



# QUALITY CONTROL CHECKLIST

FOR

## PROCESSING, SCORING, AND REPORTING

*January 2003*

This Checklist was developed based on contributions from several states that are members of the CCSSO Technical Issues in Large-Scale Assessment SCASS. The project was chaired by Fen Chou, Louisiana Department of Education, who compiled the state's experiences and procedures in quality control of their assessment programs. This checklist is appropriate for use with high-stakes tests as well as most other types of assessments. It is important to note that states are responsible for overall quality assurance of their assessment systems, and vendors should meet the quality control requirements of the state.

This checklist focuses on the quality control procedures for processing, scoring, and reporting. Some, but not all, quality control procedures for test form development are included in order to complete the steps for Criteria 1. This checklist does not address other quality control processes (e.g., other test form development issues, demographic edits, and security checks) that may be handled at the same time while processing, scoring, and reporting tasks are being conducted.

This checklist is based on the assumption that test items are developed according to an assessment framework and other test design specifications. All items should measure knowledge and skills indicated in the content standards document. During item development, the curriculum and assessment staff in each content area should ensure that each developed item measures one or more content standards/benchmarks (or objectives) and verify the accuracy of item-to-standard codes.

This checklist is also based on the assumption that item selection procedures are conducted according to the test blueprint and other psychometric and content requirements. All items selected for each test form have been reviewed and approved by the content staff and the psychometric/research staff, and all items, passages, graphics, and directions have been proofed by editors.

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

**1. Test keys and items to standards/benchmarks (or objectives) maps are accurate.**

Check √	Stage/ Task Purpose	Task	Recommended Expertise *				Recommended Materials
			C	A	P	D	
<input type="checkbox"/>	<u>During test form development</u> To ensure the match of items and content standards/ benchmarks	<b>1A.</b> During test form development, test items are selected based on the test blueprints. Each selected item should align to the content standard and/or benchmark. Department staff in each content area should confirm the match of each item to the content standard/benchmark and confirm that the test as a whole matches the test blueprint and meets the content coverage requirement.	C	A	P		Content standards, test blueprints, assessment framework
<input type="checkbox"/>	<u>During test form development</u> To ensure accuracy of information on the test maps	<b>1B.</b> Test maps are created to document the information for each item selected for final test production. The information includes each item's correct response, the maximum number of score points, the item's intended standard/benchmark, and other attributes (e.g., location, field test data). Department staff should (1) verify the accuracy of all information on the test maps before they are sent to the vendor for test booklet production, and (2) ensure that all items for each test form meet the statistical requirements.	C	A	P		Test maps, test blueprints, item records, field test results, psychometric rules for test development
<input type="checkbox"/>	<u>After test form development</u> To ensure accuracy of answer keys for multiple-choice items and accuracy of information for open-ended items	<b>1C.</b> To verify the accuracy of the answer key information for multiple-choice items on the test maps, department staff should take the test for their respective content area and compare their answers to the answer key on the test map. They should also verify item-to-standard information on the test maps. To verify the accuracy of the information for open-ended items, department staff should review the items and scoring rubrics and approve the maximum score point assigned to each open-ended item, the item-to-standard classification for each item, and other item attributes.	C	A			Test maps, test items, scoring rubrics, hand scoring rules
<input type="checkbox"/>	<u>After scanning answer documents</u> To verify accuracy of answer keys and item-to-standard codes in the data file that is used to produce reports	<b>1D.</b> Using a sample of actual student response data prior to the operational scoring process, department staff should verify the accuracy of answer keys in the data file provided by the vendor. Frequency distributions of item responses or point-biserial information can be used to check multiple-choice answer keys; frequency distributions of score points for each item could be used to check students' performance on the open-ended items. Observations of unexpected results should be investigated and resolved.			P	D	Sample student responses file, test maps

\* Key for Professional Expertise Recommendations: C = Content A = Assessment P = Psychometric/Research D = Data Management

