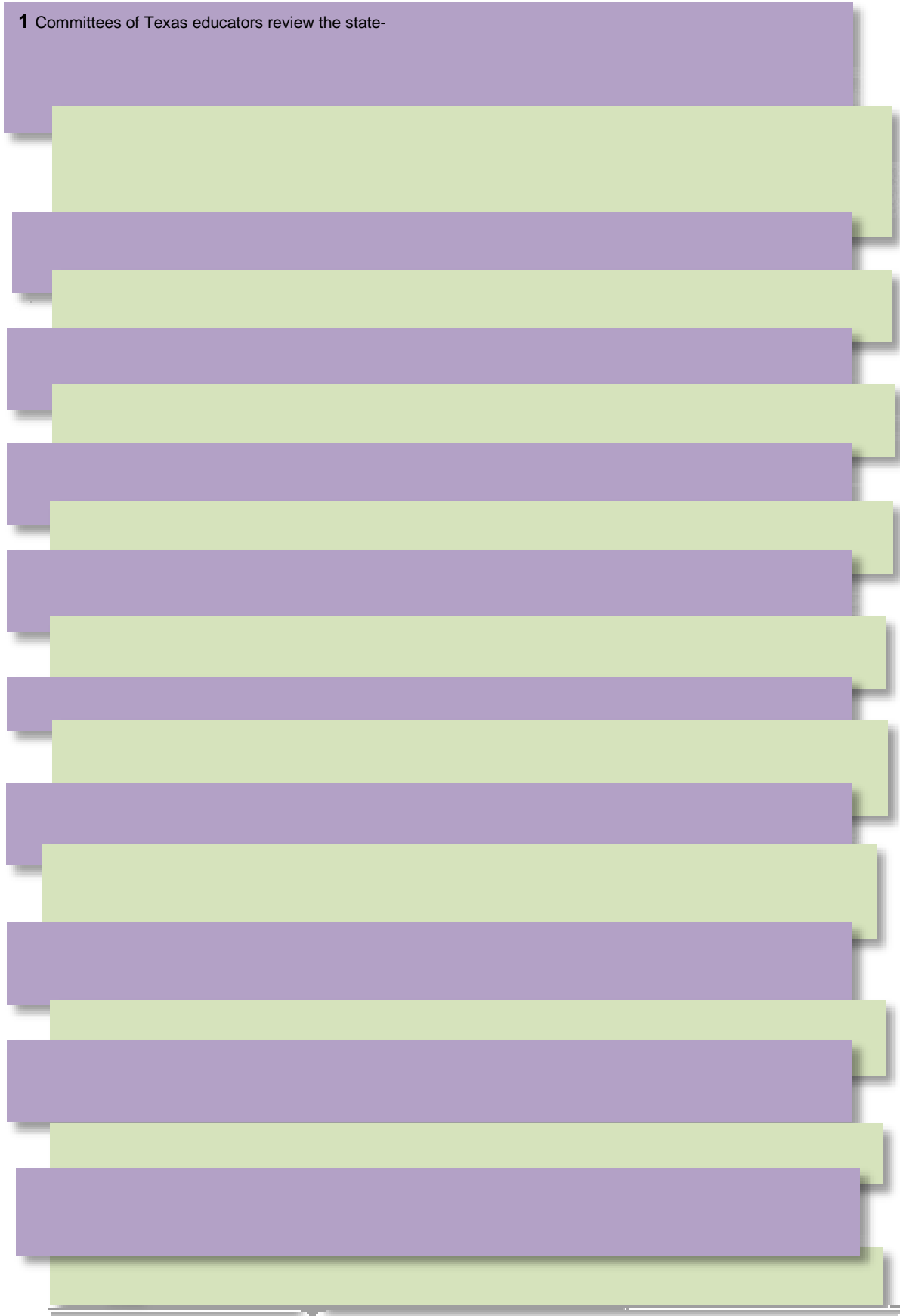




**Figure 2.1. Test-Development Process**







## Item Writers

Pearson and its subcontractors employ item writers who have extensive experience developing items for standardized achievement tests, large-scale criterion-referenced measurements, and English language proficiency tests. These individuals are selected based on their content-area knowledge, their teaching or curriculum development experience in the relevant grades, or their experience teaching students with special needs and English learners.

