IOWA STATEWIDE ASSESSMENT OF STUDENT PROGRESS

Development and Review of Items and Materials for ISASP

Iowa Testing Programs
Prepared by Catherine Welch and Stephen Dunbar

December 2025

Development and Review of Items and Materials for ISASP

Since 2019, the Iowa Statewide Summative Assessment of Student Progress (ISASP) has been administered annually by school districts. Iowa Code Standards* identify the purposes of the assessment as "accurately describe student achievement and growth for purposes of the school, the school district, and state accountability systems; provide valid, reliable and fair measures of student progress toward college or career readiness."

This document describes the item development and review processes undertaken to measure the Iowa Core Standards. Documentation of the completion of these processes is a required component of the U.S. Department of Education's Assessment Peer Review Process.

Development Processes

Sound item development is critical for providing quality and consistency across forms of the ISASP assessments. Items and stimulus/item sets (reading passages, graphs, maps, tables, etc. that support a group of items) are created according to the test specifications. The content domains, number of items per domain, cognitive levels, and item types are defined in the test specifications and serve as a basis for item writing. The initial development of items and related testing materials is the first critical step in an extensive, iterative process of drafting, rewriting, editing, aligning, and reviewing items. Only at the end of this extensive process are items considered eligible for inclusion on an ISASP form.

Item writers for the ISASP program are educators who are knowledgeable about the Iowa Core Standards and about Iowa students. Iowa Testing Programs (ITP) works with Iowa educators to identify, select, and contract with individuals for item writing assignments. Hundreds of Iowa educators have contributed to the item writing process for ISASP. These individuals are representative of the state teacher population and have extensive experience with students who are representative of Iowa's student test-taking population in terms of geographic region, demographics, and district size. Educators from 73 counties in Iowa have contributed to date. New educators are recruited to this process each year through workshops, seminars, and presentations across the state.

ITP content specialists routinely convene item writing workshops and train educators on sound item writing practices. Specific guides for writing test materials for each ISASP content area summarize general item writing principles and provide support resources for item writers. Educators are assigned to write items in the content areas and grade levels that best align with the Iowa Core Standards consistent with their expertise and experiences. ITP employs procedures for item writers that protect the security of the assessment materials as well as the confidentiality of the item writing assignments using secure file transfer protocols and nondisclosure agreements.

**The Iowa Core Standards were in place for the development and assembly of ISASP forms between 2019 and 2025. Beginning with the 2026 ISASP tests, the Iowa Academic Standards for English Language Arts and Mathematics will be used to guide development in these areas.*

To assist in the evaluation of open-ended items, writers who are developing such items also draft scoring criteria simultaneously. The scoring criteria are used to evaluate each item's alignment to the Iowa Core Standards as well as understand the cognitive demands required by the item given the rubric by which it will be scored. The complementary process for writers of selected-response items is that the item writer is expected to provide substantive rationales for the keyed response as well as distractors. Reviewers use these rationales in alignment and other validation activities during the item development process.

Item production goals ensure an "overage" of items across assessment areas so that the pool of available items for each ISASP assessment contains far more material than is needed to build each form. This overage allows content experts to discard those items that do not survive internal and external item review or post field test data review.

Review Processes

After items are written, content specialists review them individually and collectively for issues related to content accuracy, balance of topics, fairness, universal design, and alignment to the Iowa Core Standards. The goal of these reviews is to ensure items are accurate and accessible to all Iowa students. Consistent with Peer Review Guidelines, items are developed by individuals with content area expertise, experience as educators, and experience with the student population in Iowa.

The review processes check for alignment to the Iowa State Standards, the level of cognitive complexity identified in the Iowa Core Standards, and construct-relevance. Construct-relevance for ISASP means that the assessment measures the Iowa Core Standards. If items and testing materials are consistent with and aligned to the Iowa Core Standards, they are construct-relevant. Aspects of the items and testing materials that are not measuring the Iowa Core Standards are considered construct-irrelevant. For example, verbal ability is construct-relevant to the reading test, but construct-irrelevant to some parts of the math test. Critical to the review processes is the elimination of construct-irrelevant sources. *Table 1* summarizes the overall various steps in the development of ISASP assessments.

