

assess²know™

Item and Test Development Process for *Assess2Know*®



Riverside Publishing

A HOUGHTON MIFFLIN COMPANY



Table of Contents

Passage and Item Development Workflow.....4

Evaluating Test Materials..... .6

Guidelines for Passage Development.....9

Review Checklist for Passages..... .13

Guidelines for Item Development.....14

Review Checklist for Multiple-Choice Items.....18

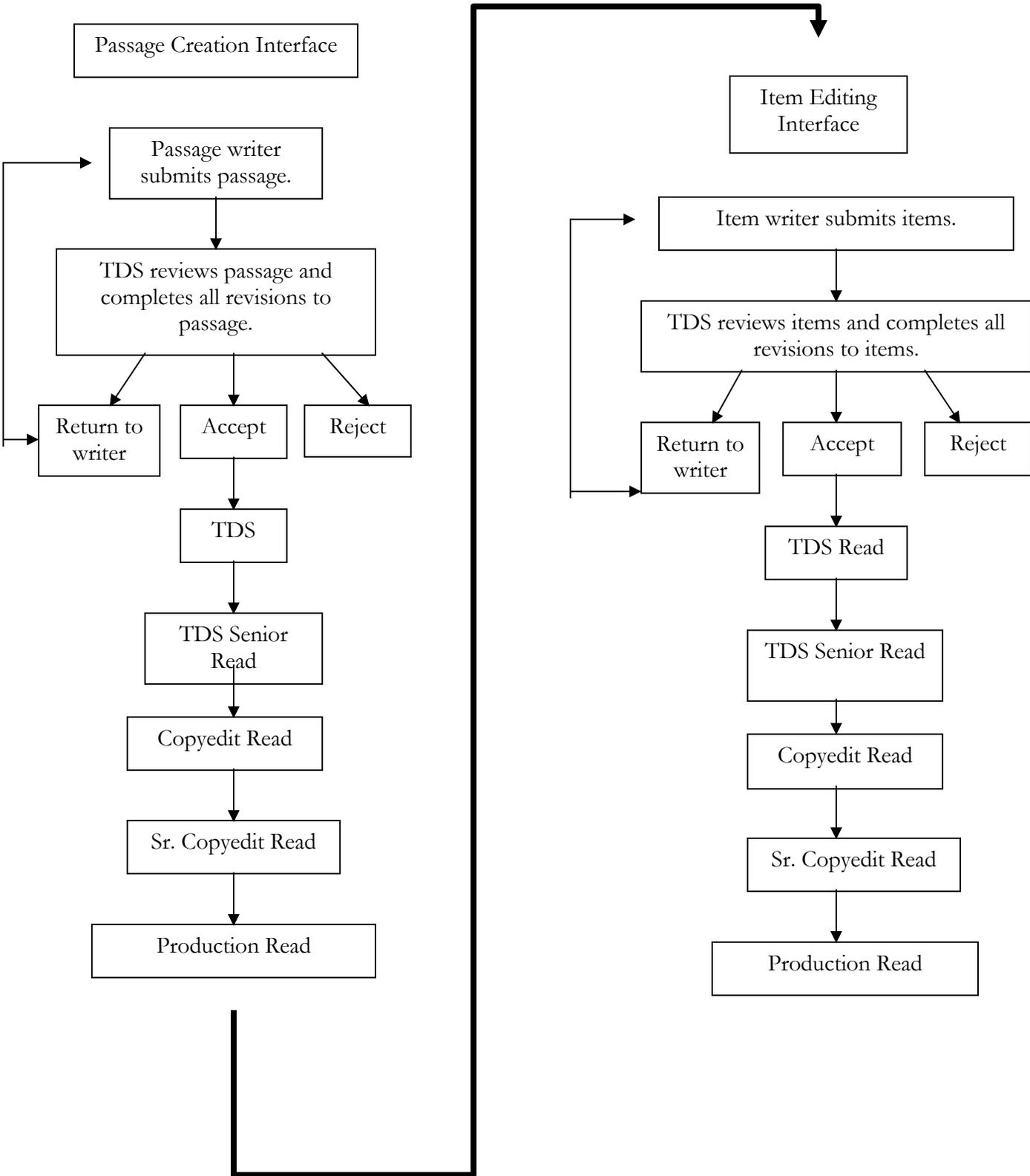


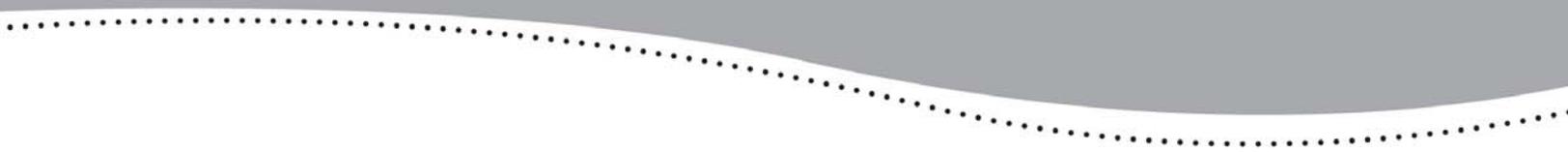
Passage and Item Development Workflow

To ensure the creation of high quality passages, items, and assessments time after time, Riverside has established processes and procedures with accompanying checklists that guide item and test development. Periodically the processes and procedures are modified to keep pace with changes in assessment or educational philosophies or to adjust to improved technology. Each time changes are implemented, checklists and other forms of documentation are revised to maintain quality and stay current with today's standards in test publishing.

The following chart and accompanying descriptions provide a high-level overview of the test development process and describe the qualifications and roles of the staff involved.

Development Flow Chart





Passage Writers: Freelance writers who contract with Riverside to create passages that meet criteria established by the Development team. Many have published articles and books for youth.

Item Writers: Freelance writers who contract with Riverside to create stand-alone, prompt/scenario driven items, or passage-driven items that meet specifications provided by the Development team. Writers are required to have applicable classroom teaching experience and undergo a formal training program at Riverside before receiving assignments for their content area and grade level of expertise.

Test Development Specialists (TDS): Full-time employees of Riverside with specialized training in assessment development. In addition to being knowledgeable about item and test construction, TDSs have classroom teaching or other education-related experience.

Senior Test Development Specialists (TDS Senior): Advanced-level Test Development Specialists with proven skills in ensuring that items, passages, and assessments meet stringent quality standards. The Senior TDS specialist possesses extensive experience in item development and test form review, typically on a variety of assessment programs.

Copyeditors: Copyeditors correct errors in grammar, syntax, spelling, punctuation, style, and formatting. Copyeditors also check for problems with logic, clarity, conciseness, consistency, factual accuracy, clueing, and bias and sensitivity.

