



A State Collaborative on Assessment and Student Standards

QUALITY CONTROL CHECKLIST

FOR

ITEM DEVELOPMENT AND TEST FORM CONSTRUCTION

July 2003

This quality control checklist has been developed for state personnel to review developed test items and test forms submitted by vendors or item development committees. This checklist can also be used to train new office assessment staff, and item content and bias review committees. Because each state has its own unique test design and assessment plan, states are encouraged to modify this checklist to meet their needs. The ultimate goal is to ensure that all test forms used in large-scale assessment programs meet the content specifications set forth in the test blueprints and guiding documents, as well as meet professional and psychometric standards for excellence.

Regular development procedures between states and vendors regarding item development, item content and bias review, field test, and test form development are not the focus of this document. Most states include that information in the technical reports.

For more information on the TILSA SCASS, or to provide feedback on this Quality Control Checklist, please contact Arthur Halbrook at arthurh@ccsso.org.



Council of Chief State School Officers

Use of the information in this document is permitted with appropriate attribution to the CCSSO TILSA SCASS

Development and Review of Test Items

A. General Guidelines

- | | |
|--------------------------|---|
| <input type="checkbox"/> | 1A. Each item should assess content standard(s) and/or benchmark(s) as specified in the test blueprint or assessment frameworks. |
| <input type="checkbox"/> | 2A. Items must measure appropriate thinking skills as specified in the test blueprint or assessment frameworks. |
| <input type="checkbox"/> | 3A. Items should be written at appropriate cognitive levels and reading levels according to the item specification guidelines. |
| <input type="checkbox"/> | Items should be grade-level appropriate. |
| <input type="checkbox"/> | Items should include accurate content information. |
| <input type="checkbox"/> | Items should not contain information, instructions, or stimuli that are biased based on gender, ethnicity, race, social class, disability, geography, etc. |
| <input type="checkbox"/> | Items should not require students to learn new knowledge beyond assessment frameworks to respond correctly. |
| <input type="checkbox"/> | Items should be clear, concise, complete, and grammatically correct. |
| <input type="checkbox"/> | Cueing should be avoided within an item and between items. |
| <input type="checkbox"/> | Items should not assess opinions (e.g., avoid items of the form "What would you do?"). |
| <input type="checkbox"/> | Items must be fair to all students (e.g., avoid items that trick, entertain, disgust, or insult students). |
| <input type="checkbox"/> | Items that require students to estimate or round should contain appropriate qualifying modifiers (e.g., "most reasonable," "most accurate," "closest"). |
| <input type="checkbox"/> | Items should adhere to the style guide approved by the individual departments of education. |
| <input type="checkbox"/> | Graphics, formulas, diagrams, and figures should be legible, clean, necessary, relevant, accurate, and unbiased. |
| <input type="checkbox"/> | Items should be reviewed and approved by the content review committee and the bias/sensitivity review committee before field testing. |
| <input type="checkbox"/> | The state should require the vendor to have a professional editor review items before field testing and verify copyright for passages and stimuli (e.g., charts, maps). |

B. Specific Guidelines for Multiple-Choice Items

- | | |
|--------------------------|--|
| <input type="checkbox"/> | Stems on multiple-choice items should pose a single problem, although the solution to that problem may require more than one step. |
| <input type="checkbox"/> | For passage-based items, avoid stems that allow students to determine the correct response without reading the passage. |
| <input type="checkbox"/> | Restrict the use of negatives (e.g., "not," "none") in the multiple-choice stem. |

- Avoid negative distractors following a negative stem (double negatives).
- Avoid repeating words from the stem in the answer.
- Distractors on multiple-choice questions must include only one correct answer and several (e.g., three) distractors containing common errors students would make.
- Distractors should be in a logical order.
- Avoid items having “none of the above” and “all of the above” as distractors.
- Distractors must be grammatically responsive to the stem.
- Avoid superfluous information in the distractors.
- Avoid clues in the distractors.
- Make certain distractors are comparable in length, complexity, and grammatical form to the answer.
- Make certain none of the distractors is a possible answer.
- Avoid overlapping distractors.

C. Specific Guidelines for Constructed-Response Items

- Items should clearly state how the student should respond (e.g., “include details and examples to support your answer”).
- Items should allow for a variety of acceptable responses.
- Constructed-response stems should use direct, concise sentence structure.
- Constructed-response items should allow students enough time (varies depending on item type) to respond and enough space in the answer document to accommodate the response.
- Extended constructed-response items should conform to the test blueprint (or assessment frameworks) and item development specifications.
- Extended constructed-response items should include an entry point that is accessible to most students and if scaffolded, should include more difficult tasks in the latter part of question.
- Scenarios that involve gender or ethnic bias should not be used.
- Scoring rubrics should measure the response as objectively as possible.
- Scoring rubrics should be defensible.