A series of reviews are conducted for all ISASP items. Each review is intended to focus on a different attribute of the assessment. Provided below is a description of the various reviews. The intention is that all test takers are treated respectfully and impartially throughout the testing process.

Table 1. ISASP Review Procedures in Test Development

Test Development Stage	Considerations for Valid Interpretations
Articulation of test purpose and constructs to be measured	Delineation of the construct to be measured Review of the curricular standards for appropriateness or accessibility issues
Test specifications	Educators who are representative of the test taker population for which the achievement test is designed
Item development	Item writers who represent the test takers for which inferences will be made materials Educators who have experience teaching the core content that the achievement test is designed to measure
Alignment	Aligners who represent the test takers for which inferences will be made Educators who have experience teaching the core content that the achievement test is designed to measure to a diverse and representative sample of test takers
Item review	Reviewers who represent the test takers for which inferences will be made Reviewers provided guidance and training about what to look for when reviewing items Reviewers made aware of potential issues with respect to cultural stereotyping, irrelevant characteristics of an item, sensitive topics, and offensive language Reviewers made aware of the principles of universal design
Pilot testing	Test takers who are representative of the total test taking population All delivery modes pilot tested on a representative sample All item formats pilot tested on a representative sample
Field testing	Proportional representativeness of the test taker population for which the achievement test is designed
Generation of item-level and test-level statistics	Disaggregated item-level statistics to allow for comparison of performance across students (reliability estimates, precision estimates, relationships between domains)
Assembly of forms/pools	Balance of forms with respect to content using items that have successfully cleared previous steps in the test development process
Review of forms/pools	Reviewers who represent the test takers for which inferences will be made Reviewers provided guidance and training about what to look for when reviewing items Reviewers made aware of potential issues with respect to cultural stereotyping, irrelevant characteristics of an item, sensitive topics, and offensive language Reviewers made aware of the principles of universal design
Linking, equating, and scaling	Special studies designed to collect evidence for any post-administration adjustments or links should be designed to select samples that are representative of the total test taker population

Content and Alignment Review

Once the items have been reviewed internally, ITP convenes panels of Iowa educators to review the items and associated stimuli (reading passages, tables, graphs, maps, etc.). After a formal training session in the review process, educators evaluate the items for grade level alignment, content relevance, and accuracy. Since these external reviewers have not been involved in the development process up to this point, they provide an objective "cold read" of test materials for potential concerns and unintended interpretations. A main goal of the content review is to confirm that the items are aligned at the appropriate grade level, content standard, and cognitive level. ITP development staff processes the information obtained for each item and determines whether further editorial work is needed. This review focuses on any edits made to the items throughout the process and again checks for content accuracy, fairness, and universal design.

Fairness Review

For review purposes, the term fairness can be defined as the extent to which test scores are valid for all test takers. Careful consideration of the issues related to fairness is required at each step of the test development process for the ISASP. Reviewers follow guiding principles as they consider each item, including suggested revisions to avoid construct-irrelevant variance and to allow all students the same opportunity to show what they know. Specifically, to make items accessible to all students, reviewers are asked to consider whether items contain the following:

Unnecessarily difficult language. It is best practice to keep testing language simple and direct. The test should use accessible language. While the use of accessible language is particularly important for test takers who have limited English skills, it is beneficial for all test takers when linguistic competence is not relevant to the construct the test intends to measure.

Unfamiliar language/vocabulary. The test should use language that is common and consistent with the level identified in the Iowa Core Standards. Items should avoid words or phrases that are associated with irrelevant content or topics.

Regionalisms. Test language should not require knowledge of words, phrases, or concepts more likely to be known in some areas of Iowa than in others. It is best practice to use words and phrases that are understood across the state in rural, urban, and suburban school settings.

Jargon. Items should not contain specialized language that is difficult for others to understand. Test language should avoid technical terms relating to finance, politics, specific professions, cultures, or regions.

Emotional topics. Test content that is unnecessarily controversial, offensive, or upsetting should be avoided when possible. It is best practice to avoid topics that may evoke feelings of discomfort, fear, sadness, or anxiety in test takers.

Stereotypes. Test content should be respectful of all students. Stereotypes attempt to classify people based on a single aspect, such as age, race, ethnicity, religion, income level, geographic region, or gender.