**2. Student responses are completely and accurately scanned and captured.**

Check √	Stage/ Task Purpose	Task	Recommended Expertise*			Recommended Materials	
<input type="checkbox"/>	<u>Before scanning answer documents</u> To ensure the vendor's scanning system works appropriately and accurately	<b>2A.</b> Prior to the scanning and scoring of actual response documents, the Department should require the vendor to provide a "test deck" file and a copy of the "test deck" answer documents. The test deck procedures involve processing hand-bubbled answer documents with a variety of scenarios through the scanning process allowing simulated student test data to be processed and scored. Department staff should compare the information on the test deck file with the bubbles on the simulated answer documents and review the reports provided by the vendor.			<b>P</b>	<b>D</b>	Test deck file, test deck reports, simulated answer documents
<input type="checkbox"/>	<u>Before scanning answer documents</u> To ensure the vendor's imaging system works appropriately and accurately	<b>2B.</b> Prior to the scanning and scoring of actual response documents, the Department should require the vendor to (1) demonstrate that student responses are captured completely and are readable on-screen and when printed, (2) specify procedures for re-scanning documents if images are determined to be unreadable during scoring, and (3) demonstrate that individual response images distributed for scoring are accurately linked to the correct student record.		<b>A</b>			Quality assurance report on the imaging system
<input type="checkbox"/>	<u>During the scanning process</u> To ensure ALL student responses are scanned and captured	<b>2C.</b> The Department should require the vendor to provide information regarding the status of the test documents throughout the scoring and editing process. All answer documents and school/class headers for all expected schools should be received and scanned. Actual counts based on scanned records should be compared with expected counts from the department student and school file or list. Any discrepancies should be researched and resolved. Department staff should review the documents provided by the vendor and assist with the investigation of any discrepancies.		<b>A</b>	<b>P</b>	<b>D</b>	Scanning and editing status reports
<input type="checkbox"/>	<u>After the image scanning process</u> To ensure accuracy of scanned images of responses for open-ended items	<b>2D.</b> The Department should require the vendor to deliver (1) a set of answer documents produced from scanned images and (2) photocopies of the same set of answer documents. Department staff should verify the accuracy of information on the scanned images against the photocopies of the original documents and also ensure that the student response images are matched back with the proper student's record.		<b>A</b>	<b>P</b>	<b>D</b>	Imaging materials
<input type="checkbox"/>	<u>After the scanning process</u> To ensure accuracy of scanning data for multiple-choice items	<b>2E.</b> The Department should require the vendor to deliver (1) a scanning file including scanned multiple-choice responses and (2) a copy of a set of answer documents for each test form. Department staff should verify the accuracy of information on the scanning file against the bubbles on the answer documents.		<b>A</b>	<b>P</b>	<b>D</b>	Scanning file, copies of answer documents

\* Key for Professional Expertise Recommendations: C = Content A = Assessment P = Psychometric/Research D = Data Management

