Texas
Educator Certification
Examination Program

TEXAS EDUCATOR CERTIFICATION EXAMINATION PROGRAM

Annual Technical Development Manual
2020-2021

March 2023

Prepared for:
Texas Education Agency
Educator Certification Division
William B. Travis Building
1701 N. Congress Avenue
Austin, Texas, 78701

Developed and Produced by

Pearson

Evaluation Systems, Pearson, 300 Venture Way, Hadley, MA 01035

(intentionally blank)

Table of Contents

Preface

This technical manual provides information on examination development activities undertaken for the Texas Educator Certification Examination Program during the 2020-2021 program year. Specifically, the manual provides validity evidence to support the use of new and redeveloped examinations that became operational between September 1, 2020, and August 31, 2021. This manual also provides evidence to support the validity of score interpretations for these examinations.

This manual is intended for policy makers, state educators, and other interested stakeholders who would like to learn more about the:

- purpose, structure, and composition of the Texas Educator Certification Examination Program generally;
- exam item development and validation processes generally, and those related to the examinations in particular that became operational in 2020-21;
- content and bias review processes related to the examinations that became operational in 2020-21; and
- establishment of Texas educator passing standards for certification related to examinations that became operational in 2020-21.

In 2017, The Texas Education Agency (TEA) awarded the Evaluation Systems group of Pearson (Pearson) the contract to support the Texas Educator Certification Examination Program, beginning on September 1, 2018. As part of the transition process from the previous vendor, all the existing examinations and preparation materials at that time were delivered to Pearson for use as-is. Specific details about the development of these materials by the previous vendor can be found on the TEA website at https://tea.texas.gov/sites/default/files/texas_technical_manual_8.31.18.pdf. As described in that manual, those examinations were developed in accordance with the practices recommended by the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) (*Standards*). The *Standards* require a clear definition of content domain and a rationale to support a claim that the knowledge, skills, and abilities being assessed in a licensure test are required for credential-worthy performance. The *Standards* also require that the assessments are fair, valid, and reliable with administration conditions that are equitable for all examinees.

The primary purpose of this technical manual is to document the evidence and exam development activity that Pearson facilitated in support of launching each of the following four* new and redeveloped examinations in the 2020-21 program year:

- Science of Teaching Reading (293)
- Early Childhood: PK-3 (292)

- Educational Diagnostician (253)
- Core Subjects: Early Childhood-Grade 6 (391)*
    - Core Subjects: EC-6 English Language Arts and Reading (901)
    - Core Subjects: EC-6 Mathematics (902)
    - Core Subjects: EC-6 Social Studies (903)
    - Core Subjects: EC-6 Science (904)
    - Core Subjects: EC-6 Fine Arts, Health, and Physical Education (905)

    *Exam 391 is composed of 5 subtests

As such, the exam development activities referred to in this technical manual reflect **all** tasks of the exam development process that were followed in development of these four examinations. Note the steps to operationalize Core Subjects: Early Childhood-Grade 6 (391) were streamlined to address a specific goal, and the circumstances and steps are further described in this manual.

In addition, in 2020-21 Pearson conducted exam development activities in support of the following educator certification exams that was ongoing and did not result in an exam launch during 2020-2021:

- School Counselor (252)
- Pedagogy and Professional Responsibilities for Trade & Industrial Education (370)
- English Language Arts and Reading 4-8 (217)
- Physical Education EC-12 (258)
- English Language Arts and Reading 7-12 (331)
- Health EC-12 (257)
- Deafblind EC-12 (185)
- Special Education EC-6 (186)
- Special Education 6-12 (187)

Those activities that Pearson conducted specifically during the 2020-21 program year for these educator certification examinations (which were not operational) are also identified throughout the manual.

Purpose of the Texas Educator Certification Examination Program

Introduction

The purpose of the Texas Educator Certification Examination Program is to verify that each educator has the prerequisite content and professional knowledge, skills, and abilities necessary for an entry-level position in Texas public schools and has the capability of becoming a highly effective educator. In the 2020-21 academic year, the Texas Educator Certification Examination Program, administered by Pearson, included five categories of examinations: the Texas Examinations of Educator Standards (TExES), the Texas Assessment of Sign Communications (TASC), the Texas Assessment of Sign Communications—American Sign Language (TASC-ASL), the American Association of Family and Consumer Sciences (AAFCS) exams, and the Pre-Admission Content Tests (PACT).

Program Policy

Texas law requires that educators pass appropriate examinations to become certified. The examination(s) required for certification are specified in Texas Administrative Code, §230.21(e): http://ritter.tea.state.tx.us/sbecrules/tac/chapter230/ch230c.html.

The examinations are aligned to standards for beginning educators. These standards are created and approved by the State Board for Educator Certification (SBEC)[1]: https://tea.texas.gov/about-tea/leadership/state-board-for-educator-certification. The standards are focused upon the Texas Essential Knowledge and Skills (TEKS), the required statewide public-school curriculum. They reflect current research on the developmental stages and needs of children from Early Childhood (EC) through Grade 12. Additionally, the Texas Commissioner of Education has adopted rules pertaining to Texas educator standards in Texas Administrative Code Chapter 149, accessible through the TEA website: https://tea.texas.gov/texas-educators/preparation-and-continuing-education/approved-educator-standards.

Texas law also allows for the administration of an appropriate content certification exam to an applicant seeking admission to an EPP, if that applicant does not meet a minimum GPA requirement, as specified in Texas Administrative Code §227.10: https://tea.texas.gov/sites/default/files/ch227a-New-Fig-Link.pdf

[1] SBEC was created by the Texas Legislature in 1995 to recognize public school educators as professionals and grant educators the authority to govern the standards of their profession. The Board oversees all aspects of the preparation, certification and standards of conduct of public-school educators. SBEC's mission statement is "SBEC is dedicated to improving student achievement and ensuring the safety and welfare of Texas school children by upholding the highest level of educator preparation, performance, continuing education, and standards of conduct."

Composition of the Texas Educator Certification Examination Program

As of August 31, 2021, the Texas Educator Certification Examination Program comprised 75 examinations across the categories of TExES, TASC, TASC-ASL, and AAFCS, as well as 38 PACT examinations. The available examinations are listed by category and examination code in Table 1.

The Texas Educator Certification Examination Program examinations are computer-administered and delivered through a national network of Pearson testing centers. Pearson offers several web-based resources to help candidates prepare for the examinations. These resources include online preparation manuals, interactive practice exams, detailed score reports, and computer-administered testing tutorials. In addition, a suite of web-based faculty resources and interactive worksheets are available to assist in candidate preparation.

The TEA and Educator Preparation Programs (EPPs) have access to an interactive, electronic exam results database called *ResultsAnalyzer*® that allows them to create customized reports. Using *ResultsAnalyzer*®, faculty can generate tailored reports by exam and by candidate and filter these results through many factors including time period, candidate demographics, and Title II eligibility. Some analyses are available as charts and graphs. Data from *ResultsAnalyzer*® can be printed or downloaded into an Excel spreadsheet for further analyses and merging with other data sets. Candidate individual score reports (ISRs) are also available to EPPs through *ResultsAnalyzer*®.

From August to November 2018, Pearson conducted a series of webinars to train EPP representatives on the functions and features of *ResultsAnalyzer*®. In addition, individual one-on-one sessions with a *ResultsAnalyzer*® specialist were provided in-person to EPP representatives at the Consortium of State Organizations for Texas Teacher Education (CSOTTE) 2018 annual conference. Additionally, between November 2019 and August 2021, Pearson delivered multiple webinar demonstrations of enhanced features and functions to support EPPs in their data analysis.

