Supplemental Table S4. Sentence production probe effect sizes.

| Participant (Tx order) | Treated Abstract Sets | | | | Untreated Gx Sets | |
| | 1st treated set | | 2nd treated set | | Gx Abstract Set | Gx Concrete Set |
	Tx ES	Follow-up ES	Tx ES	Follow-up ES	Follow-up ES	Follow-up ES
AV-1 (A2, A1)	3.16	2.29	2.02	2.37	1.13	0.57
AV-2 (A3, A1)	4.60	4.77	8.92	8.31	4.41	1.55
AV-3 (A2, A3)	16.90	16.58	26.85	22.55	15.32	12.53

Note. Tx = treatment; ES = effect size; Gx = generalization. As a measure of the magnitude of change, *d* statistic effect sizes (Beeson & Robey, 2006; Busk & Serlin, 1992) were calculated for the sentence production probe values for each set and for each participant. These effect sizes are calculated with the following formula:

$$d = \frac{\bar{x}_{A_2} - \bar{x}_{A_1}}{S_{A_1}}$$

For the sets of stimuli that received treatment, two separate effect sizes were calculated: a treatment effect size (using data from the last five original baseline values and the last two treatment phase values), and a follow-up effect (using data from the last five baseline values and the two follow-up probe values). This means that for the set treated in the second treatment phase, the treatment effect size shows the effect of both phases of treatment (baselines from before the first treatment phase). For the generalization sets of stimuli (those not receiving treatment), only the follow-up effect sizes were calculated, and these were calculated in the same way as just described. These effect sizes reflect the change in sentence production probe performance while considering the variability of performance during the baseline phase. Note that effect sizes do not take into account the presence or lack of experimental control, and so effect sizes other than those for which experimental control was demonstrated via the CDC lines (i.e., those other than the first treatment phase effect sizes for Participants 1 and 3) should be viewed as correlative effects rather than causal effects, and may be the result of variables other than the primary independent variable (the experimental treatment. Without benchmarks for effect size interpretation, the effect sizes are interpreted only in relation with each other. All effect sizes were positive, indicating increases in performance. For the treated sets, treatment and follow-up effect sizes were similar to each other for a given participant for a given set. Effect sizes for Participants 2 and 3 were much larger for the second treatment set compared to the first. Follow-up effect sizes for the untreated low concreteness generalization sets were roughly similar to those of the first treated sets for Participants 2 and 3, and slightly smaller for Participant 1. Follow-up effect sizes for the untreated high concreteness generalization sets were the smallest effect sizes for each participant.

References

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: lessons learned from the aphasia literature. *Neuropsychology Review, 16*(4), 161–169. https://doi.org/10.1007/s11065-006-9013-7

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education*. Hillsdale, NJ: Lawrence Erlbaum Associates.