Senior Copyeditors: More experienced copyeditors review edits requested at the initial editorial stage to ensure that all necessary corrections and considerations for context have occurred.

Production Editors: Production editors proofread text and art and review all items and passages for quality of formatting. Production editors also make sure that items and passages consistently adhere to the project style guide.

Evaluating Test Materials

Creating a reliable and valid assessment requires developers and editors to carefully evaluate test materials so that students find it relevant, interesting, and engaging but not offensive, troubling, or distracting. In an effort to achieve this delicate balance, Riverside analyzes the following key elements when developing every assessment:

- Bias and sensitivity
- Representational fairness
- Language usage
- Stereotyping
- Controversial or emotionally charged subject matter
- Historical context

Bias and Sensitivity

Bias may occur when some characteristic of an item results in differential performance for two individuals of the same ability but from different subgroups.

The presence of bias in an assessment is not only undesirable from a civil rights point of view, but also from a measurement point of view. Bias can cause inaccurate test scores, so it is essential that assessments be free of bias and present all people fairly.

Riverside is committed to ensuring that its assessments acknowledge the multicultural and multiethnic nature of society and reflect thoughtful and fair consideration of the very broad character of test users. As part of the effort to attain this goal, all Riverside tests and products are designed to avoid language, symbols, gestures, words, phrases, or examples that are generally regarded as sexist, racist, otherwise offensive, inappropriate, or negative toward any group. Throughout the development process, testing materials are reviewed for potential bias and sensitivity concerns as representational fairness, language usage, and potential stereotyping are closely examined.

Assessments should also be free of subject matter that many would consider controversial or emotionally charged. Although the presence of such material may not make the test item more difficult, it could prove distracting or unsettling to the test taker. This could cause students to fail to perform at their true ability level or could hamper their ability to take the rest of the test.