Table 1. Texas Educator Certification Examination Program Composition

TExES™	Supplemental
113 English Language Arts and Reading/Social Studies 4–8	157 Health EC–12
114 Mathematics/Science 4–8	158 Physical Education EC–12
115 Mathematics 4–8	160 Pedagogy and Professional Responsibilities EC–12
116 Science 4–8	161 Special Education EC–12
117 English Language Arts and Reading 4–8	162 Gifted and Talented Supplemental
118 Social Studies 4–8	163 Special Education Supplemental
129 Speech 7–12	164 Bilingual Education Supplemental
150 School Librarian	171 Technology Education 6–12
151 Reading Specialist	177 Music EC–12
152 School Counselor	178 Art EC–12
153 Educational Diagnostician	180 Theatre EC–12
154 English as a Second Language	181 Deaf and Hard-of-Hearing

182 Visually Impaired
184 American Sign Language (ASL)
190 Bilingual Target Language Proficiency Test (BTLPT) Spanish
195 Superintendent
231 English Language Arts and Reading 7–12
232 Social Studies 7–12
233 History 7–12
235 Mathematics 7–12
236 Science 7–12
237 Physical Science 6–12
238 Life Science 7–12
240 Chemistry 7–12
241 Computer Science 8–12
242 Technology Applications EC–12
243 Physics/Mathematics 7–12
253 Educational Diagnostician
256 Journalism 7–12
268 Principal as Instructional Leader (PAIL)
270 Pedagogy and Professional Responsibilities for Trade and Industrial Education 6–12
272 Agriculture, Food and Natural Resources 6–12
273 Health Science 6–12
274 Mathematics/Physical Science/Engineering 6–12
275 Marketing 6–12
276 Business and Finance 6–12
279 Dance 6–12
283 Braille (UEB)
292 Early Childhood: PK-3
293 Science of Teaching Reading
610 Languages Other Than English (LOTE) French
611 Languages Other Than English (LOTE) German
612 Languages Other Than English (LOTE) Latin
613 Languages Other Than English (LOTE) Spanish
801 Core Subjects EC–6 English Language Arts and Reading
802 Core Subjects EC–6 Mathematics
803 Core Subjects EC–6 Social Studies
804 Core Subjects EC–6 Science
805 Core Subjects EC–6 Fine Arts, Health and Physical Education
806 Core Subjects 4–8 English Language Arts & Reading
807 Core Subjects 4–8 Mathematics
808 Core Subjects 4–8 Social Studies
809 Core Subjects 4–8 Science
901 Core Subjects EC–6 English Language Arts and Reading
902 Core Subjects EC–6 Mathematics
903 Core Subjects EC–6 Social Studies
904 Core Subjects EC–6 Science
905 Core Subjects EC–6 Fine Arts, Health and Physical Education

TASC™
072 Texas Assessment of Sign Communication (TASC)

TASC–ASL™
073 Texas Assessment of Sign Communication—American Sign Language (TASC–ASL)

PACT
700 TX PACT: Essential Academic Skills (Subtest I: Reading)
701 TX PACT: Essential Academic Skills (Subtest II: Writing)
702 TX PACT: Essential Academic Skills (Subtest III: Mathematics)
710 TX PACT: LOTE French: Early Childhood–Grade 12
711 TX PACT: LOTE German: Early Childhood–Grade 12
712 TX PACT: LOTE Latin: Early Childhood–Grade 12
713 TX PACT: LOTE Spanish: Early Childhood–Grade 12
714 TX PACT: LOTE Chinese: Early Childhood–Grade 12
715 TX PACT: Mathematics: Grades 4–8
716 TX PACT: Science: Grades 4–8
717 TX PACT: English Language Arts and Reading: Grades 4–8
718 TX PACT: Social Studies: Grades 4–8
721 TX PACT: Family and Consumer Sciences
729 TX PACT: Speech: Grades 7–12
731 TX PACT: English Language Arts and Reading: Grades 7–12
732 TX PACT: Social Studies: Grades 7–12
733 TX PACT: History: Grades 7–12
735 TX PACT: Mathematics: Grades 7–12
736 TX PACT: Science: Grades 7–12
737 TX PACT: Physical Science: Grades 6–12
738 TX PACT: Life Science: Grades 7–12
739 TX PACT: Physics: Grades 7–12
740 TX PACT: Chemistry: Grades 7–12
741 TX PACT: Computer Science: Grades 8–12
742 TX PACT: Technology Applications: Early Childhood–Grade 12
756 TX PACT: Journalism: Grades 7–12
757 TX PACT: Health: Early Childhood–Grade 12
758 TX PACT: Physical Education: Early Childhood–Grade 12
771 TX PACT: Technology Education: Grades 6–12
772 TX PACT: Agriculture, Food, and Natural Resources: Grades 6–12
776 TX PACT: Business and Finance: Grades 6–12
777 TX PACT: Music: Early Childhood–Grade 12
778 TX PACT: Art: Early Childhood–Grade 12
779 TX PACT: Dance: Grades 6–12

780 TX PACT: Theatre: Early Childhood–Grade 12
784 TX PACT: American Sign Language (ASL): Early Childhood–Grade 12 (Subtest I)
785 TX PACT: American Sign Language (ASL): Early Childhood–Grade 12 (Subtest II)
790 TX PACT: Core Subjects: Grades 4–8

AAFCS
200 AAFCS Family and Consumer Sciences, Composite
201 AAFCS Hospitality, Nutrition, and Food Science
202 AAFCS Human Development and Family Studies

Validity Evidence to Support the Texas Educator Certification Examination Program

Validity is the degree to which evidence and theory support the interpretation of test scores for proposed uses of tests. Validation techniques traditionally used to support the use of tests for licensure and certification are described in the *Standards* (AERA, APA, & NCME, 2014). The *Standards* provides professional guidelines for accumulating validity evidence. The guidelines are clear that the process for accumulating such validity evidence must be comprehensive and draw from every aspect of exam development.

For the Texas Educator Certification Examination Program, the primary validity focus is content validity. Gathering content-related validity evidence includes a comprehensive process of reviewing assessment content for alignment with state requirements for licensure, reviewing content to verify it is equitable and free from bias, validating competencies and items, and establishing an appropriate passing standard. Pearson works with the TEA, Texas educators, and educator preparation faculty to implement such a process for the development of new exams in the Texas Educator Certification Examination Program, collecting key validity evidence to support the use of the assessments for the purpose of educator licensure.

The process was designed to establish and/or support the connection between an exam and its educational purpose (i.e., educator licensure). This connection provides evidence supporting the validity of score interpretations, which is the central concern in high-stakes professional testing programs. Comprehensive validity evidence strengthens the credibility of a licensure test for state use.

Bias Prevention and Fairness in Exam Development

To create sensitive, fair, and valid examinations for test takers, Pearson makes bias prevention and equity a priority during the development and review of examination materials. Exam developers and editors are charged with detecting and removing potentially biased content, situations, language, and stereotypes throughout the exam design process. The composition of educator review committees reflects, to the extent possible, various institutions and educational philosophies and the diversity of the Texas population demographically, geographically, and by professional expertise. Sampling of participants for content validation surveys and standard setting activities also takes into consideration these demographic variables.

Additionally, Pearson exam development staff employ statistical analyses designed to detect instances where one group of candidates performs significantly better on an item than another group of equivalent ability (differential item functioning). Based on these results, any issues with the items can be addressed before the exams become operational.

Pearson's *Fairness and Diversity in Tests* (2009) manual guides these efforts. This manual was developed by psychometricians and exam development experts and

is used by educators across the country for exam development purposes. In four major sections, it provides an in-depth discussion of the dimensions of bias in test development, addresses specific bias prevention steps and methods of bias review to be taken in exam development, and includes a comprehensive understanding of equity inclusion (i.e., the inclusion of content that reflects diverse populations).

The sources of bias discussed in the manual include:

- bias due to content;
- bias in language;
- bias due to assumptions and stereotypes; and
- bias due to lack of inclusion of exam content that reflects diversity of the population for whom the exam is intended.

