Section 4. Item Calibration and Test Equating

Item Calibration

All of the items in the May 2008 field test forms were calibrated concurrently using the three parameter logistic (3PL) model and the Linking sample for each subject area. As described in Section 1, this sample consisted of a subset of the May 2008 HSA examinees who also took the Mod-HSA in the same content area to provide data that could be used for linking. All students in the Linking sample took the Mod-HSAs online.

The 3PL model states that the probability that a person with ability θ will respond correctly to item j can be expressed as follows:

$$P(U_j = 1 \mid \theta) = P_j(\theta) = c_j + \frac{1 - c_j}{1 + e^{-1.7a_j(\theta - b_j)}}$$

where:

U_j is the response to item j, 1 if correct and 0 if incorrect;
a_j is the slope parameter of item j, characterizing its discrimination;
b_j is the threshold parameter of item j, characterizing its difficulty; and
c_j is the lower asymptote parameter of item j, reflecting the chance that students with very low proficiency will select the correct answer; sometimes called the "pseudo-guessing" level.

A proprietary version of the PARSCALE computer program (Muraki & Bock, 1995) was used to estimate the item parameters.

Initially all item parameters were freely estimated and placed on scale using the Stocking and Lord (1983) procedure. Comparisons of the characteristic curves for the linking and reference parameters after linking revealed some divergence at the lower end of the ability scale. A second calibration run was conducted after fixing the c-parameters of the linking items to their bank values. This approach improved the correspondence between the characteristic curves for reference and linking parameters throughout the full range of ability.

Test Equating

The Mod-HSA forms were linked to the HSA scale using two linking methods. The first involved use of common items, and the second involved use of common persons. Two approaches were planned, so that if one method was not effective, there would be an alternative approach to link the Mod-HSAs to the HSA scale. The equating methods used for each design are described in the following sections.

Linking using the Common Items Design

The 24 anchor items in each form were drawn from the HSA item bank and placed on the field test forms without modification. The linking items were placed in approximately the same positions within the test as when the items were originally field tested, to avoid position effects. The same 24 items were used in both forms administered in each content area.

The linking items had parameters estimated when the field test forms were calibrated. The banked parameters were expressed on the HSA reporting scale. The Stocking and Lord (1983) procedure was used to align the TCCs based on the two sets of parameters and to derive linear constants that could be used to transform the Mod-HSA parameters to the HSA reporting scale.

Linking using the Common Persons Design

Students in the linking samples had taken both a regular HSA and a Mod-HSA assessing the same content area. As a result, two sets of scores were available for the students; an HSA scale score and a Mod-HSA theta estimate generated after the Mod-HSA forms were calibrated.

Linear equipercentile equating was used to find linear transformation values that would minimize the differences between scores on two test forms (Yen & Fitzpatrick, 2006). The process involves using the equipercentile equating procedure to identify corresponding scores on the two forms, one of which is a target form. A linear transformation then is found that minimizes the differences between the scores of the form to be equated and the target form. The transformation is applied to the item parameters and ability estimates associated with the form to be equated to align them with the target score scale.

