

Supplemental Material S8. AUC analysis of magnitude change differences between treatment and control groups, examining partial credit scores, rather than %C, alongside model-based statistics.

Bar graphs showing the pairwise classification accuracy (i.e., the area under the receiver operating characteristic curve, AUC) for distinguishing treatment participants from control participants based on the magnitude of change in each summary statistic. Error bars representing 95% confidence intervals for AUC were estimated using 1,000 bootstrap samples with replacement. This is the same information that is presented in Figure 6 of the main text, but %C has been replaced with a partial credit score: No Attempt, Abstruse Neologism, and Unrelated errors were assigned zero points; Semantic, Mixed, and Neologism errors were assigned one point; Formal errors were assigned two points; Correct responses were assigned three points; the final score was a sum of points over all items, divided by the maximum number of possible points. Comparable to the other accuracy statistics, the partial credit score was unable to discriminate between treatment and control participants at rates significantly greater than chance (Treatment Group 1: AUC = 59%, CI = [45%, 72%]; Treatment Group 2: AUC = 57%, CI = [40%, 71%]). An alternative scoring rubric yielded nearly identical results (not pictured): Non-Naming Attempt = 0 points; Abstruse Neologism = 1 point; Neologism, Unrelated = 2 points; Semantic, Formal = 3 points; Mixed = 4 points; Correct = 5 points.