Representational Fairness

Bias may occur through over-representation or under-representation of any given subgroup. Over-representation can give a false impression of the realities of life for many members of the subgroup. Under-representation fails to acknowledge the roles and achievements of the subgroup in society.

Assessments should represent subgroups as realistically as possible and should be diverse in terms of ethnicity, age, socioeconomic background, community setting, and physical disabilities.

The testing materials should also be relevant to the life experiences of the test taker. For example, the concepts of snow and freezing winters are outside many Southern students' general frame of reference. Asking students in this geographic region questions that would require prior knowledge in order to answer the question correctly would be an example of regional bias and should be avoided.

Language Usage

Language reflects underlying assumptions that will often reinforce and perpetuate bias. Great care must be given to use language that is not offensive and that has the same basic semantic and affective content for all test takers.

Stereotyping

Stereotyping is the consistent portrayal of a group using only one attribute, characteristic, or role. All groups of society should be portrayed accurately and fairly without reference to stereotypes or traditional roles regarding gender, age, race, ethnicity, religion, physical ability, or geographic setting. Presentations of cultural or ethnic differences should neither explicitly nor implicitly rely on stereotypes nor make moral judgments. All group members should be portrayed as exhibiting a full range of emotions, occupations, activities, and roles across the range of community settings and socioeconomic classes. No one group should be characterized by any particular attribute or demographic characteristic.

Assessments should not contain materials that present any subgroup's concerns as trivial, humorous, or unimportant. All populations should be depicted as having a desire to work and contribute to society. People should be presented as individuals in terms of appearance, attitudes, food preferences, lifestyle choices, abilities, and opportunities.

Controversial or Emotionally Charged Subject Matter

The following is a list of topics generally considered to be controversial or emotionally charged. These topics should be avoided whenever possible in a testing situation unless the subject matter is directly relevant to the curriculum area being assessed.

- Abortion
- Creatures that are considered scary or dirty (e.g., scorpions, rats, and roaches)
- Death and disease
- Disrespectful or criminal behavior
- Evolution
- Gambling
- High-priced consumer goods
- Magic, witchcraft, and the supernatural

- Personal appearance (e.g., height and weight)
- Politics
- Religion
- Social problems (e.g., child abuse, animal abuse, and addiction)
- Unemployment
- Unsafe situations
- Weapons and violence

Questions that require students to reveal personal information about themselves or their home life should be avoided. These questions can make test takers feel very uncomfortable and responses to these questions are typically difficult to score.

Historical Context

When developing testing materials dealing with historical information, **accuracy** should be the goal. Present-day values should not be imposed on the past.

Riverside trains freelance passage and item writers to avoid passage and item selection that introduces elements that could negatively affect a student's performance on the assessment. In addition, Riverside includes specific entries on passage, item, and form checklists to ensure that every developer and editor review all material in an effort to make the assessment as fair, reliable, and valid as possible.

Guidelines for Passage Development

For any prompt-based assessment, one of the earliest steps is locating or writing a passage that will support the types of items necessary in the assessment. Riverside has developed a set of guidelines for passages, and it trains all of its freelance and full-time employees to follow these guidelines.

Sources

Passages may be commissioned, or they may be drawn from published sources. Passages should **not** be drawn from a current textbook series or widely used anthologies. Passages from well-known works also should be avoided. Lesser-known works by well-known authors are acceptable. If drawing from published works, material should come from a variety of sources, including books and magazines.

Prior Knowledge

Students should be answering comprehension questions because they understand what they have read, not because they already know the subject. Topics that would have been studied in school by a certain grade level or that would be familiar to students from family vacations are not good candidates.

Suitability for Generating Test Items

Passages should be rich enough to yield a variety of questions measuring students' ability to construct meaning.

Quality

Passages should be well written and factually accurate.

Interest and Appropriateness

Passages should be of high interest to students at the target grade. They should also represent a wide range of interest areas. For example, not all science passages should be drawn from biology. Topics that will appeal to only one gender should be avoided. Passages should be appropriate to grade level; professional judgment and information from such sources as *Children's Books in Print* should be used to help decide this. Appropriate readability formulas should augment this judgment.