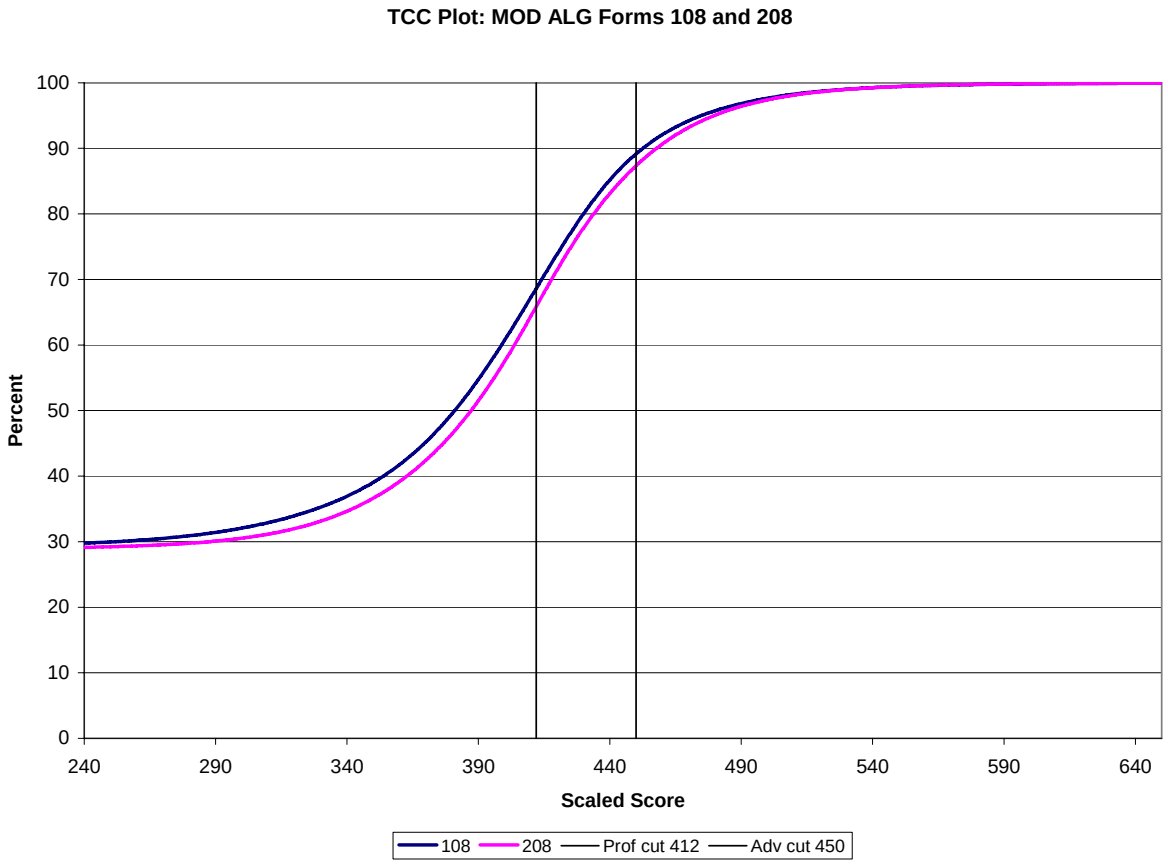
In this case the regular HSA form was the target, so student scores obtained using the Mod-HSAs were aligned to their HSA scale scores. These analyses were completed using *LinEq*, an ETS proprietary software program.

Comparison of Results from the Two Equating Methods

A summary of the equating findings is provided in Appendix C. In general, the two approaches to linking the Mod-HSA to the HSA produced very similar results. The NPC recommended to use the results obtained using the parameters of the common items design and the Stocking and Lord (1983) equating method. One reason for this decision was that IRT equating allows for the removal of misbehaving items; another was that this would allow the same equating method to be used for both the Mod-HSA and the HSA.