While bias prevention is an integral part of Pearson's test development activities and a component of each Content Advisory Committee's (CAC's) responsibility, Pearson establishes a separate and independent Equity Assurance Panel (EAP), composed of Texas educators, to specifically focus on reviewing examination materials for potential bias concerns, an exam development step recommended by the *Standards* (AERA et al., 2014).

Examination Development Process

What follows is a description of the general process Pearson follows in custom exam development, which was applied in whole to all certification exams that became operational in the 2020-21 program year (with variation for Core Subjects: EC-6 (391) as noted below). Also included are specific references to the activity associated with exam development that was ongoing in 2020-21 for certification exams that did not become operational by August 31, 2021 (with exception for Science of Teaching Reading (293) which was launched in January 2021).

Examination development and validation processes consist of defining exam materials and linking them to the most appropriate measurement tools for assessing the content. Pearson develops examinations in accordance with the guidelines specified in the *Standards* (AERA et al., 2014) for defining examination materials, developing examination questions, establishing passing standards, and collecting evidence to support the validity of the examinations. The following activities are conducted to collect validity evidence to support the use of an examination.

Task 1: Conduct Program Planning

Pearson met with representatives of the TEA in Texas and by conference calls, as needed, to discuss and formulate a plan for program development activities. Topics discussed included TEA's vision for improvements to the Texas Educator Certification Examination Program, identification of Texas and national standards

to inform development, proposed conference dates, the recruitment of Texas educators to participate in assessment development activities, and the involvement of Texas stakeholder groups. Development and maintenance topics are reviewed weekly and on an as-needed basis.

Generally, exam development activities have been split into several phases to meet the priorities determined by TEA.

Task 2: Establish Texas Advisory Committees

As part of the standard process, Texas educators and educator preparation program faculty are called upon to review and validate test materials for use in the Texas Educator Certification Examination Program. Pearson worked with the TEA to establish Equity Assurance Panels (EAP) and separate Content Advisory Committees (CACs) to review materials for each field as they are developed. Selection of committee members for each panel was guided by the *Standards* which state:

> *When appropriate to documenting the validity of test score interpretations for intended uses, relevant experts external to the testing program should review the test specifications to evaluate their appropriateness for intended uses of the test scores and fairness for intended test takers. The purpose of the review, the process by which the review is conducted, and the results of the review should be documented. The qualifications, relevant experiences, and demographic characteristics of expert judges should also be documented. (AERA, APA, & NCME, 2014, p. 87).*

Potential committee members include certified Texas educators and educator preparation program faculty as recommended by educator stakeholder groups (e.g., professional organizations, preparation program deans, school superintendents and principals). Recommended individuals are invited to complete a committee application form. In assembling each review group, Pearson worked with the TEA to provide representation in terms of ethnicity, gender, geographic region of the state, and school setting (e.g., urban, suburban, and rural areas). The TEA approved all committee participants.

Prior to beginning development, Pearson initiated the process for assembling an EAP and CAC to review draft exam materials for all the following fields in which exam development activity occurred within the 2020-21 year:

- School Counselor (252)
- Pedagogy and Professional Responsibilities for Trade & Industrial Education 6-12 (370)
- English Language Arts and Reading 4-8 (217)
- Health EC-12 (257)
- Physical Education EC-12 (258)
- English Language Arts and Reading 7-12 (331)
- Deafblind EC-12 (185)
- Special Education EC-6 (186)
- Special Education 6-12 (187)

Task 3: Align with Texas Standards

The process of gathering validity evidence to support the use of the Texas Educator Certification Examination Program examinations continues with the alignment of the examination frameworks to Texas and national standards. The TEA and Pearson verified that the content of examinations under development is appropriate as defined by these standards (see *Approved Educator Standards: https://tea.texas.gov/Texas_Educators/Preparation_and_Continuing_Education/Approved_Educator_Standards*).

In 2020-21, Pearson began the process of reviewing state and national standards in preparation for framework development in each of the following fields:

- Deafblind EC-12 (185)
- Special Education EC-6 (186)
- Special Education 6-12 (187)
- English Language Arts and Reading 7-12 (331)

Framework development for the following fields was completed prior to 2020-21:

- School Counselor (252)
- Pedagogy and Professional Responsibilities for Trade & Industrial Education 6-12 (370)
- Physical Education EC-12 (258)
- English Language Arts and Reading 4-8 (217)

Task 4: Develop and Review Examination Frameworks

Standard 11.13 of the *Standards* requires that evidence should be provided to show that the knowledge, skills, and abilities that the examination intends to assess are required for credential-worthy performance in the occupation and are consistent with the purpose of the licensure program (AERA, APA, & NCME, 2014, p. 178). For each developed examination in the Texas Educator Certification Examination Program, a Content Advisory Committee (CAC) reviews examination materials for accuracy and validates materials to include only content that is pertinent to the field and important for use in a licensing instrument. The role of each CAC is to review examination materials for content accuracy and appropriateness. The CACs provide content-related validity evidence to support the use of the examinations. CACs review frameworks and test questions and participate in standard setting activities. This section describes the major components of examination frameworks, as well as the steps and criteria for bias and content review of the frameworks.

Composition of Examination Frameworks

An examination framework defines the content knowledge, skills, and abilities important for the job of an entry-level educator in the area being assessed. Pearson develops an exam framework for each Texas Educator Certification

Examination Program examination, guided by recommendations in the *Standards*:

> *The first step in developing test specifications is to extend the original statement of purpose(s), and the construct or content domain being considered, into a framework for the test that describes the extent of the domain, or the scope of the construct to be measured. Content specifications, sometimes referred to as content frameworks, delineate the aspects (e.g., content, skills, processes, and diagnostic features) of the construct or domain to be measured.… The delineation of the content specifications can be guided by theory or by an analysis of the content domain (e.g., an analysis of job requirements in the case of many credentialing and employment tests). The content specifications serve as a guide to subsequent test evaluation. (AERA, APA, & NCME, 2014, p. 76)*

Each Texas Educator Certification Examination Program examination framework is structured to include content domains, competencies, and descriptive statements that outline the knowledge and skills to be covered by the examination. Domains, competencies, and descriptive statements are described in more detail and further illustrated in Table 2.

- **Domains**. The framework is organized into multiple domains. The domains structure the subject matter for both examination preparation and score reporting. Together, the domains indicate the main areas of subject matter knowledge and skills important for the job of a public-school educator in Texas.
- **Exam Competencies.** The examination competencies are the key elements of the examination framework. They are intended to be broad, meaningful statements of the knowledge and skills important to the job of an educator in Texas public schools. The competencies define the range of knowledge and skills to be measured by the examination.
- **Descriptive Statements.** The descriptive statements further define each competency. Descriptive statements provide more detailed information about the content of a competency by including examples of the types of knowledge and skills covered by the competency. The descriptive statements are intended to provide only examples of the subject matter of a competency and do not cover the entire range of knowledge and skills represented by the competency.

Table 2. Sample Framework Format

Domain	READING AND VOCABULARY
Competency	Understand the literal content of a variety of authentic materials.
Descriptive Statement	Includes determining the stated main idea; summarizing; identifying character, setting or events described in a selection; and determining the sequence of events.

Equity Review of Examination Frameworks

The frameworks for the Texas Educator Certification Examination Program are reviewed by the EAP for potential issues of fairness and to verify the content reflects the diversity of Texas. They are then reviewed for content by the CACs; however, CAC committee members are also instructed to review for potential issues of fairness. For the review of the frameworks, Pearson trainers and facilitators provide information to EAP committee members regarding the background, purpose, and policies of the Texas Educator Certification Examination Program, and directions for completing the framework reviews. Committee members are trained in the definition of bias as well as the inclusive and exclusive aspects of bias review. They review the competencies included in the frameworks using criteria pertaining to content, language, offense, stereotypes, fairness, and diversity. Committee members, in both EAP and CAC settings, consider each competency as fair only if it meets all criteria for fairness and diversity.

