

Summary of Technical Reports on the
Regents Examinations in Comprehensive English (English Language Arts),
Mathematics A, Global History and Geography (Global History), and
United States History and Government

Gerald E. DeMauro
March 29, 2000

Technical reports on the Regents examinations in English Language Arts, Mathematics A, Global History and United States History and Government are available. The reports indicate that these instruments have been properly developed for their use, and provide for the technical reader the steps taken to equate the tests and screen the items.

Note, the two social studies tests are a year behind English Language Arts and Mathematics A. Review of these statistics will inform all future test development.

The model used for test development requires certain assumptions about the test and items:

- 1) That one factor is measured by each test;
- 2) That students of increasingly higher ability have increasing probability of answering the questions correctly;
- 3) That the questions can be ordered to describe the probabilities on a scale in which the abilities of the students can also be located.

The analyses are directed to the validity of the instruments in terms of these assumptions, the soundness of the assumptions; instrument reliability, and fairness of the instruments to all populations. The results are summarized below.

Assumptions of the Model

Three of the tests each seem to be described by one overriding factor and several smaller factors. The U.S. History and Government test did not have a completed field test, so it was not factory analyzed.

Analyses were made of how well the test questions fit the test model. As the ability of student groups increases the probabilities of answering each question correctly should also increase in a well-specified relationship. Analyses of the fit of the questions also addresses test the dimensionality, or the factors measured by the test, by evaluating the extent of concurrence of each item with the main trait. The analyses show that very few items were

misfitting on any test form. A misfitting item would have mean square values lower than 0.7 or greater than 1.3. These fit statistics for all items show that only 9 of 300 English Language Arts, 17 of 328 Mathematics A, 4 of 486 Global History and 3 of 343 United States History items failed the criteria in field testing.

Fairness

The item analyses included item fairness using the generalized Mantel-Haenszel method. This is the first two steps process that also includes a review by content expert of any items flagged by the empirical analysis.

The first step of the process is an empirical analysis that flags items that are differentially difficult for populations matched on overall scores. The items may not actually be more difficult for the whole of one group or the other, but more difficult for groups matched by skill. The content analysis then determines if the content measured by flagged items is necessary for achievement of the standards.

Because the matching procedures limit numbers of examinees in each group, the need for sufficient sample sizes limited the comparisons to male/female and African American/European Americans. Items that were flagged were sent for further analyses to the content experts.

Equating

The items were placed on the same scale as examinee scores for equating purposes. Setting the values of overlapping items shifts the whole test scale, and shifts the student scores with it to be congruent with one form of the test. Therefore, whenever these items are used on a test form, the scores of the students can be transformed to a common scale.

Conclusion

The analyses are useful for test development and to insure equity from test form to test form. The instruments now in use have taken advantage of these analyses.