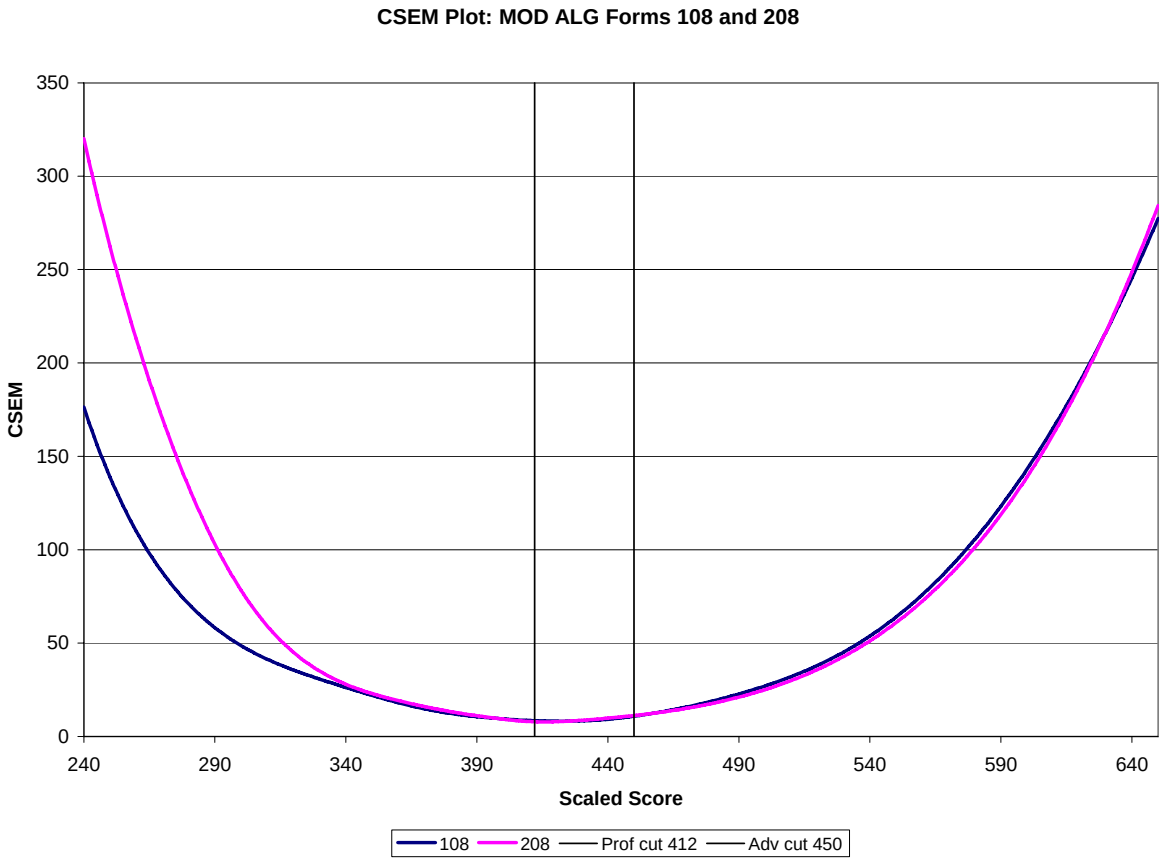
Comparison of TCCs and CSEMs

Figures 4.1 to 4.8 present the TCCs and CSEMs, resulting from the Stocking and Lord (1983) equating method, for the two forms in each content area.



