Utilizing the data acquired from the administration of items to the above-mentioned population, developers analyzed the items' characteristics for difficulty, bias, and contribution to score accuracy. The three-parameter logistic model (Lord & Novick, 1968; Lord, 1980) from item response theory was used in the analysis. The three-parameter logistic model takes into account item discrimination, difficulty, and guessing for each item. Item discrimination is an indicator of the item's power to elicit different responses from higher-achieving and lower-achieving examinees. The greater the discriminating power of the items selected, the fewer items are needed to attain a given level of accuracy in test results. Developers utilized four different procedures to reduce bias in the Test of Adult Basic Education 7 and 8: (a) editorial attention to content, (b) use of McGraw-Hill guidelines, (c) review by educational community professionals, and (d) an empirical procedure following Linn and Harnisch (1981). Test developers reported while the first three steps mentioned above improved the quality of the test and reduced bias, the Linn and Harnisch procedure examined each item for differential performance among groups of students. The procedures used not only overall group differences, but also differences among quintile groups. Use of quintile group differences, as well as overall group differences, allows for detection of items biased within one or more score groups that would otherwise be hidden in overall data. The data were analyzed for examinees identified as Black, Hispanic, Asian, and a final group labeled other. Data from male and female examinees were also compared. The selection of potentially biased items for the final version of each test was minimized, and the final version of the test was required to have a lower average bias rating than the pool of items from which it was selected. Items appearing to be biased from a content, context, or statistical standpoint were not selected for the final test. (p. 9)

Developers' statistical goals for item selection were (a) to minimize the standard error of measurement over the ability range for which the test level was intended, nominal range, (b) maximize fit, (c) minimize bias, and (d) items whose answer choices performed in a reasonable manner. Items were selected with minimum measurement error throughout