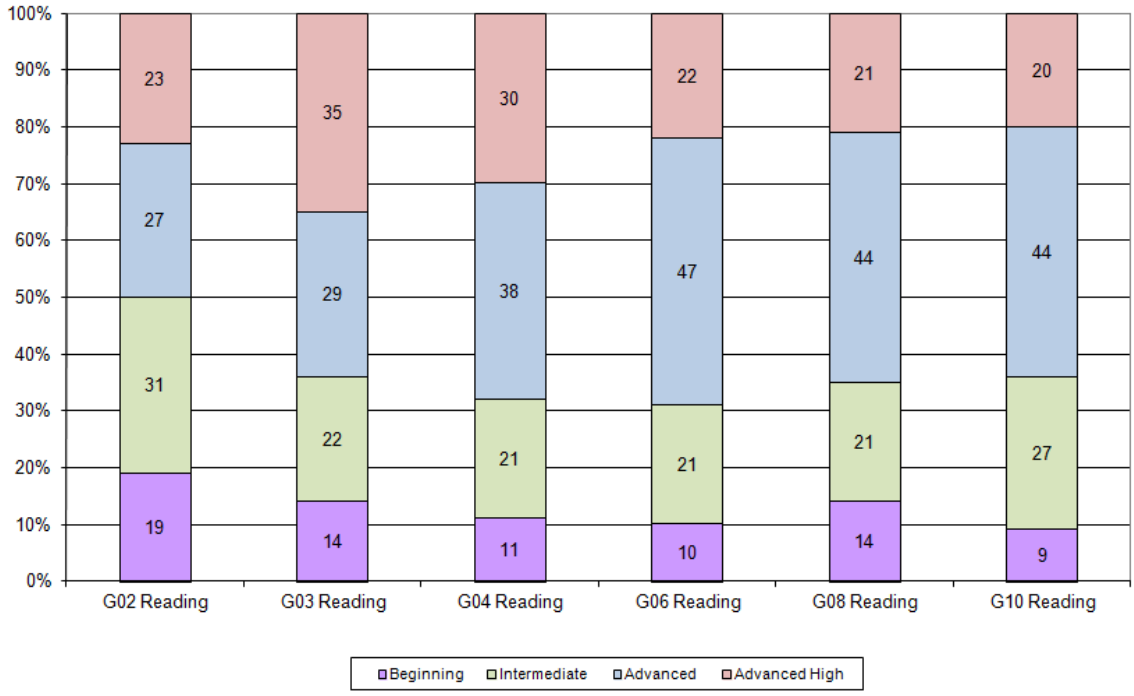
Please note that the impact data are computed based on the performance of ELLs on the TELPAS reading assessments administered in spring 2013.



TELPAS Reading Impact Data Across Grades - Articulation



TELPAS Reading Impact Data Across Grades - Reasonableness Review



References

Beimers, J. N., Way, W. D., McClarty, K. L., & Miles, J. A. (2012). Evidence based standard setting: Establishing cut scores by integrating research evidence with expert content judgments. Pearson Bulletin, January 2012, Issue 21. Retrieved from www.pearsonassessments.com

Cook, G., Linquanti, R., Chinen, M., & Jung, H. (2012). National evaluation of Title III implementation supplemental report—exploring approaches to setting English language proficiency performance criteria and monitoring English learner progress. Retrieved December 1, 2013 from http://www2.ed.gov/rschstat/eval/title-iii/implementation-supplemental-report.pdf

Ferrara, S., Lewis, D., Mercado, R., D'Brot, J., Barth, J., & Egan, K. (2011, April). *A method for setting benchmarked performance standards: Workshop procedures, panelist judgments, and empirical results*. Paper presented at the annual meetings of the National Council on Measurement in Education. New Orleans, LA.

Lewis, D. M., Mitzel, H. C., Green, D. R., & Patz, R. J. (1999). *The bookmark standard setting procedure*. Monterey, CA: McGraw-Hill.

O'Malley, K., Keng, L., & Miles, J. (2012). Using validity evidence to set performance standards. In G. J. Cizek (Ed.), *Setting performance standards* (2nd ed. pp.301–322). New York: Routledge.

Phillips, S. E. (2002). *Setting standards on the TAKS test: A modified item mapping procedure*. BETA, Inc., NCS Pearson, and TEA.

Rasch, G. (1980). *Probabilistic models for intelligence and attainment tests*. Chicago: The University of Chicago Press.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3), 461-481.