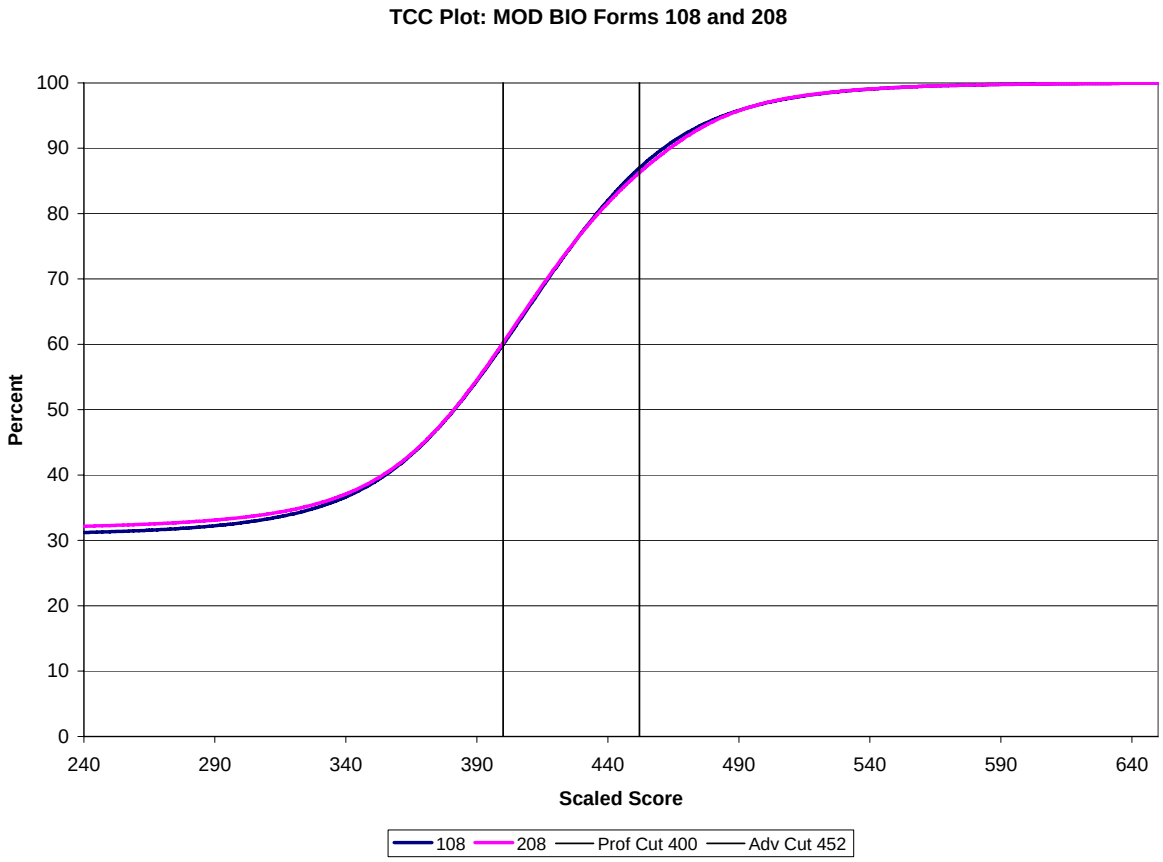
Note Algebra Cut Scores: Proficient 412, Advanced 450

Figure 4.1 TCCs for the Mod-HSA Algebra Forms



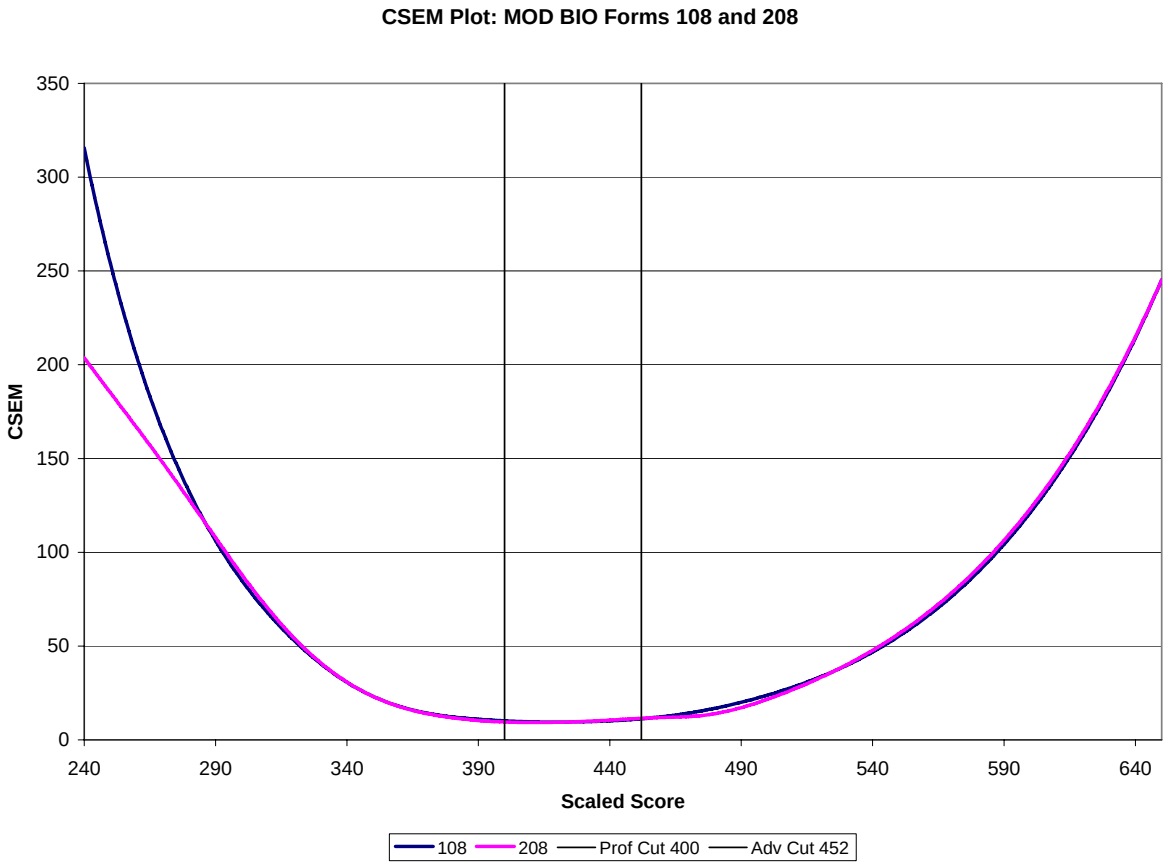
Note Algebra Cut Scores: Proficient 412, Advanced 450