The following criteria are provided to EAP and CAC members as they review the exam frameworks.

Content

Does the framework contain content that disadvantages a person because of her or his gender, race, nationality, national origin, ethnicity, religion, age, sexual orientation, disability, or cultural, economic, or geographic background?

Language

Does the framework contain language that disadvantages a person because of her or his gender, race, nationality, national origin, ethnicity, religion, age, sexual orientation, disability, or cultural, economic, or geographic background?

Offense

Is the framework, presented in such a way as to offend a person because of her or his gender, race, nationality, national origin,

ethnicity, religion, age, sexual orientation, disability, or cultural, economic, or geographic background?

Stereotypes

Does the framework contain language or content that reflects a stereotypical view of a group based on gender, race, nationality, national origin, ethnicity, religion, age, sexual orientation, disability, or cultural, economic, or geographic background?

Fairness

Taken as a whole, are the frameworks fair to all individuals regardless of race, gender, cultural background, or other personal characteristics?

Diversity

Taken as a whole, do the frameworks include content that reflects the diversity of the Texas population?

Content Review of Examination Frameworks

As required by the *Standards* (AERA et al., 2014), examination frameworks for licensure need to focus on knowledge, skills, and abilities necessary for safe and effective practice in the profession. Therefore, the role of the CACs in Texas is to consider if the frameworks are aligned with expectations for Texas educators, address important areas of Texas educator knowledge, skills, and abilities clearly and appropriately, and are free from potential bias.

Pearson provides information regarding the background and purpose of the Texas Educator Certification Examination Program and directions for completing the content review. Committee members review the examination frameworks for alignment, completeness, language and terminology, and freedom from bias. The criteria used to determine if revisions are needed to the framework included the following:

- improving alignment to Texas and/or national standards
- adding emerging content
- addressing potential bias
- enhancing job-relatedness
- increasing or decreasing the emphasis of one component of content versus another component to align with Texas needs
- incorporating terminology commonly used in Texas
- increasing representativeness of content with Texas educator preparation program curricula

CAC members were instructed to ask themselves a set of organized questions when reviewing the content of the exam framework. The questions relate to the framework structure: Program Purpose, Organization, and Inclusiveness.

Program Purpose

Is the framework consistent with the purpose of the Texas Educator Certification Examinations (i.e., to determine whether prospective educators have the knowledge and skills to perform the job of an educator in Texas)?

Organization

Is the framework organized in a reasonable way?

Inclusiveness

Is the content of the framework complete?

Does the framework reflect the knowledge and skills an educator should have to perform the job of an educator?

Is there any content that should be added?

The following questions were also considered by the CAC when reviewing the framework competencies and related sets of descriptive statements within the framework, as associated with Significance, Accuracy, Freedom from Bias, and Job-Relatedness.

Significance

Do the competencies and descriptive statements describe knowledge and skills that are important for educators to have?

Accuracy

Do the competencies and descriptive statements accurately reflect the content, as it is understood by educators in the field?

Are the competencies and descriptive statements stated clearly and accurately, using appropriate terminology?

Freedom from Bias

Are the competencies and descriptive statements free of elements that might disadvantage an individual because of her or his gender, race, ethnicity, nationality, national origin, religion, age, sexual orientation, disability, or cultural, economic, or geographic background?

Job-Relatedness

Do the competencies and descriptive statements cover important knowledge and skills that an educator should have to perform the job of a Texas educator?

In 2020-21, draft exam frameworks for the following fields were reviewed by an EAP and CAC:

- English Language Arts and Reading 7-12 (331)

- Deafblind EC-12 (185)
- Special Education EC-6 (186)
- Special Education 6-12 (187)

Framework development for the following fields was completed prior to 2020-21:

- School Counselor (252)
- Pedagogy and Professional Responsibilities for Trade & Industrial Education (370)
- Physical Education EC-12 (258)
- English Language Arts and Reading 4-8 (217)
- Core Subjects EC-6 English Language Arts and Reading (901)*

** Note: Frameworks for subtests 902-905 remained identical to 802-805 and did not go through framework review.*

Task 5: Conduct Content Validation Surveys

Content-related validity evidence is important in licensure testing because it provides evidence that the examination adequately represents the content domain of the occupation for which the examination is developed (AERA et al., 2014). The *Standards* (AERA et al., 2014) require a clear definition of content domain and a rationale to support a claim that the knowledge, skills, and abilities being assessed in a licensure test are required for credential-worthy performance. The sections below describe how Pearson typically selects participants and the criteria used to evaluate results of the content validation surveys.

Survey participants and ratings

Pearson targets practicing public school teachers and/or administrators and educator preparation faculty for participation in the Content Validation Surveys of examination frameworks in the Texas Educator Certification Examination Program. Using an interactive, online survey instrument, participants use a five-point scale to independently rate the components of the examination framework.

The survey consists of three questions pertaining to importance, representativeness, and completeness of the knowledge and skills presented in the frameworks.

Criteria for evaluating survey results

Pearson reviews overall ratings to determine the final status of the examination framework components. In general, a rating of 3.00 or higher is considered to be a clear indication that the content addressed by a competency is of the appropriate level of importance for an educator certification examination. Ratings below 3.00 may also be included if they are deemed to address topics that are articulated in the relevant state standards.

In 2020-21, Pearson conducted a content validation survey for the following fields:

- Core Subjects EC-6 English Language Arts and Reading (901)
- English Language Arts and Reading 7-12 (331)
- Deafblind EC-12 (185)

** Note: Frameworks for subtests 902-905 remained identical to 802-805 and did not go through a content validation survey.*

Framework development, including the content validation survey, for the following fields was completed prior to 2020-21:

- School Counselor (252)
- Pedagogy and Professional Responsibilities for Trade & Industrial Education (370)
- Physical Education EC-12 (258)
- English Language Arts and Reading 4-8 (217)

Task 6: Develop Examination and Item Specifications

Texas Educator Certification Examination Program items are developed in accordance with the final, validated examination frameworks and examination designs. Pearson assembles a team of content specialists, exam development specialists, editors, content reviewers, and equity advisors to develop the examination items and associated scoring rubrics to support a close link between the examination materials produced and the examination frameworks and to verify they meet Pearson's standards for editorial quality.

Pearson develops Exam Specifications for each field to inform the details and features of each examination and to guide item development needs. This includes the number and types of examination items on each form, the proportion or weighting of items from each competency, the allotted time to complete the examination, the proportion of scorable and non-scorable items and other guidelines relevant to form development. Information about each examination is located on the Texas Educator Certification Examination Program website: http://www.tx.nesinc.com/PageView.aspx?f=GEN_Tests.html.

Pearson prepares items for the Texas Educator Certification Examination Program by drawing from existing item banks and by drafting additional items as necessary. Examination item and material development involves a series of activities designed to produce an examination that is technically sound, reliable, and valid.

Pearson develops Item Specifications for each field to provide explicit content limits and requirements by competency in the exam framework to guide item development and provide the specifications by which items can be approved. Pearson prepares draft Item Specifications for TEA review, updates the specifications based on TEA feedback – including direct meetings as needed to further develop and clarify the specifications – until TEA approves the document for use in new item development. Any items drawn from existing banks and proposed for adoption into the new bank must meet the new approved Item Specifications for that bank or they are revised and reviewed again until they do.