**3. Student multiple-choice responses are accurately scored.**

Check √	Stage/ Task Purpose	Task	Recommended Expertise*			Recommended Materials	
<input type="checkbox"/>	<p><u>After scanning and scoring multiple-choice (MC) items</u></p> <p>To ensure accuracy of the scoring data for multiple-choice items</p>	<p><b>3A.</b> Prior to the final processing of student test results, the Department should require the vendor to provide the preliminary data file for quality control of the scoring and reporting processes. If the state file is extremely large, the vendor should provide a sample student responses data file that consists of student test data from several districts and special schools (e.g., charter and lab schools). Department staff should ensure (1) the file format and data elements match the final data file layout and specifications that have been approved by the state and (2) the scoring data match the processing rules/specifications.</p>			<b>P</b>	<b>D</b>	Preliminary data file, file layout, processing rules/specifications document
<input type="checkbox"/>	<p><u>After scoring MC items</u></p> <p>To ensure accuracy of the answer keys and item-to-standard data in the data file</p>	<p><b>3B.</b> The accuracy of answer keys and item-to-standard data for multiple-choice items in the preliminary data file should be verified. Department staff should score and check the answer keys and standard sub-score in the data file by using an independent statistical program (e.g., SAS, SPSS). The results should be compared with the information on the test maps to ensure that the answer keys and item-to-standard data are applied accurately by the vendor during the scoring process.</p>		<b>A</b>	<b>P</b>	<b>D</b>	Preliminary data file, test maps
<input type="checkbox"/>	<p><u>After scoring MC items</u></p> <p>To ensure ALL multiple-choice items on each test are captured</p>	<p><b>3C.</b> Using the preliminary data file, department staff should verify the accuracy of the score range for multiple-choice items on each content area test since different tests have different numbers of multiple-choice items. A SAS or SPSS program is a good tool for running frequency distributions of raw scores for multiple-choice items for each test. The results should be compared with the information on the test maps or test design document.</p>		<b>A</b>	<b>P</b>	<b>D</b>	Preliminary data file, test maps
<input type="checkbox"/>	<p><u>After scoring MC items</u></p> <p>To detect possible errors with multiple-choice scoring</p>	<p><b>3D.</b> The p-value or point biserials for each multiple-choice item is a good indicator to use in checking if any item has been scored incorrectly. Prior year operational test data for anchor items or field test data for non-anchor items should be compared with preliminary current year data to assess the reasonableness of current student test results.</p>		<b>A</b>	<b>P</b>	<b>D</b>	Preliminary data file, test maps, prior year data or field test data

\* Key for Professional Expertise Recommendations: C = Content A = Assessment P = Psychometric/Research D = Data Management

**4. Student open-ended responses are accurately scored pursuant to specifications (e.g., scoring rubrics).**

Check √	Stage/ Task Purpose	Task	Recommended Expertise*				Recommended Materials
<input type="checkbox"/>	<u>Before handscoring open-ended (OE) items</u> To ensure the training of scorers is conducted appropriately	4A. If state tests contain open-ended items, department staff should (1) verify the rules that the vendor uses to hire readers are followed, (2) review readers' qualifications, (3) approve the training procedures and materials that the vendor uses to train handscoring readers, and (4) oversee the training process through on-site visits, conference calls, or the review of status reports and any reference documents.	C	A			Reader qualification materials, scoring rubrics, training materials
<input type="checkbox"/>	<u>During handscoring OE items</u> To ensure the handscoring process is conducted appropriately	4B. The Department should require the vendor to deliver the reader reliability reports on a regular basis during the handscoring process. The procedures designed to establish the reliability of the read-behind papers should be applied using random sampling (invisible read-behinds). These reports should be reviewed carefully to ensure that open-ended items are scored accurately and the reliability of the readers meets the criteria. If any unusual data are observed, the Department should require the vendor to investigate the issue and resolve the problem.		A	P		Scoring rubrics, reader reliability reports
<input type="checkbox"/>	<u>After handscoring OE items</u> To ensure hand-scored data are accurate	4C. Students' scores for the open-ended items in the preliminary data file should be used to verify the reasonableness and accuracy of the scored data. Department staff should check the score range for each open-ended item and compare the results to the scoring rubrics, handscoring criteria/rules, and information on the test maps. Prior year data or field test data should be compared with preliminary current year data to assess the reasonableness of current student test results.			P	D	Scoring rubrics, hand scoring rules, test map, preliminary data file
<input type="checkbox"/>	<u>After handscoring OE items</u> To detect possible errors with hand-scored data.	4D. The percentage of students at each score point for the open-ended items is a good indicator for checking if the items are scored correctly. Prior year data or field test data should be compared with preliminary current year data to assess the reasonableness of current student test results.			P	D	Preliminary data file, prior year data or field test data

\* Key for Professional Expertise Recommendations: C = Content A = Assessment P = Psychometric/Research D = Data Management

**5. Student scored multiple-choice records and scored open-ended records are correctly matched to create a complete student record.**