Biographies

The type of biography that simply lists a selection of important events from a person's life should be avoided. Biographies should provide a special note of human interest about the person profiled or a unique perspective on the person's life. Also, biographies of living persons should be avoided.

Gender/Ethnic Balance

The subject matter and authorship of passages should represent both genders and a variety of ethnic and cultural groups.

Stereotypes

Content should show males and females, various ethnic and cultural groups, the disabled, and the elderly in a variety of positive roles. Riverside does not publish any material that reinforces a negative stereotype (for example, elderly persons who suffer from physical deterioration, unemployed members of a particular group, or people from other cultures living in “primitive” conditions). Elderly people can be depicted as aging without using harmful stereotypes. Writers should avoid reinforcing stereotypes that attribute human characteristics exclusively to one group or another (for example, women as nurturing, men as active problem solvers). References or labels that tend to demean, stereotype, or patronize any persons engaged in any particular occupation or vocation, whether essentially entrepreneur, management, or labor, should not be used. Passages that break stereotypes are good (for example, a story about a dad who bakes bread). Passages should be balanced (males and females as heroes, in different professions, and in different roles).

Negative or Sensitive Material

Riverside does not publish negative or sensitive material. Negative material includes, but is not limited to, parents quarreling and children mistreating each other, disobeying their parents, or generally showing disrespect for authority. Sensitive material includes references to Satanism, paganism, parapsychology, magic, extraterrestrials, Halloween, ghosts, witches, and similar subjects, even in a fantasy context. Pumpkins and masks have become tainted by their association with Halloween and should be avoided. References to gambling are not acceptable. Topics dealing with nudity or implied nudity, pregnancy, and giving birth to animals or people should be avoided, as should topics related to controversial styles of music, such as rap or rock-and-roll.

Politics and Religion

References to religion and to controversial political issues (gun control, for example) should be avoided. Passages should not make even an incidental reference to the religion of the character. Passages about birthday or religious holiday celebrations (including Thanksgiving) cannot be used. References to religious traditions (animal gods, plants that have a spirit, etc.) in Native American materials should be avoided.

Evolution

Riverside does not publish any discussion of evolution and the origins of the universe. Dinosaurs and fossils can be referenced in passages (e.g., an article about finding dinosaur bones or fossils) as long as the passage does not imply a belief in evolution as a fact or other inappropriate generalizations.

Attitudes, Beliefs, and Values

Passages should not suggest that Riverside is delving into attitudes, beliefs, or values or advocating a particular value that may not be embraced by all. The content of fables, which sometimes end on a cynical note or have a pragmatic moral that some people would find offensive, should be carefully monitored. Situation ethics, secular humanism, and anything that smacks of “new age” religion are taboo. Topics that involve attitudes, beliefs, or value systems which are essential in the legal, economic, and social structure of society and to the personal and social responsibilities of citizens in society (i.e., democratic values) are appropriate.

Nonstandard English

Passages containing nonstandard English (slang, dialect, trendy expressions, etc.) should be avoided. Also, topics that use words or phrases that are likely to have negative connotations (such as “grass”) should not be used.

Socioeconomic Status

Passages should include children from a variety of socioeconomic backgrounds. Content should not suggest that everyone lives in a house with a sprawling lawn or that the average teenager can afford to buy a CD player or a brand-new 10-speed bicycle. Vacations in faraway places, ski trips, and other upper- or middle-class amenities are out of the reach of many children and do not make good subject matter for test passages.

Healthful Habits

Content should model healthful personal habits. Riverside does not use passages that make even passing reference to someone’s smoking or drinking unless the reference can be cut. Passages should not refer to the consumption of junk food. References to coffee and/or tea should be used cautiously, and references to even common drugs such as aspirin should be avoided.

Safety Practices

Content should demonstrate a concern for good safety practices. Riverside does not use passages that discuss children keeping dangerous animals or that describe children doing dangerous things, no matter how good the moral of the story is.

Obsolete Content

Particularly in science, content should not be outdated or be likely to become outdated in the next few years.

Branded Products

Riverside does not use passages about the development of a branded product (say, the invention of Coca-Cola), nor should there be references to brand names in passages. Sometimes, permission to cut a passing reference to a brand name can be obtained.

Regional Bias

Content should reflect all regions of the United States and, where possible, other countries, too.