In 2020-21, Pearson drafted new exam specifications only (not item

specifications) for a new Core Subjects EC-6 (391), as follows:

- Core Subjects EC-6 English Language Arts and Reading (901)
- Core Subjects EC-6 Mathematics (902)
- Core Subjects EC-6 Social Studies (903)
- Core Subjects EC-6 Science (904)
- Core Subjects EC-6 Fine Arts, Health, and Physical Education (905)

In 2020-21, Pearson drafted new exam specifications and item specifications for the following fields:

- English Language Arts and Reading 7-12 (331)
- Deafblind (185)
- Special Education EC-6 (186)
- Special Education 6-12 (187)

Development of exam and item specifications for the following fields was completed prior to 2020-21:

- School Counselor (252)
- Pedagogy and Professional Responsibilities for Trade & Industrial Education (370)
- Physical Education EC-12 (258)
- English Language Arts and Reading 4-8 (217)

Task 7: Prepare and Review Examination Items

Equity Assurance Panel (EAP)

Pearson conducts item review and validation conferences with the Texas EAP for items in new and re-developed item banks. The purpose of the meetings is to review draft examination items to verify that the materials are free from potential issues of fairness and reflect the diversity of Texas according to established review criteria. As the *Standards* state:

> *Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics. (AERA, APA, & NCME, 2014, p. 64)*

Pearson trainers and facilitators provide EAP committee members with information regarding the background, purpose, and policies of the Texas Educator Certification Examination Program, and directions for completing the review. Committee members are provided training in the definition of bias as well as the exclusive and inclusive aspects of bias review. If the committee indicates that an item contained a potential issue, they are asked to suggest possible revisions to address it. The EAP concerns and suggested revisions are shared with the content advisory committees.

Content Advisory Committee (CAC)

For new and re-developed examinations, Pearson conducts item review and validation activities with CACs. Committee members participate in a consensus

review of each item in the item bank. Once committee consensus is reached on an item, with or without revisions, each committee member provides an independent item validation judgment. Committee members rate each item as either "valid" or "not valid" according to the review criteria listed below.

- Match to competency or content domain
- Accuracy
- Freedom from bias
- Job-relatedness

An item is rated "valid" if it matches all four of the review criteria; an item is rated "not valid" if it fails to match one or more of the review criteria. When rating items "not valid," committee members are instructed to indicate which of the four criteria were not met and provided a written reason for the "not valid" rating. Pearson analyzes the item validation ratings and reviews the committee members' item validation comments.

In 2020-21, Pearson organized and facilitated Equity Assurance Panels and Content Advisory Committees to review, revise, and approve draft exam items in the following fields:

- Science of Teaching Reading (293)
- School Counselor (252)
- Pedagogy and Professional Responsibilities for Trade & Industrial Education (370)
- Physical Education EC-12 (258)
- English Language Arts and Reading 4-8 (217)

Task 8: Conduct Pilot Testing

Newly developed or revised examination items should be tried out through an established field test or pilot test process to determine whether the items function as intended and to assess statistical characteristics of new examination questions or forms. This step in the examination development process relies on a sample of test-takers to try out one or more aspects of a new examination or new examination items such as item response formats or options and new item types. The process also provides a check of the adequacy of testing procedures such as the clarity and accuracy of examination directions and the appropriateness of the allotted testing time (AERA, APA, & NCME, 2014).

Pilot testing provides another source of validity evidence by gathering data regarding the performance characteristics of the examination items. When testing volumes and candidate populations permit, Pearson conducts pilot testing to collect this data. For examinations with constructed-response sections, pilot testing also serves the purpose of obtaining authentic examinee responses for the establishment of markers and scorer training.

In 2020-21, Pearson conducted pilot testing for the following fields:

- School Counselor (252)
- Pedagogy and Professional Responsibilities for Trade & Industrial Education (370)
- English Language Arts and Reading 4-8 (217)
- Science of Teaching Reading (293)

Task 9: Build Operational Examination Forms

Strict measures of quality control supported the process of assembly, production, preparation, and transmission of the first operational exam form, including definition of examination form content, review of item data, and examination form blueprinting and assembly.

In 2020-21, Pearson created new operational examination forms for the following fields:

- Science of Teaching Reading (293)
- Early Childhood: PK-3 (292)
- Educational Diagnostician (253)
- Core Subjects EC-6 English Language Arts and Reading (901)
- Core Subjects EC-6 Mathematics (902)
- Core Subjects EC-6 Social Studies (903)
- Core Subjects EC-6 Science (904)
- Core Subjects EC-6 Fine Arts, Health, and Physical Education (905)

Task 10: Conduct Standard Setting

The process of establishing passing requirements on an examination is referred to as standard setting. Standard setting relates directly to the validity of the interpretations made about candidates based on their test scores because the process produces a recommended passing score. The recommended passing score defines the boundary line between the acceptable level of knowledge, skills, and abilities required of an entry-level educator and an unacceptable level of knowledge, skills, and abilities. The *Standards* require that passing scores be set high enough to distinguish adequate from inadequate performance, but not too high to be unreasonably limiting (AERA, APA, & NCME, 2014).

Hambleton and Pitoniak (2006) recommend that standard setting processes include the following nine steps:

1. Selecting standard setting method
2. Choosing panelists
3. Preparing performance-level descriptors
4. Training panelists
5. Collecting item ratings
6. Providing feedback to panelists
7. Compiling panelists' ratings to obtain performance standards
8. Conducting panelists' evaluation, and
9. Compiling validity evidence and preparing documentation

Standard setting method
The modified Angoff method (Angoff, 1971) is used to set passing scores for the PAIL selected-response items section of the first operational exam form. Using this method, subject matter experts review each item and marked the proportion of the target population that would provide a correct response. An extended-Angoff standard setting method is used for constructed-response items. Details about the two methods are provided in a later section below.

Standard setting panel
The committee established to participate in standard setting consists of public-school educators who are certified and practicing in the field, and faculty from Texas colleges and universities who are currently preparing (or have prepared) prospective Texas educators. Panel members are selected to include educators from public schools and preparation programs that represent districts and colleges across the state and generally reflect the diversity of the state of Texas.

Performance-level descriptor
Performance-level descriptors are descriptions of the skills and knowledge that candidates in a performance category should possess. Performance-level descriptors form the basis for making judgments and providing ratings in standard setting. Appropriate threshold candidate description(s) are referenced when establishing a passing score.

For the purpose of Texas standard setting, this description is referred to as the "Just Acceptably Qualified Candidate (JAQC).

The TEA defines the Just Acceptably Qualified Candidate (JAQC) for Texas teachers generally as:

> *An individual who is just at the minimum level of knowledge and skills needed to be an effective educator in Texas and positively contribute to student learning*

Training of panelists
Panelist training is a critical component in setting performance standards. Training allows panelists to receive information pertaining to the testing program, test development procedures, scoring, and the task the panelists are required to complete. Panel members for the standard setting meeting are given an orientation that explained the standard setting recommendation process, the materials they will use, the concept of the Just Acceptably Qualified Candidate, and the judgments about examination items that they are asked to provide.

Panelists also complete a training exercise, which includes items with a range of item difficulty and cognitive complexity, to prepare them for the actual rating activity. The roles of the TEA in establishing the final passing score are also explained.

Training also involves a simulated test-taking activity, which allows panelists to

become familiar with examination content. Panel members are asked to review the framework and then "take the exam." Each panel member is provided with a copy of an examination form and is asked to read and answer the questions on the exam without referring to the answer key. After responding to all questions, panelists are provided with the answer key and asked to score their own answers as circled in the booklet. Panelists are encouraged to write and make notes as needed in their copies of the exam booklet to assist them with their review.

Collecting item ratings

Standard setting panelists provide performance-level judgments for each item on the exam form. The facilitator provides training in this step of the standard setting process, in which panel members make item-by-item judgments using a modified Angoff procedure. Referring to the exam form they had reviewed and taken earlier, panel members are asked to provide their individual, independent judgments regarding the expected performance of the JAQC on each of the selected-response and constructed-response items.

Panel members are asked to envision a group of Texas educators who are just at the minimum level of knowledge and skills an educator needs to be an effective educator in Texas and positively contribute to student learning. Panel members are asked to refer to the concept of these candidates in making their judgments throughout the standard setting process. For the constructed-response items, panelists review the sample "marker" responses from the constructed-response items on the examination form and descriptions of performance at each score point on the scoring scale. An iterative procedure is used, in which standard setting ratings are gathered in multiple rounds.

