Study Title: Treatments of Acquired Apraxia of Speech

NCT #: NCT01483807

Document: Statistical Analysis

Date of last study approval by IRB: November 22, 2015

Statistical Analyses

Effect Sizes

Six separate sets of effect size values were calculated for each participant as an indication of magnitude of change for *treated items*: 1) pre-treatment/baseline vs. end of SPT-R treatment phase – Random Phase Items – both subsets; 2) pre-treatment vs. 2-week follow-up – Random Phase Items – both subsets; 3) pre-treatment vs. 10-week follow-up – Random Phase Items – both subsets, 4) pre-treatment vs. end of SPT-B treatment phase – Blocked Phase Items – both subsets; and 5) pre-treatment vs. 2-week follow-up – Blocked Phase Items – both subsets; and 6) pre-treatment vs. 10-week follow-up – Blocked Phase Items – both subsets. This resulted in twelve Δ-values for treated items per individual. Six additional sets of effect size values were also calculated for untreated (generalization) items with comparisons being analogous to those of treatment items.

For the pre-treatment versus end of treatment comparisons, the five baseline probes immediately preceding application of treatment were used along with the final two probes completed during the treatment phase. For the pre-treatment versus two-week follow-up comparisons, the five baseline probes immediately preceding treatment were used with the 2-week follow-up probe that followed completion of that phase of treatment. For the 10-week comparisons, the five baseline probes immediately preceding the treatment phase were also used with the probe that occurred 10 weeks following the treatment phase. Because the experimental design was a cross-over design with only a two-week no-treatment interval, the 10-week follow up probes for the first treated sets fell at approximately two weeks after the conclusion of all treatment (i.e., there was not a 10-week no treatment interval and 10 weeks after the first treatment phase occurred about 2 weeks after the second treatment phase). For the second treated sets, the 10-week follow up

probes were those conducted at 10 weeks following all treatment. For the two participants with shortened treatment phases, the 10 week follow-up point for the first phase of treatment fell at approximately 6 weeks post all treatment.

The Δ-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 comparison values (i.e., end of treatment, 2 week follow up, or 10 week follow up).

For each participant, the two Δ-values for each subset within each experimental set were averaged to obtain one effect size per participant for each comparison condition. Tables 1 and 2 display the average effect size, SD of effect sizes, minimum and maximum effect sizes for the group of 20 participants for the treated items and generalization items, respectively. Relative to SPT benchmark effect sizes recently described by Bailey, Eatchel, and Wambaugh (2015) medium effect sizes were found for baseline to treatment phase comparisons for treated items for the group for both SPT-R and SPT-B (SPT-R = 9.203 and SPT-B = 7.389). For the two week follow-up phase comparison, a medium effect size was found for SPT-R (8.643) and a small effect size was found for SPT-B (6.861). For the 10 week follow-up comparison, small effect sizes were found for SPT-R (6.84) and SPT-B (6.144). For untreated items, small effect sizes were found for both SPT-R and SPT-B in all instances for the group (Table 4). 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 20 Participants*

Condition	**Comparison**	**Mean Δ-values**	**SD**	**Minimum**	**Maximum**
Random	BL vs Tx	9.203	4.344	2.766	18.244
	BL vs 2 wk FU	8.643	5.204	1.469	21.019
	BL vs 10 wk FU	6.84	3.99	1.66	18.78
Blocked	BL vs Tx	7.389	3.754	1.341	13.868
	BL vs 2 wk FU	6.861	3.533	2.109	15.199
	BL vs.10 wk FU	6.144	3.299	.830	14.378

BL = baseline; Tx = treatment; wk = week

Bailey et al. (2015) benchmarks: small = 5.23, medium = 6.98, large = 9.65 for baseline to treatment phase comparisons and 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* ***Generalization Items*** *by Treatment Condition for the Group of 20 Participants*

Condition	**Comparison**	**Mean Δ-values**	**SD**	**Minimum**	**Maximum**
Random	BL vs Tx	3.68	3.162	-.274	10.757
	BL vs 2 wk FU	3.452	2.711	-.483	8.367
	BL vs. 10 wk FU	2.556	2.615	-.293	9.839
Blocked	BL vs Tx	3.083	3.011	.0424	10.168
	BL vs 2 wk FU	3.074	2.865	-.771	8.721
	BL vs. 10 wk FU	2.107	2.466	-.771	7.826

BL = baseline; Tx = treatment; wk = week

Bailey et al. (2015) benchmarks: small = 2.25, medium = 3.75, large = 6.66 for baseline to treatment phase comparisons and small = 2.59, medium, = 4.23, large = 6.47 for baseline to follow-up phase comparisons

Paired t-tests with were utilized to compare the effect sizes obtained for SPT-R versus SPT-B. Normality tests were passed and one-tailed *p* values were obtained. There was not a significant difference in the sample means for ***treated item*** effect sizes associated with the end of treatment (p = .08) , 2-week follow-up phases (p=.105), or 10-week follow up phases (p=.254) for SPT-R treated items versus the SPT-B treated items.

