**SUPPLEMENTARY MATERIAL**

**Appendix 1.**

**University of California, San Francisco Scanning parameters.** T1 images were acquired on a 3T Siemens TrioTim syngo MRI scanner using a MPRAGE sequence (160 sagittal slices; slice thickness = 1 mm; field of view = 256 mm2; matrix = 256 x 240; voxel size 1.0 x 1.0 x 1.0 mm; repetition time = 2300 msec; echo time = 2.98 msec; inversion time = 900 msec; flip angle = 9°).

**University of Texas, Austin Scanning parameters.** T1 images were acquired on a Siemens 3T Magnetom Skyra MRI scanner using a MPRAGE sequence (192 sagittal slices; slice thickness = 0.90 mm; field of view = 240 mm2; matrix 256 × 256; voxel size .94 × .94 × .90 mm; repetition time = 1900 msec; echo time = 2.49 msec; inversion time = 900 msec; flip angle = 9°).

**Appendix 2.** Reliability of coding for primary outcome measure and treatment fidelity

In addition to on-line coding by the treating clinician, spoken productions were transcribed and coded by a naïve listener blind to treatment condition (trained versus untrained script) and time point (pre-treatment, post-treatment or follow-up) using CHAT and CLAN software (MacWhinney, 2000; MacWhinney et al., 2011). Each spoken word was coded as intelligible or unintelligible, and as present or omitted relative to the script target. Productions that were deemed unintelligible were transcribed phonetically. If the participant self-corrected, the repaired response was counted as correct. Interrater reliability, comparing the clinician’s online data collection and second coder’s transcription, was conducted for 25% of the total number of treatment sessions for each participant. Point-by-point agreement (Kadzin, 1982) was used to calculate reliability for each participant, which was then averaged across participants. For each participant, two raters determined if each scripted word was produced or omitted during probing. Subsequently, the number of agreements and disagreements was calculated between the two raters. The number of agreements was then divided by the number of agreements plus disagreements, and the derived value was then multiplied by 100. Overall, reliability was high, at 94.33%.

Two trained undergraduate raters reviewed video recordings of 25% of the total number of treatment sessions in order to determine whether the clinicians adhered to the treatment protocol. Fidelity scores revealed that the clinicians correctly administered treatment steps in 99.71% of opportunities.

Reference:

Kazdin A. Single-case experimental designs: Methods for clinical and applied settings. New York: Oxford 1982.

**Supplementary Table 1.** Descriptive statistics for primary and secondary outcome measures and standardized tests.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Minimum | Maximum | Mean | Std. Deviation |
| Primary outcome measure: | Pre-treatment | 11.78 | 57.86 | 38.01 | 13.70 |
| Trained scripts | Post-treatment | 56.04 | 98.65 | 89.78 | 14.67 |
|  | 3 months post | 39.34 | 99.11 | 87.52 | 19.52 |
|  | 6 months post | 12.13 | 96.97 | 79.89 | 27.19 |
|  | 1 year post | 6.24 | 99.53 | 68.64 | 35.15 |
| Primary outcome measure: | Pre-treatment | 18.41 | 70.90 | 36.07 | 14.71 |
| Untrained scripts | Post-treatment | 22.50 | 88.11 | 43.07 | 19.89 |
|  | 3 months post | 10.60 | 81.15 | 35.62 | 19.74 |
|  | 6 months post | 3.75 | 75.41 | 30.06 | 20.14 |
|  | 1 year post | 5.43 | 85.25 | 31.06 | 26.12 |
| % Intelligible words: | Pre-treatment | 84.62 | 100.00 | 95.75 | 5.09 |
| Trained scripts | Post-treatment | 94.20 | 100.00 | 99.25 | 1.85 |
|  | 3 months post | 96.85 | 100.00 | 99.29 | 1.17 |
|  | 6 months post | 95.78 | 100.00 | 99.10 | 1.34 |
|  | 1 year post | 81.88 | 100.00 | 96.62 | 5.74 |
| % Intelligible words: | Pre-treatment | 87.54 | 100.00 | 97.38 | 3.83 |
| Untrained scripts | Post-treatment | 94.24 | 100.00 | 98.38 | 2.36 |
|  | 3 months post | 99.15 | 100.00 | 99.92 | 0.27 |
|  | 6 months post | 87.50 | 100.00 | 98.08 | 3.80 |
|  | 1 year post | 70.00 | 100.00 | 94.34 | 9.41 |
| Grammatical errors per hundred words: | Pre-treatment | 1.22 | 30.40 | 9.23 | 11.72 |
| Trained scripts | Post-treatment | 0.00 | 11.98 | 1.77 | 3.66 |
|  | 3 months post | 0.00 | 17.97 | 3.87 | 5.87 |
|  | 6 months post | 0.27 | 17.51 | 3.76 | 5.24 |
|  | 1 year post | 0.44 | 15.67 | 4.08 | 4.84 |
| Grammatical errors per hundred words: | Pre-treatment | 0.00 | 21.30 | 7.41 | 8.40 |
| Untrained scripts | Post-treatment | 0.00 | 15.38 | 5.39 | 5.84 |
|  | 3 months post | 0.00 | 22.22 | 5.09 | 6.62 |
|  | 6 months post | 0.00 | 17.59 | 5.35 | 6.42 |
|  | 1 year post | 0.82 | 13.89 | 5.44 | 4.57 |
| Northwestern Anagram Test | Pre-treatment | 36.67 | 90.00 | 63.70 | 21.57 |
|  | Pre-treatment (short version)\* | 40.00 | 83.33 | 65.55 | 16.69 |
|  | Post-treatment | 46.67 | 100.00 | 74.33 | 19.94 |
|  | Post-treatment (short version)\* | 41.67 | 100.00 | 72.83 | 21.10 |
|  | 3 months post\* | 25.00 | 100.00 | 66.50 | 27.05 |
|  | 6 months post\* | 10.00 | 100.00 | 54.34 | 34.28 |
|  | 1 year post\* | 0.00 | 100.00 | 42.62 | 37.65 |
| Western Aphasia Battery | Pre-treatment | 76.20 | 96.90 | 84.30 | 6.36 |
|  | Post-treatment | 78.80 | 97.70 | 85.69 | 6.08 |
|  | 3 months post | 71.60 | 97.30 | 81.66 | 8.33 |
|  | 6 months post | 66.50 | 97.70 | 79.93 | 8.89 |
|  | 1 year post | 55.70 | 97.40 | 75.53 | 11.89 |