Figure 4.2 CSEMs for the Mod-HSA Algebra Forms



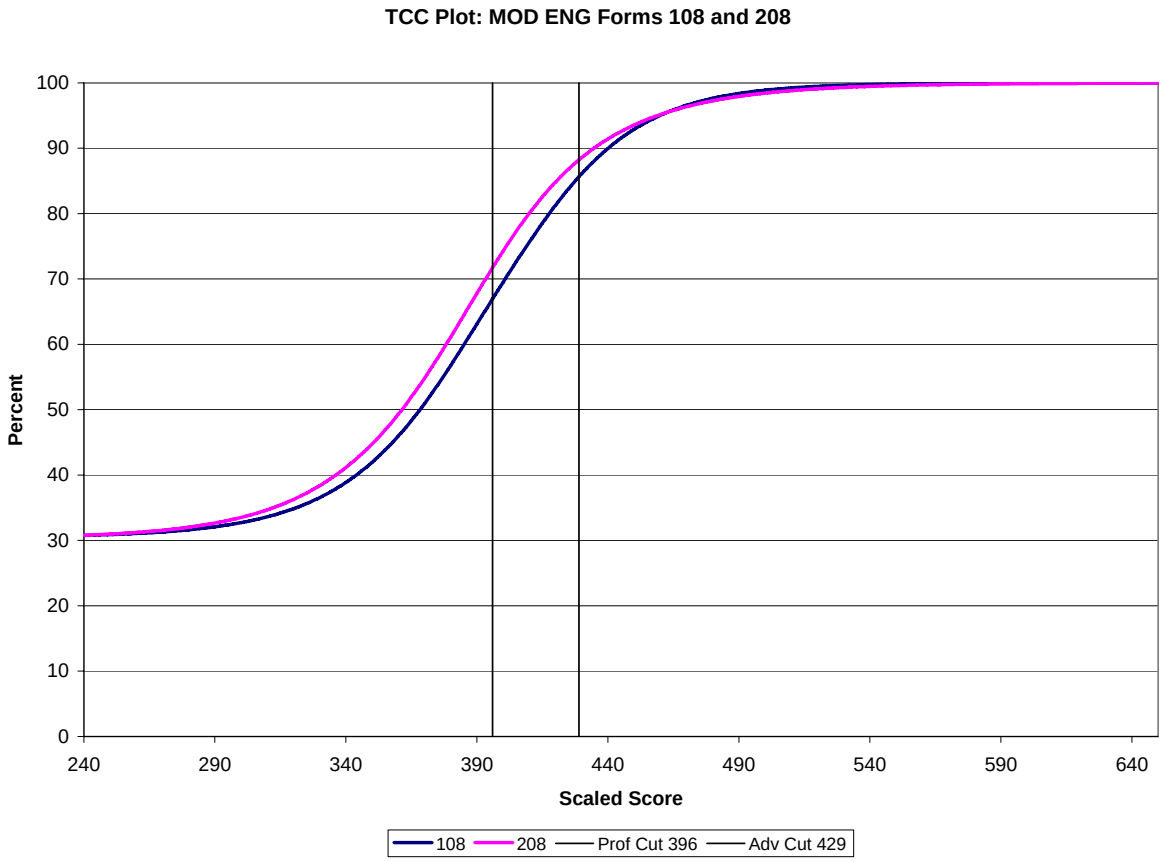
Note Biology Cut Scores: Proficient 400; Advanced 452