For the ***untreated*** (generalization items), there were also no significant differences in the sample means for the three measurement conditions: end of treatment – p = .186; 2-week follow-up – p = .236; and 10-week follow-up – p = .212.

For all paired t-tests describe above, power was below the desired power of .80 and negative findings should be interpreted cautiously.

Percent Change Over Highest Baseline Performance

In order to provide clinically interpretable indicators of degree of change associated with treatment, percent change scores were calculated. Separate calculations were made for each experimental set of items separated by subgrouping (SPT-R treated, SPT-R untreated, SPT-B treated, SPT-B untreated). For each participant, the highest percent accuracy score achieved in baseline was subtracted from the a) last treatment probe, b) the 2 week follow-up probe, and c) the 10 week follow-up probe. Tables 3 and 4 depict the average change scores for the group for the treated and untreated items. A negative value in Tables 3 and 4 indicates that the follow-up value was less than the highest baseline value. Note that the median scores for the group were all positive as were the 25th percentile values.

Table 3

*Percent Change Over Highest Baseline: Descriptive Statistics for **Treated Items** by Treatment Condition for the Group of 20 Participants*

Condition	Comparison	Mean	SD	Minimum	Maximum	Median	25th %tile	75th %tile
Random	BL vs Tx	65.5%	20.7%	20%	100%	60%	50%	80%
	BL vs 2 wk FU	58.5%	29%	0%	100%	60%	40%	87.5%
	BL vs 10 wk FU	46.25%	32.1%	-20%	100%	50%	20%	85%
Blocked	BL vs Tx	51%	29.8%	-10%	100%	50%	30%	80%
	BL vs 2 wk FU	45.8%	29.5%	-20%	100%	40%	20%	70%
	BL vs 10 wk FU	38.0%	26.3%	-20%	80%	40%	12.5%	60%

BL = baseline; Tx = treatment; wk = week

Table 4

*Percent Change Over Highest Baseline: Descriptive Statistics for **Generalization (Untreated) Items** by Treatment Condition for the Group of 20 Participants*

Condition	Comparison	Mean	SD	Minimum	Maximum	Median	25th %tile	75th %tile
Random	BL vs Tx	24.3%	41.7%	-67%	100%	20%	0%	60%
	BL vs 2 wk FU	23.8%	40.5%	-67%	100%	37%	0%	55%
	BL vs 10 wk FU	14.15	35.2%	-60%	100%	0%	0%	30.5%
Blocked	BL vs Tx	24.5%	41.4%	-33%	100%	20%	0%	55%
	BL vs 2 wk FU	22.7%	33.8%	-40%	100%	20%	0%	40%
	BL vs 10 wk FU	12.5%	32.7%	-60%	80%	0%	0%	34%

Comparisons of average group percent change were completed using parametric statistics (i.e., dependent t-tests, one-tailed) when possible and the nonparametric analogs when normality tests were failed (i.e., Wilcoxon Signed Rank tests). The Wilcoxon Signed Rank test indicated that the percent change in SPT-R treated items (65.5%) was statistically significantly greater than the change in SPT-B treated items (51%) for the last probe of the treatment phase: $p = .022$. Similarly, the sample mean percent change for SPT-R treatment items at 2-week follow-up (58.5%) was statistically significantly greater than the sample mean percent change for SPT-B treatment items at 2-weeks follow-up (45.8%): $p = .016$. For the 10-week follow-up comparison, a paired t-test was employed; the sample mean percent change of the SPT-R treated items (46.25%) exceeded the mean percent change of SPT-B items (38%) by an amount that is greater than would be expected by chance: $p = .043$. For the generalization items, no statistically significant differences were found for any of the measurement times: last treatment probe ($p = .488$), the 2-week follow-up probe ($p = .477$) and the 10-week follow-up probe ($p = .396$). As with the effect size data comparisons, the comparisons involving the percent change data were underpowered.