\*A short (15-item) version of the Northwestern Anagram Test was administered at follow-up assessments. The full version was used for pre-post treatment comparisons; the short version was used for paired comparisons between pre-treatment and all follow-up assessments.

**Supplementary Table 2.** Results of paired permutation tests examining the primary outcome measure (percent correct, intelligible scripted words) for trained and untrained scripts. *t =* t-value derived from conventional test; *P* = *p*-value derived from permutation test. Bonferroni correction was applied across time points within each outcome measure, with adjusted alpha of *P* = 0.0125.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Pretreatment performance vs.** | | | | | | | | | |
|  | **Post-treatment (N=10)** | | **3 Months post**  **(N=10)** | | **6 Months post (N=10)** | | **1 Year post**  **(N=9)** | |
|  | **t** | ***P*** | ***t*** | ***P*** | ***t*** | ***P*** | ***t*** | ***P*** |
| Trained Scripts (1-tailed) | -8.762 | **0.001** | -7.199 | **0.002** | -4.555 | **0.002** | -2.540 | **0.012** |
| Untrained Scripts  (2-tailed) | -1.685 | 0.043 | 0.097 | 0.680 | 0.998 | 0.680 | 0.765 | 1.000 |
| **Trained vs. untrained scripts** | | | | | | | | |
| Trained vs. Untrained  (1-tailed) | 8.358 | **0.002** | 7.155 | **0.002** | 7.270 | **0.002** | 4.242 | **0.004** |

**Supplementary Table 3.** Results of paired permutation tests examining overall percent intelligible words for trained and untrained scripts. *t =t*-value derived from conventional test; *P* = *p*-value derived from permutation test. Bonferroni correction was applied across time points within each outcome measure, with adjusted alpha of *P* = 0.0125.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Pretreatment performance vs.** | | | | | | | | | |
|  | **Post-treatment (N=10)** | | **3 Months post**  **(N=10)** | | **6 Months post (N=10)** | | **1 Year post**  **(N=9)** | | |
|  | **t** | ***P*** | ***t*** | ***P*** | ***t*** | ***P*** | ***t*** | ***P*** |
| Trained Scripts  (1-tailed) | -2.020 | **0.006** | -2.087 | **0.006** | -1.923 | **0.002** | -0.861 | 0.879 |
| Untrained Scripts  (2-tailed) | -1.027 | **0.012** | -2.124 | **0.002** | -1.875 | **0.012** | 1.359 | 0.359 |
| **Trained vs. untrained scripts** | | | | | | | | | |
| Trained vs. Untrained  (1-tailed) | 0.929 | 0.988 | -1.892 | 1.000 | 0.782 | 0.940 | 1.504 | 0.551 |

