

Issues and Technical Considerations Related to Transitioning to a New Test According to Common Core State Standards (CCSS)

(MARCES Team)

Issues	Technical Considerations
<p>Psychometric Related</p> <ul style="list-style-type: none">• Scaling<ul style="list-style-type: none">• <i>Item parameter drift</i>: Item parameters (e.g., item difficulty and discrimination) do not remain invariant over occasions due to factors other than sampling error. When invariance does not hold, the item is considered to be drifting from its original parameter value.• <i>Item parameter shift</i>: When a state changes its tests from paper-and-pencil tests to computer-based tests, item parameters may shift due to mode effect.• <i>Item parameter change</i>: When items are being re-aligned to another grade, item parameters will change from previous estimated values obtained from a different grade’s test (population).• <i>Score scale drift</i>: Scale drift is a shift in the meaning of score scale that alters the interpretation which can be attached to score points along the scale.• <i>Score scale shift</i>: Score scale shift may occur when states develop or modify the current assessments in the transition to CCSS for the next generation assessments.<p>Source: https://ccsso.confex.com/ccsso/2013/webprogram/Session3598.html</p>• Linking<p>The CCSS test could be treated as a new test without any connections with the original test. In this case,</p><ul style="list-style-type: none">• The continuity may be lost from the original test/assessment• A new scale and new cut scores are needed for reporting test scores and student performance<p>Source:</p>	<ul style="list-style-type: none">▪ Use the Validity Framework by Embretson (2007) as a guideline to collect evidence through theoretical and empirical analyses.▪ Decisions should be made based on validity evidence, state policy, and other logistical considerations.▪ Extreme caution must be taken for cases that use a vertical scale, change testing mode, and/or switch from a linear test to an adaptive test.▪ Establish statistical relationships between the state standards-based assessment and the CCSS-based assessments through the approach of linkage.▪ Revisiting performance standards is necessary. Examine the impact on student performance by using operational data and then determine to reset or modify performance standards or cut scores. <ul style="list-style-type: none">▪ Through statistical linkage, the relationships between the old and the new tests could be established for continuity and comparisons:<ul style="list-style-type: none">▪ Equating▪ Prediction▪ Scale alignments▪ Other methods: projection, calibration, moderation

<https://ccsso.confex.com/ccsso/2013/webprogram/Session3598.html>

- DIF issues across states
- Unidimensionality in the newly developed items with various forms
- How to decide on cut scores for CAT data

- New and innovative item types:
 - Items assessing complex skills
 - Psychometric issues with the new item types
 - Challenge: Next generation items involve complex student interactions moving beyond conventional response types
 - How to train content experts to write innovative items?

Source:

<https://ccsso.confex.com/ccsso/2013/webprogram/Session3546.html>

- If issues present themselves; statistical analysis will be employed.
- If issues present themselves; statistical analysis will be employed.
- See MARCES's summary on standard setting and cut score determination.
- Collaborate with other educators to better understand new standards and how to create innovative test questions to assess the new standards.
- Utilize the latest technology to create new items in an intuitive interface without necessitating programming skills.
- Review processes can be developed so that items are reviewed by experienced curriculum staff before the items are added to a form.

Technology Related

- Technology readiness:
Funding, bandwidth, devices are biggest concerns
At least three states have submitted funding requests to their legislature
Mixed interest in AI scoring; states realize its importance to cost reduction
- AI/machine scoring of student essays is possible, but contraverial
- Technology enhanced items
- Test Security

Source:

<https://ccsso.confex.com/ccsso/2013/webprogram/Session3482.html>

- Not sure of the state capacity and state needs.
- State needs to evaluate the technology readiness across the Districts.

Implementation Related

- Test delivery

- Deliver items on any supported device with the same level of fidelity. Allow for a test question (including technology enhanced

- Scoring

- Reporting

Source:

<https://ccsso.confex.com/ccsso/2013/webprogram/Session3546.html>

questions) to be written just once and subsequently delivered on multiple devices and in varying testing scenarios.

- Enhance automated scoring to include support for the innovative items as well as standard item types.
- Provide human scoring for new items that are too complex to be rated by using automated scoring.
- Student roster data can be made available as students submit their tests. Summary data can be made available at the end of the testing window.
- Item level data can also be prepared easily.
- Test results need to be reported in a timely manner.

Policy Related

- Student growth and trend
- Teacher and school effectiveness
- State proficiency levels

- Statistical and psychometric issues need to be resolved.
-