Figure 4.3 TCCs for the Mod-HSA Biology Forms



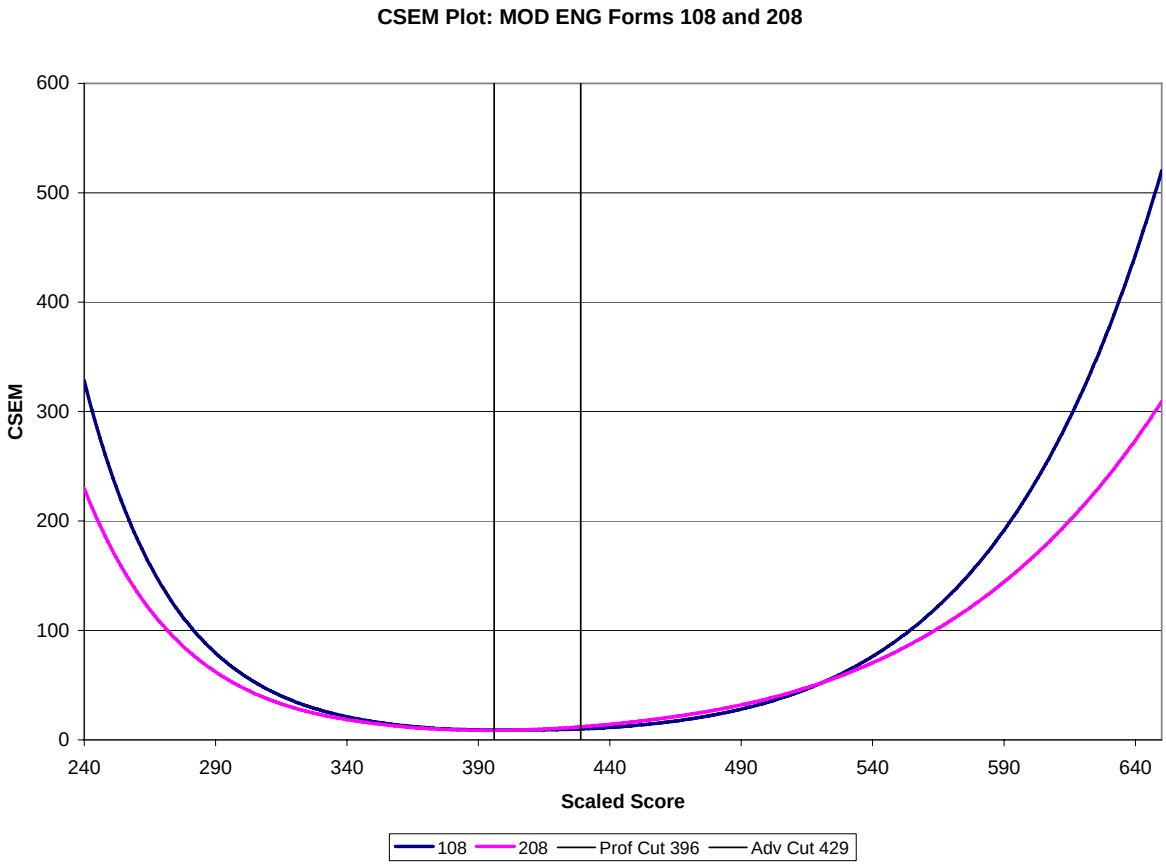
Note Biology Cut Scores: Proficient 400; Advanced 452