**Supplementary Table 4.** Results of paired permutation tests examining grammatical errors per hundred words for trained and untrained scripts. *t = t*-value derived from conventional test; *P* = *p*-value derived from permutation test. Bonferroni correction was applied across time points within each outcome measure, with adjusted alpha of *P* = 0.0125.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Pre Treatment performance vs.** | | | | | | | | | |
|  | **Post-treatment (N=10)** | | **3 Months post (N=10)** | | **6 Months post**  **(N=10)** | | **1 Year post**  **(N=9)** | |
|  | **t** | ***P*** | ***t*** | ***P*** | ***t*** | ***P*** | ***t*** | ***P*** |
| Trained Scripts  (1-tailed) | 2.085 | **0.002** | 1.689 | 0.021 | 1.592 | 0.092 | 1.611 | 0.180 |
| Untrained Scripts  (2-tailed) | 2.163 | 0.184 | 1.475 | 0.680 | 0.947 | 0.680 | 1.113 | 1.000 |
| **Trained vs. untrained scripts** | | | | | | | | | |
| Trained vs. Untrained  (1-tailed) | -2.402 | **0.001** | -1.421 | 0.021 | -1.032 | 0.092 | -2.034 | **0.012** |

**Supplementary Table 5.** Results of paired permutation tests for standardized tests. NAT = Northwestern Anagram Test; WAB = Western Aphasia Battery; *t = t-*value derived from conventional test; *P* = *p*-value derived from permutation test. Bonferroni correction was applied across time points within each outcome measure, with adjusted alpha of *P* = 0.0125.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Pretreatment performance vs.** | | | | | | | | |
|  | **Post-treatment (NAT: N=9**  **WAB: N=10)** | | **3 Months post**  **(NAT: N=9**  **WAB: N=10)** | | **6 Months post**  **(NAT: N=9**  **WAB: N=10)** | | **1 Year post**  **(NAT: N=8**  **WAB: N=9)** | |
|  | **t** | ***P*** | ***t*** | ***P*** | ***t*** | ***P*** | ***t*** | ***P*** |
| NAT  (2-tailed) | -2.603 | 0.023 | 0.320 | 0.086 | 1.073 | 0.086 | 2.467 | 0.172 |
| WAB  (2-tailed) | -1.543 | 0.043 | 2.396 | 0.184 | 2.701 | 0.184 | 3.387 | 0.023 |

**Supplementary Table 6.** Results of voxel-based morphometry revealing areas of significant gray matter loss in nfvPPA patients relative to healthy controls within the inferior frontal gyrus-seeded network from Mandelli et al., 2016 (see Fig. 4).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | Peak MNI Coordinates | | |
| Brain regions (AAL) | Extent (mm3) | Max t-value | x | y | z |
| Left Putamen | 119 | 6.5081 | -21 | -1.5 | 4.5 |
| Left and right supplementary motor area, left and right cingulate, left medial superior frontal gyrus | 1098 | 5.6636 | 6 | 12 | 54 |
| Left postcentral gyrus | 64 | 5.6601 | -46.5 | -13.5 | 37.5 |
| Left inferior frontal gyrus, precentral gyrus, rolandic operculum | 2259 | 5.6384 | -45 | 12 | 27 |
| Right precentral gyrus and inferior frontal gyrus | 324 | 4.883 | 48 | 9 | 31.5 |
| Left postcentral gyrus | 110 | 4.1726 | -42 | -16.5 | 42 |

**Supplementary Table 7.** Partial correlation analyses examining the relation between ROI volumes and treatment effect size (*d*-statistic), controlling for aphasia severity (*Western Aphasia Battery* AQ). Bonferroni correction was applied, with adjusted alpha of *P* = 0.0045.

|  |  |  |
| --- | --- | --- |
| **Region of interest (ROI)** |  |  |
| left caudate/putamen | Correlation | 0.245 |
|  | Significance (1-tailed) | 0.28 |
| left cuneus | Correlation | -0.028 |
|  | Significance (1-tailed) | 0.473 |
| left IFG | Correlation | 0.603 |
|  | Significance (1-tailed) | 0.057 |
| left IPL | Correlation | 0.579 |
|  | Significance (1-tailed) | 0.066 |
| left MTG/ITG | Correlation | 0.875 |
|  | Significance (1-tailed) | **0.002** |
| right caudate | Correlation | -0.005 |
|  | Significance (1-tailed) | 0.496 |
| right IFG | Correlation | 0.01 |
|  | Significance (1-tailed) | 0.491 |
| right ITG | Correlation | 0.446 |
|  | Significance (1-tailed) | 0.134 |
| right parietal | Correlation | 0.175 |
|  | Significance (1-tailed) | 0.34 |
| right precuneus/occipital | Correlation | 0.226 |
|  | Significance (1-tailed) | 0.295 |
| SMA | Correlation | 0.528 |
|  | Significance (1-tailed) | 0.089 |