For each STAAR, STAAR Alternate 2, and TELPAS assessment, TEA receives an item inventory that indicates the number of test items to be developed for each reporting category and TEKS student expectation (for STAAR and STAAR Alternate 2 assessments) or ELPS student expectation (for TELPAS assessments). Item inventories are used throughout the item-review process. If necessary, additional items are developed by the vendor to provide the requisite number of items per student expectation.

For the TELPAS Alternate assessment, the observable behaviors were developed by Texas educators during a series of TEA-led meetings. The educators were guided by Pearson and TEA staff to develop an inventory of items that aligned to the ELPS and covered the alternate Proficiency Level Descriptors (PLDs).

## Training

Pearson provides extensive training for item writers prior to item development. During these trainings, Pearson reviews in detail the content expectations and item guidelines and discusses the scope of the testing program; security issues; adherence to the measurement specifications; and avoidance of possible economic, regional, cultural, gender, or ethnic



**Table 2.1.** Item-Review Guidelines

Passage and Item-Review Guidelines	
Reporting Category/Student Expectation Item Match	<ul style="list-style-type: none"> <li>• The item measures what it is supposed to assess.</li> <li>• The item poses a clearly defined problem or task.</li> </ul>
Appropriateness (Interest Level)	<ul style="list-style-type: none"> <li>• The item or passage is well written and clear.</li> <li>• The point of view is relevant to students taking the test.</li> <li>• vd      cobbpreve</li> </ul>

## Pilot Testing

The purpose of pilot testing is to gather information about test-item prototypes and administration logistics for a new assessment and to refine item-development guidelines as needed. Pilot testing can be conducted to accomplish varying objectives. If the purpose is to pilot items of differing types and ranges of difficulty, piloting might occur before the extensive item-development process described on the preceding pages. If the purpose is to pilot test administration logistics, the pilot might occur after major item development but before field testing. There was no pilot testing in the 2020–2021 school year.

## Field Testing and Data Review

Field testing is conducted prior to a test item being used on an operational test form. However, when there are curriculum changes, newly developed items that have not been field-tested may be used on an operational test form. This is referred to as operational field testing, which was seldom used on the Texas Assessment Program.

## Field-Test Procedures

Whenever possible, TEA conducts field tests of new items by embedding them in multiple forms of operational tests so that the field-test items are randomly distributed to students across the state. This results in a large representative sample of responses gathered on each item. Periodically, TEA conducts standalone field tests of new items (e.g., writing prompts) by administering them to a purposefully selected representative Texas student sample. Experience has shown that embedded field testing yields



To ensure that each item is examined for potential ethnic bias, the sample selection is designed so that the proportions of African American and Hispanic students in the samples are representative of their respective total student populations in Texas. Data obtained from the field test include

- the number of students by ethnicity and gender in each sample;

- the percentage of students choosing each response;

- the percentage of students, by gender and by ethnicity, choosing each response;

- point-biserial correlations to determine the relationship between a correct response on a particular test item and the score obtained on the total content-area test;

- Rasch statistical indices to determine the relative difficulty of each test item; and

- Mantel-Haenszel statistics for dichotomous items and standardized mean difference (SMD) for Constructed Response (CR) items to identify greater-than-expected differences in group performance on any single item by gender and ethnicity.

## **Data-Review Procedures**

After field testing, TEA curriculum and assessment specialists provide feedback to Pearson on each test item and its associated data regarding reporting category and student expectation match; appropriateness; level of difficulty; and potential gender, ethnic, or other bias; and then recommend acceptance or rejection of each field-test item. Items that pass all stages of development—item review, field testing, and data review—are placed in the item bank and become eligible for use on future test forms. Rejected items are marked as such and eliminated from consideration for use on any summative assessment.

## **Item Bank**

ETS and Pearson each maintain an electronic item bank for their respective portion of the assessment program. The item banks store each test item and its accompanying artwork.

Each electronic item bank also stores item data, such as the unique item number (UIN), grade or course, subject, reporting category, TEKS or ELPS student expectation measured, dates the item was administered, and item statistics. Each item bank also warehouses information obtained during data-review meetings, which specifies whether a test item is acceptable for use. TEA, ETS, and Pearson use the item statistics and other information about items during the test-construction process to maintain constant test



other items in the bank. Consequently, items are selected not only to meet sound content and test-construction practices but also to ensure that tests are approximately comparable in difficulty from administration to administration. Refer to [chapter 3, “Standard Technical Processes,”](#) for detailed information about Rasch scaling.

