

Supplemental Material S7. Statistical support for aggregated baseline naming data.

All word classes

In assessing variation across the baseline period, the 3-way Condition (treated, untreated) \times Diagnosis (svPPA, lvPPA, A) \times Baseline Assessment (baseline 1, baseline 2, baseline 3) interaction was not significant ($F[4,48] = 1.766, p = .151$), and neither was the 2-way Condition \times Baseline Assessment interaction ($F[2,48] = 0.921, p = .405$). The Diagnosis \times Baseline Assessment interaction was significant, but the simple main effects of Baseline Assessment for the three diagnostic groups were all non-significant (all $ps > .05$).

Nouns

There was a significant 3-way Condition \times Diagnosis \times Baseline Assessment interaction ($F[4,48] = 4.894, p = .002$). In order to understand the 3-way interaction, the simple 2-way Diagnosis \times Baseline Assessment interactions were subsequently analyzed for each of the two conditions. For the untreated condition, the Diagnosis \times Baseline Assessment interaction was not significant ($F[4,24] = 1.621, p = .202$), indicating that the non-significant Baseline Assessment component ($F[2,24] = 1.299, p = .291$) could be generalized to all three diagnostic groups. There was, however, a significant Diagnosis \times Baseline Assessment interaction for the treated condition ($F[4,24] = 15.541, p < .001$). The source of this interaction could be traced to a significant improvement in performance from baseline 1 to baseline 2 for the lvPPA diagnostic group ($t[24] = 8.083, p < .001$). Although the improvement was statistically significant, it only represented a relatively small increase (Cohen's $d = .28$), such that an aggregated baseline for nouns was used.

Verbs

The 3-way Condition \times Diagnosis \times Baseline Assessment interaction was not significant ($F[4,48] = 2.105, p = .095$), and neither was the 2-way Condition \times Baseline Assessment interaction ($F[2,48] = 0.092, p = .912$). The Diagnosis \times Baseline Assessment interaction, however, was significant ($F[4,48] = 4.242, p = .005$). The source of this interaction could be traced to a significant improvement in performance from baseline 1 to baseline 3 for the lvPPA diagnostic group ($t[48] = 3.266, p = .002$). Although the improvement was statistically significant, it only represented a relatively small increase (Cohen's $d = .24$), supporting the use of an aggregated baseline for verbs.

Adjectives

The 3-way Condition \times Diagnosis \times Baseline Assessment interaction was not significant ($F[4,48] = 2.353, p = .067$), and neither was the 2-way Condition \times Baseline Assessment interaction ($F[2,48] = 1.100, p = .341$) nor the 2-way Diagnosis \times Baseline Assessment interaction ($F[4,48] = 0.825, p = .516$). Because Baseline Assessment did not interact with the other factors, its non-significant main effect ($F[2,48] = 2.871, p = .066$) can be generalized across both conditions and all three diagnostic groups, rationalizing the use of an aggregated baseline for adjectives.