Unfair Advantage

Passages should not contain content that would provide unfair advantage to certain subgroups.

Passage Readability Requirements

Passages must be written with readability in mind. Flesch-Kincaid readability levels are calculated for all passages. All passages also receive a certified Lexile Measures, as well. Lexiles are the only readability measure that uses the Rausch psychometric model to assign a level to text. Certified Lexile measures provide value to educators in matching text levels to reader levels.

Genre

The *Assess2Know* item banks must contain a wide variety of genres across expository and narrative text. Some examples of passage subgenres include biography, autobiography, historical fiction, mystery, science fiction, poetry, essay, functional text, plays, and folktales. In addition, single, paired, and triplet (where appropriate) passages can be found in the item banks.

Review Checklist for Passages

Passage Title:

Task	OK/Edits	Comments
Ensure that passage content is appropriate for target audience.		
Ensure that passage is clearly written for target audience.		
Ensure that passage length is appropriate for target audience.		
Ensure that passage is grammatically and structurally sound.		
Ensure that headings, captions, and glossaries exist, where appropriate.		
Ensure that vocabulary is appropriate for target audience and that difficult vocabulary is defined in context.		
Ensure that passage is free of bias and sensitivity issues.		
Ensure that author sources have been listed and used appropriately (if any FACT in an informational passage cannot be confirmed from sources given, passage should be returned to author for clarification).		
Ensure that factual information used within a passage is highlighted for content editor. For example, when passage writer provides a source, writer should list page numbers and paragraphs referred to in passage. If copy of entire resource is provided, TDS should highlight relevant section for Copyeditor. If this information is not provided, passage may be returned to writer.		
Ensure that passage does not include information subject to change in a short period of time.		
Ensure that passage has enough information to yield a minimum of eight substantially different items.		
Ensure that passage is free of misspellings or typos.		

Guidelines for Item Development

As with the development or location of suitable passages, Riverside trains freelance item writers and its Test Development Specialists in good item writing practices. To make Riverside's expectations clear, the information below is distributed to and reviewed with every freelance writer, and Test Development Specialists refer to the guidelines when reviewing submitted items.

This section outlines the requirements for items to be included within the *Assess2Know* item bank. In order for items to be accepted, all of these requirements must be followed. Riverside reserves the right to accept or reject items at the discretion of its Development team.

General Item Requirements

Items that are developed for the item bank will be both stand-alone and passage driven. Items written covering many language arts standards and mathematics standards should stand alone and not rely on information from a prompt to answer the question. Items written for reading should be passage driven; they should rely on information from the passage to answer the question.

All items will be selected response/multiple choice. A selected response/multiple-choice item is one for which students select a response from a group of three or four possible answer options.

Stand-alone Item Requirements

All items developed for the content area of mathematics are stand-alone items. Stand-alone items are not linked to any other items or stimulus material. For example, if an item requires information stored in a map, no other items may use the information in that map.

Prompt/Scenario Item Requirements

Items developed for the content area of science will either be stand-alone items or prompt/scenario driven items. A prompt from which the items should be developed will be provided. Items are developed in groups, with special attention to clueing. Each prompt will have a set of associated items that assess a variety of science standards.

Passage-Driven Item Requirements

Most items developed for the content area of reading are passage-driven items. A passage will be provided from which the items should be developed. Items are developed in groups, with special attention to clueing. Each passage should have a minimum number of items based on the grade level assessed (i.e., passages at grade three would generally have fewer associated items than passages at grade eight).

Selected-response items

Selected-response (multiple-choice) items written for Riverside assessments must possess the following characteristics:

- Contain a stem (question) with three to four answer options
- Measure what it is intended to measure
- Do not include grammatical errors, incorrect spelling, or slang that may distract students from answering the question
- Are clearly and simply written. Avoid trick questions that require students to figure out a particular angle or hidden agenda on the part of the item writer.
- Contain vocabulary appropriate for the grade level being tested
- Function independently; the answer to one item is not required as a condition for answering subsequent items
- Do not provide information that may be used to answer other items
- Have a consistent format
- State things in the positive whenever possible. Negative statements in the stem should be avoided.
- Use accurate, high quality maps, graphs, and pictures
- Are free of bias and sensitivity issues
- Avoid information that is subject to change in a short period of time
- Require students to read the passage in order to answer the question (for reading tests)
- Are interesting and engaging to students
- Have only one correct answer
- Have answer options that are plausible and attractive distractors
- Do not use “all of the above” or “none of the above” as answer options
- Present answer options in a logical order (e.g., numerical, alphabetical, sequential, etc.) whenever possible. Options should be parallel in structure and length.
- Have answer options that grammatically match the stem