Tests are constructed to meet a blueprint for the required number of items o6 (r)-6 (ed num)-5.9 (be)10.6 o

## Security

TEA places a high priority on test security and confidentiality for all aspects of the statewide assessment program. From the development of test items to the construction of tests, and from the distribution and administration of test materials to the delivery of students' score reports, special care is taken to promote test security and confidentiality. TEA ensures that every allegation of cheating or breach of confidentiality are properly investigated.

Maintaining the security and confidentiality of the Texas Assessment Program is critical for ensuring valid test scores and providing standardized and comparable testing opportunities for all students. TEA has implemented numerous measures to strengthen test security and confidentiality, including the development of various administrative procedures and manuals to train and support district testing personnel.

### Test Administration Manuals

Test security for the Texas Assessment Program has been supported by an aligned set of test administration documents that provide clear and specific information to testing personnel. In response to the statutes and administrative rules that are the foundation for policies and documentation pertaining to test security, TEA produces and updates detailed information about appropriate test administration procedures in the [test administrator manuals](#).

#### MANUALS

The Coordinator Resources and test administrator manuals, including the TELPAS Rater Manual, provide guidelines on how to train testing personnel, administer tests, create secure testing environments, and properly store test materials. They also instruct testing personnel on how to report to TEA any confirmed or alleged testing irregularities that might have occurred in a classroom, on a campus, or within a school district. Finally, the manuals provide training and guidelines relative to test security oaths that all personnel with access to secure test materials are required to sign. The manuals give specific details about the possible penalties for violating test procedures. In addition, TAC §101.3031 includes specific language detailing the requirements of school districts and charter schools to maintain security and confidentiality of assessment instruments, including a list of violations and actions that may result from a violation.

### Online Training

TEA provides training materials that cover test administration best practices and the maintenance of test security. The online training is divided into three modules: 1) active monitoring, 2) distribution of test materials, and 3) proper handling of secure materials. Although completion of these modules is not a requirement, it is, however, strongly recommended that districts and charter schools use them to help supplement the mandatory training required of all personnel involved in testing. Training modules can be accessed from the [training webpage](#) on the [District and Campus Coordinator Resources](#) website.



Erasure information and descriptive statistics for each group (usually by grade level in







## Performance Assessments

The STAAR and TELPAS tests include constructed-response items, which require scoring by trained human raters on the following operational assessments:

STAAR grade 4 and 7 writing

STAAR Spanish grade 4 writing

STAAR English I, English II, and English III

TELPAS grades 2–12 speaking

The Texas Assessment Program uses written compositions on STAAR and STAAR Spanish, which are a direct measure of the student's ability to synthesize the component skills of writing; that is, the composition task requires the student to express ideas effectively in writing for a specified purpose. To do this, the student must be able to respond in a focused and coherent manner to a specific prompt while organizing ideas clearly, generating and developing thoughts in a way that allows the reader to thoroughly understand what the writer is attempting to communicate, and maintaining a consistent control of the conventions of written language.

For the STAAR and STAAR Spanish assessments, the types of writing required vary by grade and course and represent the learning progression evident in the TEKS.

Written compositions for STAAR are evaluated using a holistic scoring process, meaning that the essay is considered as a whole. It is evaluated according to pre-established criteria: organization/progression, development of ideas, and use of language/conventions. These criteria, explained in detail in the writing scoring rubrics for each grade and type of writing, are used to determine the effectiveness of each written response. Each essay is scored on a scale of 1 (a very limited writing performance) to 4 (an accomplished writing performance). A rating of 0 is assigned to



by demonstrating a high level of mastery before any student responses are scored. Pearson's content supervisor



Raters first complete a rubric overview training before training on item specific responses. The Item Specific Training (IST) modules are a set of student compositions that have already been scored by assessment specialists and TEA staff. The training materials are selected to clearly differentiate student performance at the different rubric score points and to help raters learn the difference between score points. The training materials also contain responses determined to be borderline between two adjacent score points to help raters determine the difference between adjacent score points.