Figure 4.4 CSEMs for the Mod-HSA Biology Forms



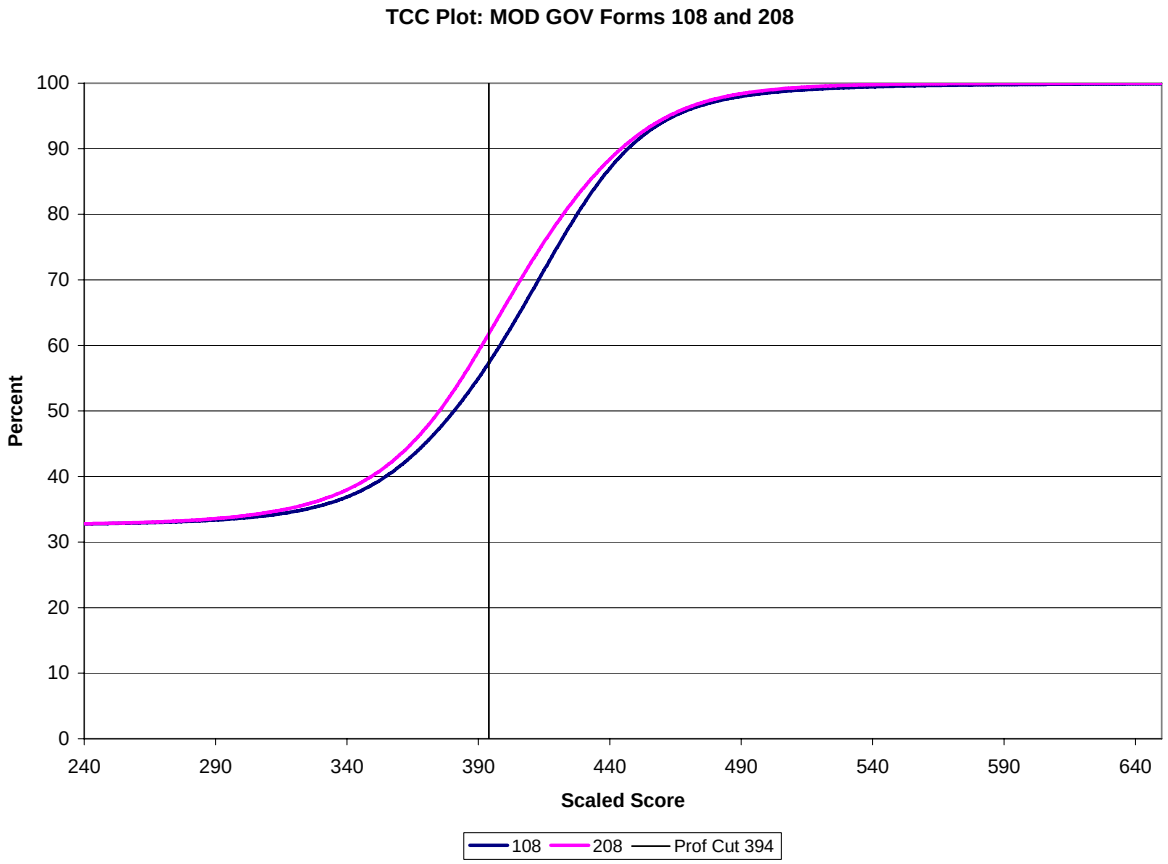
Note English Cut Scores: Proficient 396, Advanced 429