Following the training for round one, panelists provide item by-item judgments of the performance on the selected-response items from the examination form. For constructed-response items, panelists provide passing score judgments that range from "2" to "8" points.

In round two, panelists are given an opportunity to change their ratings considering feedback from round one.

In some cases, standard setting may include a round three, in which panelists provide individual recommendations for an exam-level minimum passing score, based on feedback from rounds one and two.

Providing feedback

Giving feedback to panelists is advantageous in that it helps to improve consistency of ratings and reduces variability among panelists because the panelists can develop a clearer and common understanding of expected performance. Feedback also allows identification of inconsistent and outlier panelists who can be asked to review or justify their ratings.

Therefore, prior to the second round, panelists review results from the initial round of ratings as well as item difficulty data from the pilot test. Panel members are then given an opportunity to revise their individual round-one item ratings. For constructed-response items, as with the selected-response items, following

their first round of judgments, panelists are provided with a summary of their own and other panelists' first round ratings. This information is used together to provide a final recommendation in the second round.

If a third round is planned, panelists review results from the round two selected-response item-based passing score judgments, which are calculated into an exam-based passing score judgment for the set of scorable selected-response items. Panelists are then asked to provide an exam-level judgment regarding the number of scorable selected-response items to which a JAQC would respond correctly.

Following a group discussion, panelists are asked to consider the information presented, the purpose of the Texas Educator Certification Examination Program, the framework and exam items, and the requirements for the entry level principal. Panelists are then asked to provide individual recommendations for an exam-level minimum passing score.

In all rounds and throughout the ratings process, panelists are provided multiple opportunities to ask for further clarification of how to read and interpret the information.

Compiling ratings to get cut scores
After panelists complete their selected-response ratings, their rating forms are analyzed to produce a panelist-specific Round One Item Rating Summary to be given to each panelist. The report is prepared for each selected-response item, including the panelist's rating, the median rating provided by all panelists rating the item, and the distribution of ratings across all panelists. Similar calculations are made after round two ratings. The exam-level ratings provided in round three, if needed, are used to compute a median score of the committee, which is then presented to the TEA for approval.

Conducting panelists' evaluations of the standard setting process and procedures
Panelists' evaluations provide data that could be used to improve the process and that serve as a good source of procedural validity evidence. Panelists that participate in a standard setting meeting are given a chance to evaluate the standard setting process.

The panelists complete evaluation forms to provide their professional judgments about the standard setting meeting. On a five-point scale, panel members are asked to rate the various aspects of the meeting. Panel members are also provided space to make additional comments regarding the standard setting meeting proceedings.

Implementation of the Modified-Angoff Procedure for Selected-response Questions

Panelists provide standard setting ratings based on their professional judgment, their knowledge of their examination field, their understanding of the qualifications of prospective educators, the content of an examination form, and examination data, when available, about candidate performance on the examination form. This process conforms to Standard 5.22 of the *Standards*, which states:

> *When cut scores defining pass–fail or proficiency levels are based on direct judgments about the adequacy of item or test performances, the judgmental process should be designed so that the participants providing the judgements can bring their knowledge and experience to bear in a reasonable way. (AERA, APA, & NCME, 2014, p. 108)*

The modified-Angoff procedure requires panelists to use their professional judgment and understanding of the knowledge and skills of the target group to estimate the proportion of target examinees that would provide a correct response to each item. For standard setting for the certification of teachers in Texas, panelists are asked to respond to the following question:

> *Imagine a hypothetical group of individuals who are just at the minimum level of knowledge and skills needed to be an effective educator in Texas and positively contribute to student learning.*
>
> *What percent of this group would answer the item correctly?*

0% - 10% = 1	51% - 60% = 6
11% - 20% = 2	61% - 70% = 7
21% - 30% = 3	71% - 80% = 8
31% - 40% = 4	81% - 90% = 9
41% - 50% = 5	91% - 100% = 10

Panelists are instructed to keep in mind the hypothetical reference group. Individuals in the hypothetical reference group are defined as having a sufficient level of knowledge, skills, and abilities needed to serve as initially licensed educators or instructional leaders.

Implementation of the Extended-Angoff Procedure for Constructed-response Questions

For a description of focused holistic scoring, see the next section "Scoring and Reporting."

Standard setting panelists review the sample "marker" responses from the constructed-response item(s) on the exam form and descriptions of performance at each score point on the scoring scale. The marker responses function as guides for focused holistic scoring. Panelists then provide a passing score judgment that ranges from "2" to "8" points for each constructed-response item. As with the selected-response items, following their first round of judgments, committee members are provided with a summary of their own and other committee members' first round ratings. This information is used together to provide a final recommendation in the second round.

In 2020-21, Pearson conducted standard setting for the following fields:

- Science of Teaching Reading (293)
- Early Childhood: PK-3 (292)

- Educational Diagnostician (253)
- Core Subjects EC-6 English Language Arts and Reading (901)

Task 11: Establish Passing Standards

A Standard Setting report and presentation are provided to TEA including the outcomes of the Standard Setting Conference. This includes the recommended passing score made by the committee, which is then considered in the process of establishing the final approved passing standard.

Texas Education Code, §21.048(a) requires the commissioner of education to determine the satisfactory level of performance for each educator certification examination and a satisfactory level of performance on each core subject covered by an examination.

In 2020-21, TEA established new passing standards for the following fields:

- Science of Teaching Reading (293)
- Early Childhood: PK-3 (292)
- Educational Diagnostician (253)

The design of the Core Subjects EC-6 (391) subject exams was modified slightly from the 291 designs, and a new standard was set for each of the new subtests:

- Core Subjects EC-6 English Language Arts and Reading (901)
- Core Subjects EC-6 Mathematics (902)
- Core Subjects EC-6 Social Studies (903)
- Core Subjects EC-6 Science (904)
- Core Subjects EC-6 Fine Arts, Health, and Physical Education (905)

Scoring and Reporting

This section of the report addresses the scoring and reporting processes and procedures for the Texas Educator Certification Examination Program generally.

Selected-response Item Scoring

Scoring verification of selected-response items delivered within a computer-administered testing environment is comprehensive and continuous. After the complete set of examination forms for a score reporting date have been scored, a comprehensive quality control system verifies accurate computation of candidate scores. As part of the quality control process, Pearson prepares a preliminary item analysis report and flags items that do not perform within defined statistical parameters and the overall form performance for each test form administered. Score review is not available to candidates for exams that have only selected-response items because of the quality-control measures in place to deliver accurate scoring of selected-response items.

Constructed-response Item Scoring

The responses to the constructed-response items are scored using a method

known as focused holistic scoring. In this method, scorers judge the overall effectiveness of each response while focusing on a set of performance descriptions that have been defined as important. These performance descriptions guide scorers in the assignment of holistic scores in order that uniform criteria are used to assign a score to each response.

Though this method focuses on specific descriptions, it is holistic in that each assigned score describes the overall effectiveness of these descriptions working in concert. Each response is rated on a scale of "1" to "4," with a "1" representing a response that reflects very weak or no understanding of the relevant knowledge and skills and a "4" representing a response that reflects a thorough understanding of the relevant knowledge and skills. The performance descriptions are used to form the basis of the four-point scale.

Each examinee response to a constructed-response item is evaluated independently by two or more trained and calibrated scorers. Each response is assigned a numerical score on the four-point scale from each scorer. The two independent scores are combined, resulting in a total constructed-response item score that ranges from 2 to 8. For example, if two scorers both assigned a score of "3" to an examinee response, the total raw score for the response would be "6" ("3" + "3").

Exam takers who do not pass may request a score review for exams with a written- or spoken-response section (including an interview), and only for that written- or spoken-response section (or interview).

