

# A Comparison of Outcome Measures for Speech Motor Learning in Acquired Apraxia of Speech Using Motor Learning Guided Treatment

Rachel K. Johnson, Aileen Lott, Jessica Prebor

*Department of Communication Disorders & Special Education, Old Dominion University, Norfolk, United States*

**Purpose:** The purpose of this study was to investigate potential benefits of using a qualitative and quantitative outcome measure of articulation accuracy and suprasegmental characteristics in isolation for speech motor learning in acquired apraxia of speech (AOS).

**Methods:** Baseline, retention, and maintenance measures from an oral reading task of 2 speakers with chronic AOS and aphasia were rated using an 11-point multidimensional rating scale accounting for articulation and immediacy and a hybrid scale measuring number of correctly produced words, presence of distortions in correctly produced words, and immediacy of the production. Participants received motor learning guided treatment two days a week for eighteen sessions.

**Results:** The multidimensional rating scale and the hybrid scale comparably represented speech motor changes related to articulation accuracy and immediacy of the production across the duration of the intervention. The hybrid scale provided a sensitive measure for individual differences in immediacy and presence of distortions not represented in the multidimensional rating scale.

**Conclusions:** The results of this pilot study provide evidence to support the benefit of using a qualitative and quantitative outcome measure for speech motor changes in acquired AOS. The individual differences identified through the hybrid scale have clinical and research implications.

**Keywords:** Apraxia of speech, Speech motor learning, Treatment outcome measures, Intervention



**Received:** February 15, 2018

**Accepted:** April 13, 2018

**Correspondence:**

Rachel K. Johnson

Communication Disorders & Special Education, Old Dominion University, Child Study Center, 4501 Hampton Blvd, Norfolk, VA 23529, United States

Tel: +757-683-6285

Fax: +757-683-5593

E-mail: R1johnson@odu.edu

© 2018 The Korean Association of Speech-Language Pathologists

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## INTRODUCTION

Acquired apraxia of speech (AOS) in isolation following a stroke is a rare occurrence; it is often accompanied by aphasia and/or dysarthria. Distinguishing characteristics of AOS are sound substitutions and sound distortions, an increase in these distortions with an increased rate and/or increased complexity, articulatory inaccuracy related to place and manner, and a disproportionate number of words per minute relative to maximum sustained vowel duration [1,2]. Most AOS treatment protocols are characterized as articulatory-kinematic approaches based on the theoretical framework that AOS is a phonatory-motor disorder. This distinction is relevant given the diversity of outcome measures reported across treatment studies.

Outcome measures reported in the literature vary from binary systems to multidimensional rating scales. Each system comes with its own advantage and disadvantage

to capture speech motor changes that best illustrate treatment effectiveness for AOS. The binary system is simple and straightforward, it primarily accounts for the articulatory accuracy of the production. While the binary system is straightforward, differences in the operationalization of accuracy are reported across research studies. Wambaugh and colleagues [3-5] and Knock et al. [6] operationalize a correct production as one with no errors in consonant or vowel productions (including distortions) and no syllable addition, omissions, or repetitions. Other variations of accuracy have been operationalized as error free phonemes (no distortions), appropriate stress pattern and normal rate [7] or simply accurate and intelligible productions [8]. Another variation of a binary outcome measure is the agreement of two independent judge's broad transcription of the target in which the distortions were considered correct if both judges transcribed the distortion as the intended phoneme [9]. Researchers commonly report binary outcome measures as a percentage of accuracy. The advantage to the binary system is its simplicity; however, AOS, by definition, is more than an articulation disorder. By not accounting for suprasegmental characteristics, such as length of intersyllabic intervals, the binary system alone is inadequate to fully capture the speech changes characteristic of AOS.

Reported outcome measures using an interval scale include an additional descriptive element of change and/or account for partial accuracy. Both acquired and developmental AOS treatment studies report a three-pronged measure described as correct, approximation/partial accurate response, or incorrect [10,11]. More complex scales such as multidimensional rating scales operationalize groupings of intelligibility, articulation, immediacy, and/or fluency across a continuum to describe changes in speech across time [12-15]. Studies using interval scales report outcome measures as a mean score or a percentage of targets that met a designated criterion. The advantage of a multidimensional rating scale is the inclusion of articulation and other speech characteristics affected by the impaired sensorimotor system. The challenge with using the multidimensional rating scale is the subjectivity of the rating when the changes in speech behavior do not occur congruently as grouped in the scales. The subjective nature of the measure increases the likelihood for clinical disagreement particularly when considering influential factors of co-existing aphasia behaviors and familiarity of the stimuli.

