Text: 1200 words

Introduction

Treatment efficacy is an important component of evidence-based research. The efficacy of speech treatments are typically evaluated by measuring changes in speech production: phonemic production, acoustic output, speech intelligibility, comprehensibility, and/or physiology. One component of treatment efficacy is social validity. Social validity is a measurement of the social significance of a treatment. Few treatments for aphasia and/or apraxia of speech have established social validity. One study was the social validity of a treatment technique called Script Training (Youmans et al., 2005).

Script training is a relatively new treatment approach that is used to inject automatic "islands" of speech into the communication of individuals with aphasia and apraxia. A patient selects three or four scripts that are personally meaningful to her at the time of treatment. The participant then practices each script one phrase at the time until it becomes automatic. The practice consists of repeating the phrase together with the clinician, choral production of the phrase, initial phonemic cuing of the phrase and then independent production of the phrase. When the participant masters the phrase and can generate it spontaneously, a new phrase is selected and learned.

To help establish social validity, researchers often randomize audio or video clips of a client taken both before treatment and after treatment, and ask several individuals to rate a specific behavior across the clips. To date, social validity has been largely conducted by having a group of "blind" raters judging pre- versus post-treatment samples of a client's speech in a randomized order. This method allows the investigators to determine the overall social benefit of a treatment compared to baseline. However, this approach does not enlighten us as to how the specific changes in a client's speech individually changed the raters' perceptions. For example, what aspect of the client's speech improved the raters' judgments about her naturalness?

The purposes of the study were to: Determine if the script

training treatment applied to a woman with apraxia of speech was socially valid; Determine which specific aspects of the woman's speech (correctness, number of errors, speech rate) corresponded to varying rater judgments regarding the understandability, ease of production, naturalness, and overall quality of her speech; Determine which of the variables (understandability, ease of production, naturalness) corresponded to the highest overall quality rating

Method

Participants

One-hundred-twenty-four raters from 20-40 participated. Their naïve status regarding aphasia, apraxia of speech and neurogenic speech disorders was established through a brief questionnaire, which also requested their age, level of education, profession or area of study, and self-reported hearing status.

Stimuli

Twelve audio clips were used as stimuli. The clips were extracted samples of recordings made during various phases of baseline and treatment using a Script Training method to improve the speech of a woman with apraxia. The clips were selected based on the subject's percentage of the script she produced correctly (%correct), errors produced while attempting to produce the script (including repetitions and unintelligible words), and rate. Three levels of %correct were included: low (0-32% correct), medium (33-66% correct), and high (67-100% correct). Two levels of errors were included: relatively low number of errors (0-15) and high (16 – 30). Two levels of rate were included: slow (0 – 26 words per minute) and faster (26+ words per minute).

The 12 audio clips represented a sample from each of the conditions. For example, one clip included an excerpt with a low %correct, low number of errors, and a slow rate. Another sample included an excerpt with a high %correct, high number of errors, and a fast rate. The 12 audio clips were then randomized and put on a compact disk with a space between clips.

Procedures

Following informed consent and administration of the questionnaire, raters were presented with the 12 audio excerpts.

Instructions were similar for all raters. Raters were asked to mark, on a continuous line, their subjective rating of the behavior in question. Behaviors of interest were understandability, ease of production, naturalness of speech, and overall quality of speech.

Data Analysis

The line length before the rater's mark was measured in millimeters to obtain quantitative values for each of the dependent measures for each of the conditions. An RM-MANOVA was computed to determine whether statistically significant main effects or interactions exist for %correct errors, and/or rate on each of the dependent variables. Additionally regression was computed between the overall quality rating and the other ratings to determine which aspects were most important.

Results

All main effects and interactions were significant. Multivariate (Wilks' Lamda) and univariate (Greenhouse-Geisser) statistics, as well as all of the pairwise comparisons were significant at the $p<0.001$ level. As %correct increased, the ratings for understandability, ease of production, naturalness, and overall quality increased significantly. Significant differences were found between low %correct, medium %correct and high %correct. As Errors increased, the ratings for understandability, ease of production, naturalness, and overall quality increased significantly. As Rate increased, the ratings for understandability, ease of production, naturalness, and overall quality increased significantly

Additionally, significant, positive correlations were found between the overall quality of speech and Understandability ($r = 0.85$; $p <0.0001$), Ease of production ($r = 0.86$; $p <0.0001$), and Naturalness ($r = 0.85$; $p <0.0001$). Regression analysis indicated an R2 change of 0.83 ($p <0.0001$). All effects and partial correlations were significant at the $p <0.0001$ level. Effect estimates and partial correlations for each variables are as follows: understandability (0.40; 0.50), ease of production (0.24; 0.24), and naturalness (0.34; 0.37).

Discussion

The results for %correct were intuitive. That is, as the patient's

ability to produce the script increased, the participants rated her better across all domains. Likewise, the results for Rate made sense. As the patient's rate of production changed from extremely slow to slow, the participants rated her better across all domains. However, at first, the Error results appear counterintuitive: as the patient produced more errors, the participants rated her higher.

A case analysis was conducted on the Error data. Not all data went in the same direction. In other words, although in general as errors increased, ratings increased, sometimes as errors increased, ratings went down. Whether errors were seen positively or negatively appeared to be dependent on the type of error. Real word repetitions, interjections, and "empty" speech appeared to be seen as positives. These errors may benefit the individual because they may look like fewer struggles due to less hesitations and more speech output. When unintelligible utterances and long pauses comprised the error set, the errors were perceived as negative based on ratings.

The regression analysis revealed that the perceived understandability of speech, the ease of speech production, and the naturalness of speech were significant predictors of overall speech quality perceptions. All of the preceding variables were highly related to ratings of overall quality; however, understandability appeared to be most important, followed by naturalness, and finally effort

Conclusions

Script training appeared to be socially valid. Naïve listeners appeared to be sensitive to the amount and the quality of speech output generated by the speaker. Naïve listeners appeared to perceive: struggle behaviors as negative, increased speech output (including repetitions and empty speech) as positive, faster speech (closer to a normal speaker's average word per minute) preferable to slow speech, and more understandable speech as better quality speech.

References

Foster, S., Mash, E.(1999). Assessing social validity in clinical treatment research issues and procedures. Journal of Consulting and Clinical Psychology, 67(2), 308.

Goldsten, H. (1990). Assessing clinical significance. In L. Olswang, C. Tompson, S. Warren and N. Minghetti (eds) Treatment Efficacy Research in Communication Disorders (Rockville, Md:ASHA Foundation), pp. 91-98.

LaPointe, L., Katz, R., Braden, C. (1999) Clinical Significance of change in Language performance:social validation of writing response improvement in aphasia. Aphasiology, 13(9-11), 787.

Schwartz, I., Baer, D. (1991). Social Validity assessments: is current practice state of the art? Journal of Applied Behavior Analysis, 24, 189.

Wolf, M. (1978). Social Validity: the case for subjective measurement or how applied behavior analysis is finding its heart. Journal of Applied Behavior Analysis,11, 203.

Youmans, G. L., Holland, A. L., Munoz, M., & Bourgeois, M. (2005). Script training and automaticity in two individuals with aphasia. Aphasiology, 9 (3,4,5).