Support Materials and Other Tools

Pearson developed multiple web-based Texas Educator Certification Examination Program preparation tools designed to help candidates prepare to take relevant examinations. The preparation tools are available on the Texas Educator Certification Examination Program website: http://www.tx.nesinc.com/. The online tools are designed to accommodate varying methods of preparation (independent study or under the direction of an instructor or faculty advisor); areas of focus (examination content); and opportunities for access (at school, at home, while traveling). Educator preparation program faculty can also gain an understanding of the examinations and how to help their candidates prepare through resources available on the Educator Preparation Program page: http://www.tx.nesinc.com/PageView.aspx?f=GEN_FacultyResources.html.

Candidates have access to the following resources to guide their preparation.

- **Examination Frameworks.** Examination frameworks include the content domains and competencies covered by each examination. In each framework, the competencies are organized into content domains that reflect the main areas of pedagogical or content area knowledge included on the examination. Descriptive statements provide details about the nature and range of content covered by each competency.
- **Examination Preparation Manuals.** Online preparation manuals are available for each exam. The preparation manuals include an overview of the examination format, number of questions, testing duration,

competencies with descriptive statements, and sample exam questions with rationales. The preparation manuals also include information to assist candidates in preparing for and taking the examinations.

  - **Note:** A preparation manual for each PACT exam was adopted along with the exam itself for use in Texas. PACT preparation manuals include the examination framework as well as other features described above.

- **Interactive Practice Exams.** Online interactive practice exams simulate the computer-administered testing experience. The interactive practice exam can be taken in real time or paused and returned to at any time. Test-takers receive a competency-level report with instant scoring of selected-response questions and explanations of correct responses.
- **Computer-administered Testing Tutorials.** Two tutorials are available to examinees on the program website. One tutorial is designed to help familiarize examinees with the navigation tools and operations of computer-administered testing. It includes information about how to navigate through an examination, select answers, and end the examination. The second tutorial is downloadable and interactive. It guides examinees on how to record, change, and review answers. The tutorial also gives examinees the opportunity to practice using various functions of the computer-administered environment, including viewing visuals and exhibits, scrolling pages, reviewing items, typing in an essay box, and using an on-screen calculator.
- **Faculty Resources.** EPP faculty have access to an array of resources, including specially designed worksheets that may be used to map the examination framework content to the program curriculum. This mapping may assist in assessing the degree of alignment between the knowledge and skills taught in the preparation program curriculum and in the content of the Texas Educator Certification Examination Program frameworks.
- **Exam Preparation Worksheets.** Candidates can complete worksheets to assess their preparedness to test. Faculty can review the completed worksheets to help assess whether a candidate may be ready to test based on the content covered in their coursework and guide further candidate preparation.

Program Year 2020-2021 Outcomes

Each of the examinations that were launched in 2020-21 was validated using industry-approved processes, including the review of content, job-relatedness, and prevention of bias. Texas educators and educator preparation program faculty and administrators made recommendations for the selected-response and constructed-response passing scores for the examination.

Pilot Testing Background

The goal of pilot testing is to gather empirical data about the statistical and qualitative characteristics of newly developed or revised examination questions that allows for a determination to be made regarding their usability on future operational examination forms.

Ideally, pilot test participants should have the same general characteristics as those of actual candidates who will take the examination to meet testing requirements for Texas certification. TEA and Texas educator preparation program faculty assisted in the recruitment of candidates or recently certified educators in the designated fields for participation in each pilot test. Potential participants were required to be actively enrolled in an EPP in Texas in a related field or had taken the existing exam within the past year. As incentive for providing genuine effort in their participation, candidates were offered a $50 Amazon.com gift card.

Pilot Testing Summary

The demographic characteristics of participants by field and by form are shown in Table 3.

Table 3. Demographics by Pilot Test Form

Educational Diagnostician (253)	
Gender	**Total N**
Male	3
Female	67
Ethnicity	**Total N**
Asian or Pacific Islander	1
Black or African American	20
Hispanic/Latino	22
White	22
Two or more races	2
Declined to answer	5
Total	**72**

Early Childhood PK-3 (292)	
Gender	**Total N**

Male	3
Female	38
Ethnicity	**Total N**
Asian or Pacific Islander	1
Black or African American	8
Hispanic/Latino	22
White	9
Two or more races	0
Declined to answer	0
Total	**41**

Science of Teaching Reading (293)	
Gender	**Total N**
Male	11
Female	151
Ethnicity	**Total N**
Asian or Pacific Islander	9
Black or African American	2
Hispanic/Latino	85
White	53
Two or more races	9
Declined to answer	3
Total	**162**

Pearson analyzed the results of the pilot test selected-response items to determine item performance under pilot testing conditions. Items were statistically evaluated using industry standards under classical test theory (i.e., item difficulty, item discrimination, and differential item functioning). The items were reviewed and verified by Pearson's psychometric and examination development staff to recommend to TEA the items' final disposition (i.e., retain for future use, or remove from the operational item bank).

The responses from constructed-response (CR) items were scored using the approved operational procedures, performance characteristics, and score scales established for the Texas Educator Certification Examination Program. Quantitative item statistics were generated, and qualitative analyses were conducted to indicate the following.

- Mean, standard deviation, and minimum and maximum scores by item;
- Score distribution as a percentage for each item;
- Items that elicited a comparatively high number of blank, short, incomplete, or low-scoring responses;
- Items that scorers identified as difficult to score, or presented other concerns;
- Items with a comparatively high number of scorer discrepancies; and

- Items that participants identified in the post-questionnaire as difficult, unfair, or generally unsatisfactory.

Based on these analyses, a CR item was flagged if the data revealed a restricted score distribution; a comparatively high number of blank, short, incomplete, or low-scoring responses; a comparatively high number of discrepant scores; and/or participant or scorer comments that identified an item as particularly difficult or generally unsatisfactory (e.g., directions unclear). All flagged items were subjected to further psychometric and content expert reviews.

Additionally, a Tukey analysis was performed on the CR items administered during the pilot test. Equality across CR items within an item bank is imperative when included as part of an examination design. Test takers should neither be advantaged nor disadvantaged based on the specific CR item that they receive. Equivalency between all CR items of a particular type must, therefore, be established and maintained. The process of establishing equivalency of CR items begins during exam development, continues through the analysis of pilot test data, and is maintained during holistic scoring of operational CR items.

Initially, Pearson content experts and holistic scoring staff complete a qualitative review of the items. One of the goals of this review is to establish an anchor item. One purpose of the anchor item is to assist in establishing equivalency with all other CR items currently in the item bank or to be developed at a future point in time. After pilot testing, the mean performance of the CR items is compared in a pair-wise analysis called the Tukey test. The purpose of this test is to establish equivalency of the CR items. CR items are placed in Tukey groups based on the statistical similarity of their performance. If the difference in performance between CR items is statistically significant, the CR items will not be assigned to the same Tukey group. Similarly, if the difference in performance is not statistically significant, the CR items will be assigned to the same Tukey group. Thus, if a CR item is not in the same Tukey group as the anchor item, it is not equivalent to the anchor item and is ineligible for operational use.

Further, it is possible for the anchor item to fall within multiple Tukey groups. If this situation occurs, Pearson test construction staff will select an appropriate Tukey group for operational use. Criteria for selecting the Tukey group may include the size of the group (number of CR items within the group), the placement of the anchor item statistically within the group (end point versus central value), and the overall statistical performance of the group (range of the CR item means, difficulty level of the CR items, etc.). CR items that are not part of the selected Tukey group are not eligible for operational use. Note that this is not the only criteria for determining the eligibility of CRIs. CRIs may also be deemed ineligible based on content expert reviews, examinee comments, or holistic scoring reviews.

Additionally, a latency analysis was performed using the pilot test data to determine whether the overall time allotted for the examination is appropriate and to determine whether any items were taking longer than anticipated to complete.

Results from the pilot tests informed the development of the operational

examination form used for standard setting. In keeping with industry standards, new items for each examination will be included on operational forms in the non-scorable slots for future evaluation.