Diagnostic criteria for AOS includes characteristics for sound distortion and lengthening of sound segments or pauses between segments [1,2,16-18]. During recovery, the

speech motor changes for intelligibility, articulation, and immediacy do not always change the same way as grouped together in the multidimensional rating scale. To better understand the recovery pattern and influential factors on speech motor changes, a more sensitive tool to measure those changes is needed. Individual differences along with severity, time post-onset, and lesion location can potentially serve as diagnostic indicators and advance therapeutic protocols. The purpose of this study was to investigate if rating articulation accuracy and suprasegmental characteristics in isolation would contribute information to advance understanding of speech motor changes during recovery by identify differences across individuals.

To do this, a hybrid scale was developed to measure changes in the primary distinguishing characteristics of AOS. The hybrid scale, uses a combination of quantitative and qualitative measures to account for articulation, presence of distortions, and immediacy of the production. The operationalization of these measures was based on binary measures previously described and outcomes measures reported in Youmans and Youmans [19]. Articulation accuracy is considered an indicator for motor programming and planning [2]. The hybrid scale uses two measures for the articulation parameter. One measure quantifies the articulation accuracy independent of sound errors to account for intelligibility and preservation of the intended meaning. A separate binary measure indicates the quality of the articulation based on the presence of sound errors secondary to distortions, additions, substitutions, or omissions. Initiation and fluency in a production is an indicator of automaticity of skilled performance [19]; therefore, the scale includes a binary measure for immediacy. This study compares the newly developed hybrid scale to a previously described multidimensional rating scale in two patients with chronic acquired AOS and aphasia receiving intervention using the motor learning guided (MLG) treatment. It is hypothesized that using a separate but simple measure that quantifies and qualifies changes in articulation and immediacy during intervention will result in a more meaningful outcome measure. This is a pilot study to explore proof of concept for research and clinical application.

## METHODS

### Participants

Participants in this study were a 68 year-old male (P1) who was 91 months post onset of a left hemisphere stroke and a 61

**Table 1.** Descriptive data and pre-treatment measures for P1 and P2

Characteristic	P1	P2
Age	68	61
Gender	Male	Male
Months post onset	91	86
Site of lesion	Unknown	Unknown
Years of Education	16	12
Former Occupation	Insurance agent	Investment broker
Premorbid Handedness	Right	Right
Apraxia Battery for Adults [20]	Mod	Mod-Severe
AOS Characteristics (Apraxia of Speech Rating Scale [2])	Score (0-4)	
Distorted sound substitutions	3	3
Distorted sound additions (not including intrusive schwa)	0	0
Increased sound distortions or distorted sound substitutions with increased utterance length or increased syllable/word articulatory complexity	3	4
Increased sound distortions or distorted sound substitutions with increased speech rate	3	4
Inaccurate (off-target in place or manner) speech AMR's (alternating motion rates, as in rapid repetition of "puh puh puh")	3	3
Reduced words per breath group relative to maximum vowel duration	0	0
Western Aphasia Battery-Revised [21]		
Aphasia Quotient (100)	72.8	66.8
Aphasia Classification	Transcortical motor	Broca
Spontaneous speech (20)	11	11
Auditory verbal comprehension (10)	9.8	9.6
Repetition (10)	8.2	5.6
Naming (10)	7.4	7.2
Reading (20)	16.8	12

year-old male (P2) who was 86 months post onset of a left hemisphere stroke (Table 1). Both participants had normal or corrected to normal vision and demonstrated functional hearing in a quiet environment. Presence and severity of AOS was determined using the Apraxia Battery for Adults-2 (ABA-2; 20) and Apraxia of Speech Rating Scale (ASRS; 2). Both participants demonstrated sound distortions, substitutions, and omissions that increased with articulatory complexity and increased speech rate, and inaccurate speech alternating motion rates (AMR's). Western Aphasia Battery-Revised (WAB-R; 21) was administered to characterize aphasia. P1 was characterized as having moderate AOS and transcortical motor aphasia and P2 was characterized as having moderate to severe AOS and Broca's aphasia. Reading competency was determined to be functional at the simple sentence level as determined by performance on the reading subtests on the WAB-R. Written informed consent according to Institutional

Review Board approval was obtained prior to participation in this study.