Figure 4.5 TCCs for the Mod-HSA English Forms



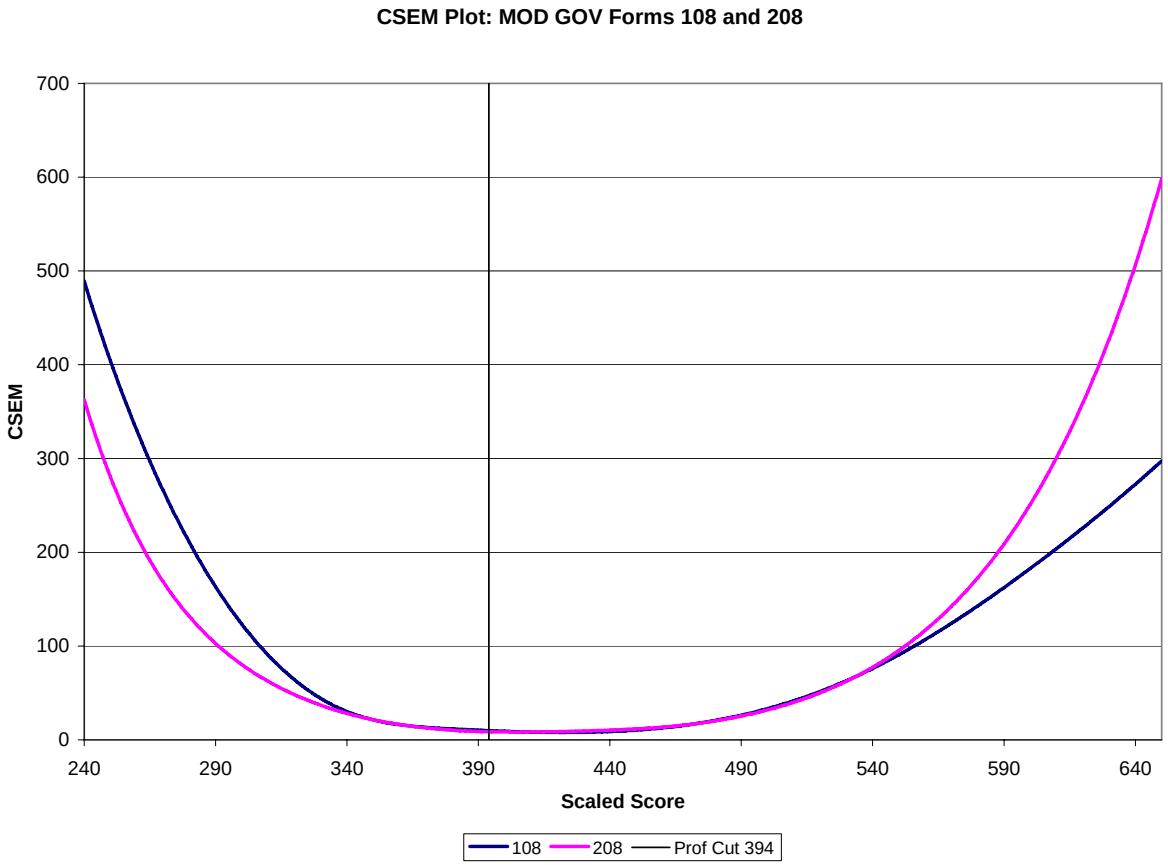
Note English Cut Scores: Proficient 396, Advanced 429

Figure 4.6 CSEMs for the Mod-HSA English Forms



Note Government Cut Score: Proficient 394

Figure 4.7 TCCs for the Mod-HSA Government Forms



Note Government Cut Score: Proficient 394

Figure 4.8 CSEMs for the Mod-HSA Government Forms