Standard Setting Background

Standard setting activities were conducted for each of these three fields with Texas educators in February 2021. The standard setting committees convened for these exam fields included experienced Texas educators and EPP faculty. Pearson used an industry accepted modified Angoff Standard Setting process described under Task 10 of this manual. The exam designs by item type (i.e., SRI and where present, CRI) for each of the exams are provided in the Appendix.

At the end of the standard setting meeting, the Standard Setting Panelists were asked to complete an evaluation of the standard setting process. An evaluation form requested their feedback about the training they received, their confidence in their ability to provide standard setting ratings, and any comments regarding the Standard Setting Conference proceedings.

Standard Setting Summary

The panelists recruited for each Standard Setting Conference were drawn from Texas educators and EPP faculty. Each Standard Setting Committee (SSC) consisted of between 6-14 members. All panelists were reviewed and approved by TEA prior to serving on the committee. The demographic distribution of members is shown in Table 4.

Note: The sum of demographic percentages may not equal 100% due to rounding.

Table 4. Standard Setting Committee Demographics

	EPP Faculty		Public School Educator		Total	
	N	**%**	**N**	**%**	**N**	**%**
Educational Diagnostician (253)	1	8%	12	92%	13	100%
Female	-	-	11	92%	11	85%
Male	1	100%	0	-	1	8%
No Response	-	-	1	8%	1	8%
African-American	-	-	3	25%	3	23%
American Indian or Alaska Native	-	-	-	-	-	-
Asian	-	-	-	-	-	-
Hispanic	-	-	3	25%	3	23%
Native Hawaiian or Other Pacific Islander	-	-	1	8%	1	8%
White	1	100%	4	33%	5	38%
No Response	-	-	1	8%	1	8%

	EPP Faculty		Public School Educator		Total	
	N	**%**	**N**	**%**	**N**	**%**
Early Childhood PK-3 (292)	4	29%	10	71%	14	100%
Female	4	100%	10	100%	14	100%
Male	-	-	-	-	-	-
No Response	-	-	-	-	-	-
African-American	1	25%	2	20%	3	21%
American Indian or Alaska Native	-	-	-	-	-	-

Asian	-	-	1	10%	1	7%
Hispanic	1	25%	2	20%	3	21%
Native Hawaiian or Other Pacific Islander	-	-	-	-	-	-
White	1	25%	3	30%	4	30%
No Response	1	25%	2	20%	3	21%

	EPP Faculty		**Public School Educator**		**Total**	
	N	**%**	**N**	**%**	**N**	**%**
Science of Teaching Reading (293)	2	33%	4	66%	6	100%
Female	2	100%	4	100%	6	100%
Male	-	-	-	-	-	-
No Response	-	-	-	-	-	-
African-American	1	50%	-	-	1	17%
American Indian or Alaska Native	-	-	-	-	-	-
Asian	-	-	-	-	-	-
Hispanic	-	-	-	-	-	-
Native Hawaiian or Other Pacific Islander	-	-	-	-	-	-
White	1	50%	2	50%	3	50%
No Response	-	-	2	50%	2	33%

Based on the results of the Round 2 judgments, the overall committee recommended passing scores were calculated for each field. The committee recommended raw passing scores were provided to TEA for review along with passing scores at 1 and 2 standard error of measurement (SEM) adjustments above and below the committee recommendations.

Each time a candidate takes an examination, there is a random chance that the score will be slightly different, and applying the SEM is one way to take this into account. The SEM allows educational analysts to determine a range of scores an examinee would receive if tested repeatedly without studying or other remediation between attempts. By adjusting the committee-recommended passing score to account for this range, a policy board can adjust for the likelihood of "false positive" or "false negative" results.

The assessment validation activities described above, and in greater detail throughout this manual, provide support that the examination fields are aligned to the state's need for a system of evaluating whether potential educator candidates have the minimum level of knowledge and skills needed to be an effective educator in Texas and positively contribute to student learning.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Angoff, W. H. (1984). *Scales, Norms and Equivalent Scores*. Princeton, NJ: Educational Testing Service.

Evaluation Systems group of Pearson. (2009). *Fairness and diversity in tests*. Amherst, MA: Author.

Hambleton, R.K., & Pitoniak, M.J. (2006). Setting Performance Standards. In R.L. Brennan (Ed.), *Educational Measurement* (6th ed., pp. 433–470). Westport, CT: American Council on Education and Praeger Publishers.

Appendix

Below are test designs for each of the exams launched in 2020-2021:

- Science of Teaching Reading (293)
- Early Childhood: PK-3 (292)
- Educational Diagnostician (253)
- Core Subjects: Early Childhood-Grade 6 (391)*
    - Core Subjects: EC-6 English Language Arts and Reading (901)
    - Core Subjects: EC-6 Mathematics (902)
    - Core Subjects: EC-6 Social Studies (903)
    - Core Subjects: EC-6 Science (904)
    - Core Subjects: EC-6 Fine Arts, Health, and Physical Education (905)

    *Exam 391 is composed of 5 subtests (Subject Exam I-V)

Each table provides the percentage of questions from each of the framework domains and competencies by item type (SRI, CRI, when present).

Note: The sum of approximate percentages may not equal 100% due to rounding.

Science of Teaching Reading (293)				
Domain	**Competency**	**Approx. % of SR Items**	**Approx. % of CR Items**	**Approx. % of TOTAL Test**
001	001-002	17%	.	13%
002	003-008	50%	.	43%
003	009-012	33%	.	24%
004	013	.	100%	20%
SRI Total Items:	90			
CRI Total Items:	1			

Early Childhood: PK-3 (292)				
Domain	**Competency**	**Approx. % of SR Items**	**Approx. % of CR Items**	**Approx. % of TOTAL Test**
001	001-003	23%	.	19%
002	004-005	15%	.	12%
003	006-007	15%	.	12%
004	008-009	15%	.	12%
005	010-013	31%	.	25%
006	014	.	100%	20%
SRI Total Items:	90			
CRI Total Items:	1			

Educational Diagnostician (253)				
Domain	**Competency**	**Approx. % of SR Items**	**Approx. % of CR Items**	**Approx. % of TOTAL Test**
001	001-003	43%	.	34%
002	004-005	29%	.	23%
003	006-007	29%	.	23%
004	008	.	100%	20%
SRI Total Items:	90			
CRI Total Items:	1			

Core Subjects: EC-6 English Language Arts and Reading (901) Subject Exam I				
Domain	**Competency**	**Approx. % of SR Items**	**Approx. % of CR Items**	**Approx. % of TOTAL Test**
001	001-010	100%	.	100%
SRI Total Items:	45			

Core Subjects: EC-6 Mathematics (902) Subject Exam II				
Domain	**Competency**	**Approx. % of SR Items**	**Approx. % of CR Items**	**Approx. % of TOTAL Test**
001	001-006	100%	.	100%
SRI Total Items:	45			

Core Subjects: EC-6 Social Studies (903) Subject Exam III				
Domain	**Competency**	**Approx. % of SR Items**	**Approx. % of CR Items**	**Approx. % of TOTAL Test**
001	001-005	100%	.	100%
SRI Total Items:	40			

Core Subjects: EC-6 Science (904) Subject Exam IV				
Domain	**Competency**	**Approx. % of SR Items**	**Approx. % of CR Items**	**Approx. % of TOTAL Test**
001	001-018	100%	.	100%
SRI Total Items:	45			

Core Subjects: EC-6 Fine Arts, Health, and Physical Education (905) Subject Exam V				
Domain	**Competency**	**Approx. % of SR Items**	**Approx. % of CR Items**	**Approx. % of TOTAL Test**
001	001-005	100%	.	100%
SRI Total Items:	40			

Developed and Produced by

Evaluation Systems

300 Venture Way, Hadley, Massachusetts 01035