Prompt/Scenario Requirements

- Prompt/Scenarios should be interesting to students
- Prompt/Scenario should not be excessive in length for the intended grade.
- Prompt/Scenario should contain all context necessary to answer all associated items. No new context should be introduced in associated items.
- Associated items should not be able to be answered without reading and understanding the prompt/scenario.
- Associated items should be short/concise/direct.
- Associated items should assess standards from several different science areas (life science, physical science, earth science).

Selected Response Characteristics

Selected-response items are composed of the stem (question) and distractors (answer choices).

Three considerations should be given when creating stems for selected-response items:

1. **Placement of questions:** Any information needed to answer the question should be given before the question is asked. This information might include charts, graphs, art, and background information. The question should be the last part of the stem.
2. **Negatives:** Stems, whenever possible, should ask positive questions rather than using the word “not” or other negatives. For example, the question “Which of the following is not a right guaranteed to all U.S. citizens?” would be restated as “Which of the following is a right guaranteed to all U.S. citizens?”
3. **Use of “most likely” instead of “always”:** “Always” is an absolute; very few answers to an item will be correct always. “Most likely” requires that the student choose the best answer from the given possibilities.

Items should have three to four distractors or answer choices. Items for grades 1 and 2 should have three distractors; items for grades 3 through 10 should have four distractors. The following considerations should be made when creating all distractors:

1. **Parallelism:** One distractor should not stand out in comparison to the others. All construction should be parallel, including the verb usage, sentence or phrase structure, and length.
2. **Numbers:** Distractors should be in ascending or descending order, typically with the same number of decimal places. An effort should be made to balance parallelism with numbers that result from common errors.
3. **Word Repetition in Stem and Distractor:** If a content word appears in the stem, it either should not appear at all in the distractors or should appear in at least two of the distractors. When a word appears in the stem and only one distractor, it makes that distractor a more attractive choice.

In addition, selected response items should be designed so that they do not clue other items. Students should not be able to discern the answer to one item based on the information in another. Also, cloned items—items that are almost identical to other items except for minor changes—will not be accepted.

Readability Requirements

Items for Riverside assessments must be written with readability in mind. In addition, vocabulary must be appropriate for the grade level being tested. Riverside uses several sources to determine the readability of items, including:

1. *Children's Writer's Word Book*
(Mogilner, Alijandra. 1992. Cincinnati, OH: Writer's Digest Books.
ISBN #: 0-89879-511-7)
2. *EDL Core Vocabularies: Reading, Mathematics, Science and Social Studies*
(1989. Austin, TX: Steck-Vaughn-EDL. ISBN#: 1-55855811-X)

Bloom's Taxonomy

All items are assigned a Bloom's Taxonomy level. Bloom's Taxonomy identifies six levels within the cognitive domain, from the simple recall or recognition of facts at the lowest level, through increasingly more complex and abstract mental levels at the highest level, which is classified as evaluation.

Cognitive Difficulty

All items are assigned a cognitive difficulty level. Cognitive difficulty levels provide a means for comparing the cognitive complexity levels of assessment items to the standards being measured. Review of student performance by standard across each of the cognitive difficulty levels can also provide valuable information for targeting instructional needs.

Difficulty Level Descriptors

- | | |
|----------------|---|
| Level 1 | Basic Skills: recall of information such as fact, definition, term, or simple one-step procedure. |
| Level 2 | Conceptual Understanding: includes the engagement of some cognitive processing beyond recalling or reproducing a response. A conceptual understanding item requires students to make some decisions as to how to approach the problem or activity and may imply more than a single step. |
| Level 3 | Extended Reasoning: requires problem solving, planning, and/or using evidence. Items require students to develop a strategy to connect and relate ideas in order to solve the problem while using multiple steps and drawing upon a variety of skills. |