Check √	Stage/ Task Purpose	Task	Recommended Expertise*			Recommended Materials
<input type="checkbox"/>	<p><u>During creating complete student records</u></p> <p>To ensure scored MC records and scored open-ended records are correctly matched</p>	<p><b>5A.</b> If a state test is composed of multiple-choice and open-ended items, the Department should require the vendor to provide status information regarding the match of scored multiple-choice records and scored open-ended records for each student. All the open-ended items for students who returned their answer documents should be scored and matched to the students' scored multiple-choice responses. Any "questionable" matches and "unmatched" student records should be examined and resolved. When necessary, department staff should check with school personnel to verify the accuracy of student records.</p>		<b>A</b>	<b>P</b>	Preliminary data file, tests, test maps
<input type="checkbox"/>	<p><u>After creating complete student records</u></p> <p>To ensure each tested student has a complete record</p>	<p><b>5B.</b> If a test is composed of multiple-choice and open-ended items, department staff should verify that all students in the preliminary data file have both sets of scores. If one set of scores is completely missing for any student, further investigation should be conducted. When necessary, the vendor must verify the accuracy of student records by reviewing the actual student test documents.</p>		<b>A</b>	<b>P</b>	<b>D</b> Preliminary data file, tests, test maps

\* Key for Professional Expertise Recommendations: C = Content A = Assessment P = Psychometric/Research D = Data Management

**6. Scoring sheets, including item parameter tables, are generated using accurately scored student response data.**

Check √	Stage/ Task Purpose	Task	Recommended Expertise*			Recommended Materials
<input type="checkbox"/>	<u>During equating and scaling</u> To ensure the equating/scaling process is accurate	<b>6A.</b> The Department should require the vendor to provide various statistical data (e.g., ICC and TCC information) for review. If any items should be deleted for the generation of scale scores, department staff should review information regarding each deletion and also ensure that item parameter tables are revised to reflect any deleted items.		A	P	Calibration data
<input type="checkbox"/>	<u>During equating and scaling</u> To ensure accuracy of the calibration sample data file	<b>6B.</b> Department staff should verify the accuracy of actual scored student response data (calibration sample) that the vendor uses to generate scoring tables. The calibration sample data file can be used to (1) verify the number of item responses (or score points) for each test against the expected number of item responses on the test maps, (2) review multiple-choice and open-ended item frequency distributions for reasonableness, and (3) review number of students achieving the highest and lowest possible score points for reasonableness. Unusual results should be investigated and resolved prior to the production of the quality check file.		A	P	Test maps, calibration data file
<input type="checkbox"/>	<u>During equating and scaling</u> To ensure accuracy of the equating process	<b>6C.</b> Department staff should check the accuracy of the equating procedures. When possible, department staff should duplicate the equating process with the same software that the vendor uses to create the scoring tables.			P	Calibration data
<input type="checkbox"/>	<u>After equating and scaling</u> To ensure accuracy of item parameters	<b>6D.</b> The Department should require the vendor to submit final item parameters generated based on the current calibration sample data on anchor items and standard errors of measurement data. These data should be reviewed for reasonableness and should not exceed expectations. Department staff should also compare the current year item parameters to the item parameters for those items generated in previous test administrations to ensure that there are no unexplained variances.		A	P	Calibration data, scoring tables, item parameters

\* Key for Professional Expertise Recommendations: C = Content A = Assessment P = Psychometric/Research D = Data Management