After receiving training in the principles outlined above, a committee of Iowa educators evaluates each item and stimulus through a formal review before the items are field tested. Each item is field tested on students in the state of Iowa to help gauge the appropriate level of difficulty for the item before it is used operationally. Committee members are educators with experience teaching the Iowa Core Standards, and those with experience teaching students with disabilities and English language learners. After the review is complete, items may be revised based on the feedback received, or they may be removed from the potential item pool. After items are field tested, items are further examined using statistical analyses to understand the difficulty level of the items.

Universal Design Review

The principles of universal design for the ISASP assessments provide guidelines for the test development and review processes to help ensure that students with special needs or incomplete language mastery are treated comparably. Universal design was a guiding principle in the creation of the publishing specifications that determine the appearance of the materials as they are experienced by students in both paper-and-pencil and online formats. Aspects of universal design including ease of navigation of test materials; clarity of typeface, graphics, and page layout; visual materials are amendable to verbal descriptions; and items can be adjusted for accommodations such as large type or increased contrast. The ISASP Accessibility and Accommodations Manual provides additional detail about the types of services available for students:
https://ia.mypearsonsupport.com/resources/manuals/IA1165188_ISASP_AccomsMan_26_WEB.pdf

External Alignment Review

ITP contracted with an external partner to conduct an external alignment study for the ISASP to establish and document evidence of consistency among the test blueprints, items, and the Iowa Core Standards. The alignment study included evaluations of assessments in ELA and Mathematics in grades 3–11, and of assessments in Science in grades 5, 8, and 10. The assessments were evaluated using an approach derived from the methodology established by the Council of Chief State School Officers (CCSSO, 2013). The evaluation of the Science assessments was further informed by criteria outlined by Achieve. The approach convened teachers and content experts to confirm the standards alignment and cognitive complexity levels of items the item writers identified (captured in item metadata in the content

management system used by ITP), and to rate all items on several other indicators of item quality.

The alignment study was conducted in two phases. During Phase 1, 34 Iowa educators representing 23 districts and eight regions of the state were convened as a panel for a two-day workshop during which they reviewed test items. Panelists were experienced Iowa educators with expertise in the content area and grade span for which they reviewed items. Panelists were organized into three groups each for ELA and Mathematics (3–5, 6–8, and 9–11; six groups total), and two groups for Science (5/8 and 10).

During this phase of the alignment process, panelists provided independent ratings for items but ultimately reached a consensus rating for each item based on group discussion. Data from Phase 1 were used to edit or replace items prior to the finalization of test forms for the ISASP. Phase 2 of the study convened a subsample of the Iowa educators who participated in Phase 1 along with one nationally recognized subject-matter expert for each content area. During Phase 2, revised and replacement items were rated using the same process implemented in Phase 1.

This study provided substantial evidence to support the content validity of the ISASP assessments in ELA, Mathematics, and Science. Across the grade/subject tests, a large majority of items were rated as measuring content outlined in the Iowa Core Standards. With a small number of exceptions, the number of aligned items fell within the ranges of items specified in the test blueprints. A similar study will be conducted with the revised state of Iowa standards.

Final Review

Although the items are reviewed extensively throughout the test development process, a specific review occurs in preparation for the assembly of final forms. This review takes place within the ISASP item banking system. This allowed for maximum test and item security, as well as allowing reviewers to experience all item types in the testing environment as experienced by test takers. Comments and ratings of the items are securely recorded within the platform. The reviewers were instructed to use the following categories for ratings:

- Approved: The item is approved as is, with no changes.
- Approved with edits: The item has a small issue but can be approved following edits to fix the issue.
- Rejected: The item has inherent flaws that cannot be fixed. The item should be removed from the item pool.

Forms Assembly

After the completion of all review processes, items that ITP has determined are available to appear on operational test forms become part of a pool of items that are eligible for selection

during forms construction. To ensure the final subject area test has adequate content coverage while at the same time being meaningful to students of varying achievement levels, the items within a typical subject area's item pool are chosen to be appropriate regarding skill alignment, cognitive level alignment, and difficulty. Items are then pulled from the item pool into test forms. During this process, careful attention is paid to item selection so that the final tests follow the predetermined test specifications and meet psychometric targets.