Review Checklist for Multiple-Choice Items

Content	Does the item clearly match the identified standard?	
	Does the item measure understanding, insight, or application of a principle rather than rote memory or factual recall?	
	Is the item free of stereotyping, bias, or advantage directed at a particular age group, gender, ethnic group, religious group, socioeconomic group, or geographic region?	
	Does the item avoid any information that may be distracting to the student? Are value statements and judgments avoided?	
	Are the vocabulary and syntax appropriate for the students at the intended grade?	
	Does the item include necessary art specifications and examples and is necessary source documentation provided?	
	Is the context based upon real-world, factual information? Are necessary sources attached to submission?	
Stem	Is the stem written without lengthy sentences?	
	Is the item clearly and concisely written? Does the stem clearly present the student with a problem to solve or a task to do?	
	Will the student clearly understand what he or she is being asked to do?	
	Is the stem direct and straightforward rather than “cute” or “tricky”?	
	Does the stem avoid the repetition of words in each of the answer options?	
Answer Options	Does the item have only one correct answer?	
	Do the distractors/options reflect incorrect knowledge or common misconceptions or errors students make? Are the rationales for these errors provided on the template?	
	Are all the incorrect options attractive and plausible to students who do not know the correct answer?	
	Are the answer options mutually exclusive?	
	Are all the answer options parallel in form?	
	Is the item free from grammatical clues or correct answers that are longer or more specific than the other answer options?	
	Are all the answer options free of repeated, extraneous words?	

TDS and Senior TDA Item Review Checklist

General Item Criteria	OK or document exceptions
Each item should assess content standard(s) and/or benchmark(s) as specified in the test blueprint or assessment frameworks.	
Items must measure appropriate thinking skills as specified in the test blueprint or assessment frameworks.	
Items should be grade-level appropriate.	
Items should be written at appropriate reading levels (vocabulary) according to the item specification guidelines. Reference the <i>EDL Core Word Book</i> and/or the <i>Children's Writer's Word Book</i> .	
Items should be written at appropriate cognitive levels according to the item specification guidelines. (Depth of Knowledge)	
Bloom's Level and Cognitive Difficulty Level should be recorded in CMS for each item.	
Items should include accurate and sufficient content information.	
Items should be clear, concise, and complete.	
Items should have correct punctuation and be grammatically correct.	
Items should not contain typos and should have all words spelled correctly.	
Items should stand-alone; they should not be based on the content of other items.	
Clueing should be avoided within an item and between items.	
Items should not contain information, instructions, or stimuli that are biased based on gender, ethnicity, race, social class, disability, geography, etc.	
Items must be fair to all students (e.g., avoid items that trick, entertain, disgust, or insult students).	
Questions should be formatted as complete sentences rather than incomplete statements.	
Graphics, formulas, diagrams, and figures should be legible, clean, necessary, relevant, accurate, and unbiased and should adhere to the style guide.	
At least two reliable sources for factual information in items and passages should be included.	
Fact checking should have been completed BEFORE item was accepted from writer.	
Items should not contain information, instructions, or stimuli that are biased based on gender, ethnicity, race, social class, disability, geography, etc. Avoid the use of "true", "creates", and "believes" in ways that suggest that science is finite or a system of beliefs.	

Multiple-Choice Item Criteria	OK or document exceptions
Each multiple-choice question must have only one correct answer that is indicated in CMS.	
Each multiple-choice question must have several incorrect unique distractors containing common errors students would make.	
Distractors should be in a logical order, according to the project style guide.	
Make certain that the distractors and the answer are comparable in length, complexity, and grammatical form.	
Item must have plausible/attractive distractors.	
Make certain that none of the distractors is a possible correct answer.	
Avoid overlapping distractors.	
Avoid negative distractors following a negative stem (double negatives).	
Avoid items having absolutes, “none of the above,” and “all of the above” as distractors.	
Avoid clues in the distractors.	
Restrict the use of negatives (e.g., “not,” “none”) in the multiple-choice stem.	
Reading items must avoid repeating words from the stem in the answer or distractors.	
Stems on multiple-choice items should pose a single problem, although the solution to that problem may require more than one step.	
Only necessary words (1 to 2 at most) should be emphasized in the stem.	



Riverside Publishing

A HOUGHTON MIFFLIN COMPANY

www.riversidepublishing.com

800.323.9540