

Appendix 3.C Establishing the HOSS and LOSS

Maryland High School Assessment

March 17, 2004

Educational Testing Service

Appendix 3.C Establishing the HOSS and LOSS

Principles for determining the HOSS and LOSS in May 2002 were described in email correspondence from Diana Marr, Research Scientist at CTB (March 16, 2004). The text of the email is printed below.

To determine the optimal HOSS and LOSS for each test form, we adopted the following principles (put forth by Wendy Yen in a 1991 memorandum):

For HOSSes,

1. The HOSS must be greater than $SS(n-1)$.
2. The HOSS must be high enough that it does not cause an unnecessary pileup of scale scores at the top of the scale.
3. The HOSS should be low enough that $SE(HOSS) < 10 * \text{Min}(SE)$.
4. The $SE(HOSS)$ should change smoothly over levels. [The HOSS gaps should also change smoothly over levels, insofar as possible, but this is less important than maintaining smooth $SE(HOSS)$ changes.]
5. The HOSS should be such that Number Correct SS and Item Pattern SS are tau equivalent.
6. The HOSS gap should be in the same ballpark as the penultimate HOSS gap.

For LOSSes,

1. The LOSS should be low enough that it does not cause an unnecessary pileup of IP scale scores at the bottom of the scale.
2. The LOSS should be high enough that $SE(LOSS) < 15 * \text{MIN}(SE)$; this criterion may be difficult to meet for some tests.
3. In general, the LOSS should be $< SS(\text{Sum } c+1)$; however, if $SS(\text{Sum } c+1)$ is poorly determined, causing violation of criterion b, then $(\text{Sum } c+2)$ may be treated as the lowest determined scale score.
4. The $SE(LOSS)$ should change smoothly over levels. [The LOSS gaps should also change smoothly over levels, insofar as possible, but this is less important than maintaining smooth $SE(LOSS)$ changes.]
5. The LOSS should be such that Number Correct SS and Item Pattern SS are tau equivalent.
6. The LOSS gap should be in the same ballpark as the penultimate LOSS gap.

After using these criteria to estimate the "optimal" HOSS and LOSS for each individual test form, results were then compared across all of the test forms within each content area to arrive at a single HOSS and LOSS for each content area. Because the 2002 test forms had been developed as field test forms, there was less consistency across forms than would be expected in a group of operational test forms. Thus, the optimal HOSS and LOSS varied considerably from form to form, and the selection of a single pair of values for each content area necessarily involved some compromises. For each content area, the final HOSS typically fell

somewhere between the lowest and highest individual test form HOSS, and the final LOSS typically fell somewhere between the lowest and highest individual test form LOSS.

References

Fitzpatrick, A. R., Link, V. B., Yen, W. M., Burket, G. R., Ito, K., & Sykes, R. C. (1996). Scaling performance assessments: A comparison of one-parameter and two-parameter partial credit models. *Journal of Educational Measurement*, *33*(3), 291-314.