Cover: Statistical Analysis Plan

NCT #: NCT02332915

Title: Effect of Intensity of Treatment on Rehabilitation of Acquired Apraxia of Speech

Date: October 3, 2020

Statistical Analyses

Effect Sizes

Sets of effect size values (d-values) were calculated for each participant as an indication of magnitude of change for *treated items* and *untreated items*: 1) pre-treatment/baseline vs. 2-week follow-up – Intense Phase Items and pre-treatment vs. 2-week follow-up – Traditional (non-intense) Phase Items.

For the pre-treatment versus follow-up comparisons, the three baseline probes immediately preceding application of treatment were used along with the probe completed at two weeks post-treatment. The d-values were derived using the following equation: $ES = (M_{A2} - M_{A1}) / SD_{A1.}$ In this equation, "A1" indicates the baseline values and "A2" indicates the 2 week follow up value.

The experimental design was a cross-over design in which each participant received each intensity of treatment in separate, sequential phases (i.e., each participant received two phases of treatment). The arms of the design are designated as Intense First and Traditional First. The arms reflect counterbalancing of order of treatment administration. Order of treatment was not an analysis of interest; there is no evidence to suggest that treatment order would be a pertinent factor in treatment response. The comparison of interest was treatment intensity – Intense versus Traditional treatment phases. Consequently, the data from the Intense phase of treatment for all 24 participants was analyzed relative to the data from the Traditional phase of treatment for all 24 participants.

Descriptive statistics are shown in Tables 1 and 2. Relative to SPT benchmark effect sizes described by Bailey, Eatchel, and Wambaugh (2015) large effect sizes were found for baseline to treatment phase comparisons for treated items for the group for both SPT- Traditional and SPT – Intense, with average effect sizes of 28.9 and 19.7, respectively. For untreated items, large effect sizes

were found for both SPT-Traditional (d=8.1) and SPT-Intense (d=7.9) in for the group. As reflected by the minimum and maximum values in Tables 1 and 2, there were wide ranges of effect sizes found across the individuals.

Table 1

Effect Size (Δ-values) *Descriptive Statistics for* ***Treated Items*** *by Treatment Condition for the Group of 24 Participants*

Condition	**Comparison**	**Mean d-values**	**SD**	**Minimum**	**Maximum**
Intense	BL vs 2 week FU	19.7	12.6	5.2	49.1
Traditional	BL vs 2 week FU	28.9	17.2	6.5	53.1

BL = baseline; Tx = treatment; wk = week

Bailey et al. (2015) benchmarks: small = 5.90, medium, = 7.12, large = 10.19 for baseline to follow-up phase comparisons

Table 2

Effect Size (Δ-values) *Descriptive Statistics for* ***Treated Items*** *by Treatment Condition for the Group of 24 Participants*

Condition	**Comparison**	**Mean d-values**	**SD**	**Minimum**	**Maximum**
Intense	BL vs 2 week FU	7.9	5.6	-4.1	34.6

Traditional	BL vs 2 week FU	8.1	8.9	-1.0	20.9

BL = baseline; Tx = treatment; wk = week

Bailey et al. (2015) benchmarks: small = 2.59, medium, = 4.23, large = 6.47 for baseline to follow-up phase comparisons

Comparisons of average group effect sizes were completed using parametric statistics (i.e., dependent t-tests, two-tailed) when possible and the nonparametric analogs when normality tests were failed (i.e., Wilcoxon Signed Rank test). For the *treated items*, results of a dependent *t*-test revealed that there was a statistically significant difference between treatment intensity conditions, with $t(23) = 2.698$, p (two-tailed) = .0128. For the untreated items, results of a dependent *t*-test revealed that there was not a statistically significant difference between treatment intensity conditions, with $t(23) = .0735$, p (two-tailed) = .942. The sample mean of treatment TRAD These tests indicate that for the *treated items*, the 2 week follow-up effect size sample mean for Traditional intensity treatment exceeded that for the Intense treatment. Conversely, for the *untreated items*, the 2 week follow-up effect size sample means did not differ by an amount great enough to exclude the possibility that any difference was due to random sampling variability.