Test Form Construction

A. General Guidelines

The first step in constructing a test form is selecting items from the item bank. Selected items can be categorized as anchor items (also called linking items or common items) and non-anchor items. The general guidelines should be applied to all selected items.

- All selected items should measure the content standards/benchmarks as specified in the test blueprint or assessment frameworks.
- Items selected to use in test forms should not have a poor model fit (or misfit) or unacceptable item statistics. If it is necessary to use an item with misfit, the content staff should first survey content specialists to determine whether the item is acceptable.
- Items should cover a wide variety of questions, and similar items on each test form should be minimized.
- Items should not have flaws or have more than one correct answer choice.
- Items should follow psychometric guidelines agreed upon by the state and the vendor (e.g., p-values are between .2 and .9, and item-total correlation coefficients are greater than .2).
- The test characteristic curve (TCC) for the complete item selection should be as similar as possible to the curve from the previous target forms.
- Items showing DIF for or against all possible subgroups (e.g., gender, ethnicity, Limited English Proficiency) should be avoided. If an item of this nature is selected, the content staff should seek input from different groups and assess the impact if this item is acceptable.
- Avoid items that were rejected by the item content review committees.
- Do not use items that were rejected by the item bias review committees.
- Limit the items with the words "best" or "most" and negative words (no, not, none) as much as possible.
- Do not use items previously released to the public.
- All the selected items should be reviewed and approved by the appropriate assessment staff and psychometrician.

B. Specific Guidelines for Anchor Item Selection

Select items from previous test forms to serve as anchor items. Anchor items are used to equate test forms, helping maintain a stable scale across test administrations.

- Anchor items should be selected from previous operational test forms.
- Anchor items should not be used in consecutive years if possible (e.g., avoid using the same item in spring 2000 and spring 2001).
- At least 14 score points are needed for the anchor item set.

- Anchor items should not be edited. If even a minor change is needed for an anchor item, this item should be removed from the anchor item set.
- Anchor items should cover a range of difficulty levels but not be extremely easy or difficult.
- Anchor items should be as representative as possible of the full-length test forms.
- Anchor items should be in a similar position to that they occupied on the previous operational test forms.
- The test characteristics curve (TCC) for anchor items should be as similar as possible to the curve from the previous target forms.

C. Specific Guidelines for Non-Anchor Item Selection

Once the anchor item set is completed, the remainder of the test items (non-anchor items) with acceptable statistical properties and content-standard constraints should be selected.

- The complete item set should match the test blueprint (or specification).
- Non-anchor items should be selected from the entire field-test item pool.
- Non-anchor items should be in a similar position to that they occupied on the field test forms.
- Minor changes (formatting, style, grammar) are permissible for non-anchor items. However, if a change may have an impact on students' performance on that item, the item should be removed from the item selection set and be field tested again.
- Items should be reviewed carefully to meet the guidelines before the item selection is completed.
- The content staff must review each item carefully to make certain that 1) the content code is correct, 2) there is no flaw in the item, 3) the answer key is accurate, and 4) the distractors are fair and are not possible answers.
- Constructed-response items should not be overly difficult. If there are more than one constructed-response item on the test form, consider placing the easiest one first.
- Constructed-response items in the same form may represent a range of difficulty and a variety of content standards.

Test Form Development

General Guidelines

After items have been selected and reviewed for both psychometric excellence and content/technical quality, the items can be used to develop new test forms.

- Each test form should include an adequate coverage, as required by the test blueprint.
- No item in a form should provide clues to the answer for any other item on that test form.
- Forms should target a balanced representation of gender and ethnic groups in terms of illustrations and names.
- Forms should include a wide range of topics and a variety of questions.
- Correct answers in multiple-choice items should be evenly distributed (i.e., avoid having all correct answers as the third of four choices).
- Forms should not include any items that have been released to the public.
- The vendor should provide test maps that include item ID, previous form location, p-value, item-total correlation coefficient, content code that includes standard and benchmark, short descriptor, etc. The content staff must ensure that all information on the test maps or form planners is accurate and matches the item cards.
- Each item on the new test form must be verified word by word on the new test form against the final item cards after the field test data review meeting.
- Each item must be verified to ensure the accuracy of the item, the answer key, the instructions to the students preceding the item, and the format.
- The assessment staff must ensure that the test form meets the test blueprint and the test maps reflect correct information.
- The editor must review formatting, directions, graphics, etc., for adherence to the approved style guide.
- All fonts should be clearly legible (i.e., at least 12-point type for regular test forms and 18-point type for large-print forms).
- The editor should ensure a consistent format across test forms for all content areas and grade levels.
- As appropriate, wording should be consistent across test forms.
- Standard proofreading marks should be used.
- Use the style guide approved by the individual departments of education to resolve editing questions.
- All changes/edits must be reviewed by the content staff and approved by the lead psychometrician.