123 456 789 1011 1213 1415 1617 1819 2021 2223 2425 2627 2829 3031 3233 3435 3637 3839 4041 4243 4445 4647 4849 5051 5253 5455 5657 5859 6061 6263 6465 6667 6869 7071 7273 7475 7677 7879 8081 8283 8485 8687 8889 9091 9293 9495 9697 9899 100101 102103 104105 106107 108109 110111 112113 114115 116117 118119 120121 122123 124125 126127 128129 130131 132133 134135 136137 138139 140141 142143 144145 146147 148149 150151 152153 154155 156157 158159 160161 162163 164165 166167 168169 170171 172173 174175 176177 178179 180181 182183 184185 186187 188189 190191 192193 194195 196197 198199 200201 202203 204205 206207 208209 210211 212213 214215 216217 218219 220221 222223 224225 226227 228229 230231 232233 234235 236237 238239 240241 242243 244245 246247 248249 250251 252253 254255 256257 258259 260261 262263 264265 266267 268269 270271 272273 274275 276277 278279 280281 282283 284285 286287 288289 290291 292293 294295 296297 298299 300301 302303 304305 306307 308309 310311 312313 314315 316317 318319 320321 322323 324325 326327 328329 330331 332333 334335 336337 338339 340341 342343 344345 346347 348349 350351 352353 354355 356357 358359 360361 362363 364365 366367 368369 370371 372373 374375 376377 378379 380381 382383 384385 386387 388389 390391 392393 394395 396397 398399 400401 402403 404405 406407 408409 410411 412413 414415 416417 418419 420421 422423 424425 426427 428429 430431 432433 434435 436437 438439 440441 442443 444445 446447 448449 450451 452453 454455 456457 458459 460461 462463 464465 466467 468469 470471 472473 474475 476477 478479 480481 482483 484485 486487 488489 490491 492493 494495 496497 498499 500501 502503 504505 506507 508509 510511 512513 514515 516517 518519 520521 522523 524525 526527 528529 530531 532533 534535 536537 538539 540541 542543 544545 546547 548549 550551 552553 554555 556557 558559 560561 562563 564565 566567 568569 570571 572573 574575 576577 578579 580581 582583 584585 586587 588589 590591 592593 594595 596597 598599 600601 602603 604605 606607 608609 610611 612613 614615 616617 618619 620621 622623 624625 626627 628629 630631 632633 634635 636637 638639 640641 642643 644645 646647 648649 650651 652653 654655 656657 658659 660661 662663 664665 666667 668669 670671 672673 674675 676677 678679 680681 682683 684685 686687 688689 690691 692693 694695 696697 698699 700701 702703 704705 706707 708709 710711 712713 714715 716717 718719 720721 722723 724725 726727 728729 730731 732733 734735 736737 738739 740741 742743 744745 746747 748749 750751 752753 754755 756757 758759 760761 762763 764765 766767 768769 770771 772773 774775 776777 778779 780781 782783 784785 786787 788789 790791 792793 794795 796797 798799 800801 802803 804805 806807 808809 810811 812813 814815 816817 818819 820821 822823 824825 826827 828829 830831 832833 834835 836837 838839 840841 842843 844845 846847 848849 850851 852853 854855 856857 858859 860861 862863 864865 866867 868869 870871 872873 874875 876877 878879 880881 882883 884885 886887 888889 890891 892893 894895 896897 898899 900901 902903 904905 906907 908909 910911 912913 914915 916917 918919 920921 922923 924925 926927 928929 930931 932933 934935 936937 938939 940941 942943 944945 946947 948949 950951 952953 954955 956957 958959 960961 962963 964965 966967 968969 970971 972973 974975 976977 978979 980981 982983 984985 986987 988989 990991 992993 994995 996997 998999 1000

**N**

the remaining pre-scored responses from the rangefinding sessions to training sets and qualifying sets for use in future rater training. Educators assist in the review and make recommendations to reach a consensus on the scores. Prior to scoring, TEA staff review and approve all scoring guides and training sets.

## **Score Reliability and Validity Information**

Throughout the years, TEA has reported on the reliability and validity of the performance-scoring process. Reliability has been expressed in terms of rater agreement (percentage of exact agreement between rater scores) and correlation between first and second ratings. Validity has been assessed by the inclusion of validity responses throughout the operational-scoring process. It is expressed in terms of exact agreement between the score assigned by the rater and the “true” score assigned by ETS and approved by TEA.

## **Appeals**

If a district has questions about the score assigned to a response, a rescore can be requested through submission of the appropriate request form. ETS provides rescore results by posting an updated STAAR Report Card (SRC) to the STAAR Assessment Management System, only if the score has changed. If the score does not change, there is a fee that districts pay. If the score changes, that fee is waived. If a district files a formal appeal with TEA related to scores reported on the consolidated accountability file, an analysis of the response in question that explains the final outcome of the appeal, and whether or not the score is correct.