

Equating the 2002 Maryland High School Assessments

Calibration and equating was performed only at the total test level. The Maryland High School Assessments have never been equated at the goal/expectation level, so the relative difficulty of the individual goals/expectations may vary substantially from one form to another.

The following procedures that were used to calibrate/equate/scale the January and May 2002 test forms in each content area:

1. January: Within each content area, items on all of the January test forms were calibrated together, and equated through common anchor items from prior administrations, using a Stocking and Lord procedure. Raw score and scale score frequency distributions and histograms were produced for each test form and compared in order to ensure that all test forms within each content area were producing comparable results.
2. January: Items were transformed from the original scale (mean of 500, s.d. of 100) to a new reporting scale with mean of 400, s.d. of 40.
3. January: Preliminary HOSS/LOSS values were established for each content area.
4. January: All old items in the item pool were transformed from the 500/100 scale to the 400/40 scale.
5. May: Within each content area, items on all of the May test forms were calibrated together, equated and placed on the new 400/40 scale using common anchor items in a Stocking and Lord procedure. Raw score and scale score frequency distributions and histograms were produced for each test form and compared in order to ensure that all test forms within each content area were producing comparable results.
6. The results from January and May were combined, and score distributions for each content area were examined within and across all of the test forms.
7. It was determined that the initial scale adjustment (linear transformation from a 500/100 scale to a 400/40 scale) had not produced the intended scale score distributions. (This was because the old scale was no longer producing the mean of 500 and s.d. of 100 that had been set in 2000.)
8. Therefore, a second linear transformation was applied to the January and May item parameters to produce the desired scale score mean of 400 and s.d. of 40.
9. Score distributions, raw-to-scale-score tables, conditional standard errors, etc. were produced for all of the January and May test forms.
10. Results were compared across test forms, and the initial HOSS/LOSS values were reexamined and adjusted as necessary on the basis of the complete set of test forms.

Note that we also carried out all of our usual data checking, classical item analysis, DIF analysis, etc., prior to the calibration/equating/scaling described above.