**7. Student raw scores are accurately converted to scaled scores.**

Check √	Stage/ Task Purpose	Task	Recommended Expertise*			Recommended Materials
<input type="checkbox"/>	<u>Before converting raw scores to scale scores</u> To verify accuracy of scoring tables	<b>7A.</b> Department staff should verify the accuracy of the information on the raw score to scale score tables. The current year scoring tables should be compared to the scoring tables for the previous test administrations. If any unusual discrepancies are observed, department staff should require the vendor to double check the equating and scaling process and also conduct a meeting with the company's psychometrician(s) to investigate possible mistakes.		<b>A</b>	<b>P</b>	Raw score to scale score table, prior year tables
<input type="checkbox"/>	<u>After converting raw scores to scale score</u> To verify accuracy of the converting process	<b>7B.</b> Department staff should recalculate student total raw scores for a sample of students using each student's scored multiple-choice responses and scored open-ended responses. Then, these total raw scores should be programmatically converted to the scale scores based on the scoring tables. Next, the results should be compared to the scale scores generated by the vendor to verify that the correct scoring tables are being used. Any discrepancies are indicative of possible violations of data integrity and should be resolved prior to generation of test results.		<b>A</b>	<b>P</b>	<b>D</b> Preliminary data file, scoring tables
<input type="checkbox"/>	<u>After converting raw scores to scale score</u> To detect possible errors with the equating data	<b>7C.</b> Department staff should conduct analyses of achievement level frequency distributions or passing/failing rates and then compare the results with prior test data to assess the reasonableness of the preliminary current test results. Any unusual differences should be investigated. Department staff should also check the score range of achievement levels (or performance standards).		<b>A</b>	<b>P</b>	<b>D</b> Preliminary data file, prior year data
<input type="checkbox"/>	<u>After converting raw scores to scale score</u> To ensure correct norms sets are applied	<b>7D.</b> Department staff should ensure that the correct norm sets are used if using the norm-referenced tests.		<b>A</b>	<b>P</b>	<b>D</b> NRT tables

\* Key for Professional Expertise Recommendations: C = Content A = Assessment P = Psychometric/Research D = Data Management

**8. Results are accurately summarized, printed, and reported at the student, school, district, and state levels.**

Check √	Stage/ Task Purpose	Task	Recommended Expertise*			Recommended Materials	
<input type="checkbox"/>	<u>Before producing final reports</u> To ensure accuracy of the data file used to produce final reports	8A. Department staff should check ALL the data elements in the preliminary data file to ensure that no data are missing, processing rules are applied correctly, and the value range for each data element is accurate.		A	P	D	Preliminary data file, file layout, processing rules/specs
<input type="checkbox"/>	<u>Before producing final reports</u> To detect possible errors with the summary reports	8B. Department staff should compare preliminary current year data with prior year data at the school, district, and state level to assess the reasonableness of the current summary test results and passing rates. Any unusual differences should be investigated.		A	P	D	Preliminary data file and sample reports, prior year reports
<input type="checkbox"/>	<u>Before producing final reports</u>	8C. Department staff should verify the accuracy of any disaggregated data before the final reports are produced.		A	P	D	Preliminary sample reports
<input type="checkbox"/>	<u>Before producing final reports</u> To verify the voiding of students' scores	8D. If a student score should be voided or converted to the lowest score due to test security violation (e.g., plagiarism, erasure analysis) or other circumstances, the student's score on the final file and reports should be checked.		A	P	D	Preliminary data file, void student list
<input type="checkbox"/>	<u>Before producing final reports</u> To verify accuracy of information on the individual student reports and summary reports	8E. Department staff should check each piece of information printed on the sample of student reports and school/district/state summary reports against the data on the data file. An electronic comparison should replace, where possible, the manual comparison to reduce the risk of human error. PC-based statistical software (e.g., SAS and SPSS) can be used to produce the same data as shown on the reports. The data from the SAS program should match the reports produced by the vendor. Any discrepancy may be indicative of a possible violation of data integrity and should be resolved immediately prior to generation of final test reports.		A	P	D	Preliminary data file, preliminary sample reports
<input type="checkbox"/>	<u>After producing final reports</u> To ensure the printing quality for the final reports	8F. Department staff should conduct final reviews of variance reports between the time the reports are printed and released for shipment to ensure the printing quality and the accuracy of student, school, district, and state reports, including all test scores, text, and titles.		A	P		Final reports

\* Key for Professional Expertise Recommendations: C = Content A = Assessment P = Psychometric/Research D = Data Management