## Procedure

### *Stimulus item selection*

Two sets of 5 phrases were used for stimuli (Table 2). The stimuli in set 1 was used for the daily oral reading retention measure (no model or feedback provided). The stimuli in set 2 was included in every fifth retention measure. Each set was created using a sentence template completed with personal information obtained in a structured interview with participant and family member. The content of the stimuli were individualized based on participant's interests. The researchers attempted to create two sets of stimuli that were similar in terms of length and complexity with the primary emphasis of functionality [16]. The level of complexity was determined based on the overall severity and performance on the multi-

**Table 2.** Stimuli for P1 and P2

P1: Set 1	P1: Set 2
In October, we spend Halloween with my four grandchildren.	In December, we go to New Jersey to visit X and his family.
My favorite TV shows are Xxx and Xxx X.	Swimming freestyle and backstroke are my favorite exercises.
X, X, and I enjoyed golfing last weekend at Xx X.	Last weekend, we went to Mass at X. X's in Xxx X.
Would you like to go to X Xx for dinner tonight?	Would you like to go to Xx X amusement park later?
I enjoy playing the lottery on Mondays and Thursdays every week.	On Fridays, I enjoy swimming at the Xx Beaches.
P2: Set 1	P2: Set 2
I would like sweet tea please.	I live by the airport.
I live with my son Xx.	I watched the news yesterday.
I watched some golf yesterday.	Last weekend I did yard work.
Xx and I went fishing last weekend.	Freddy Couples is my favorite golfer.
Golfing is my favorite sport.	I would like to work on the pool.

X used to replace any identifying information.

**Table 3.** Description of multidimensional rating scale

Rating	Articulation	Immediacy	If Aphasia characteristics present
11	Accurate articulation, intelligible	Immediate production of all elements of utterance	
10		Delayed production ( $\geq 2$ seconds) of some elements of production (searching, groping)	
9	Distortions, sound addition or omissions	Immediate	
8		Delayed	
7	Incomplete articulation (missing elements of production but does not interfere with general message)	Immediate	- Morphological inflections omitted - Use of stereotypic phrase/recurring utterance but, use is appropriate and meaning of message is maintained
6		Delayed	
5	Self-correction successful		
4	Incomplete articulation (missing crucial elements of production so that utterance is not intelligible)	Immediate	Omission of: - Articles - Conjunctions - Prepositions - Auxiliary verbs - Pronouns
3		Delayed	
2	Perseverative error, wrong target	Immediate	Inappropriate use of stereotypic phrase/recurring utterance
1		Delayed	

syllable subtests of the ABA-2 [20]. For P1, each set of stimuli ranged from 14-19 syllables, with an average of 16 syllable items for each list. For P2, each set of stimuli ranged from 6-11 syllables, with an average of 7.8 syllable items for each list.

### **Experimental design**

Multiple baseline design across participants and behaviors (oral reading) was employed. The daily retention measure across sessions was an oral reading task (no model or feedback) at the beginning of each treatment session prior to be-

ginning the treatment protocol. Treatment sessions were 30 minutes two times a week for nine weeks for a total of 18 treatment sessions. The index of speech motor learning was determined by changes in speech production performance during the retention measure. Productions were scored online and from videotaped recordings using a Panasonic HC-V750 video recorder. The initial retention measures were rated using an 11-point multidimensional rating scale (Table 3). A criterion for performance based on an 11-point multidimensional rating scale was set at a mean rating  $> 10$  on the retention mea-

sure for 3 consecutive sessions. If criterion was met for set 1 phrases, training for set 2 phrases began until criterion was met or intervention period ended (whichever came first).

### ***Probing schedule***

Baseline speech productions of set 1 and set 2 phrases were obtained in an oral reading task (no model or feedback) at the beginning and end of the second session of pre-treatment testing. The third baseline was obtained on the first day of treatment before treatment began. Treatment initiated following the third baseline measure. Baseline was extended on set 2 of phrases to serve as the control. Due to the behavioral performance and scheduled treatment period, only P1 completed training on set 2 phrases.

### ***Treatment protocol***

Each treatment session began with a retention measure in a random elicitation of the treatment phrases in an oral reading (no model or feedback) task as an index of speech motor learning. Set 2 phrases were included in every fifth session. Following the retention oral reading task, treatment began using the motor learning guided (MLG) protocol previously described [14, 24]. The written stimuli were presented via a programmed PowerPoint presentation with a set-pause time (4 seconds or 10 seconds; dependent on the protocol stage) between each visual presentation of the stimulus. Stimuli were practiced in a random order in each stage of every treatment session.

There were three stages to the MLG treatment protocol. A 2-minute break was imposed between each stage of the treatment protocol. Stage 1 began with the written presentation of the stimuli accompanied by a clinician model. After the model, the participant produced the stimuli followed by a blank screen for 4 seconds. The participant's productions and pause-time was repeated 3 times. After the third production, the clinician provided a modeled production of the stimuli followed by knowledge of results feedback (e.g. "I heard changes with each one." "which one do you think was closest?" "the x one was closest."). This was repeated for all five of the stimuli. For Stage 2 of the treatment protocol, the process was the same with the exception of no initial clinician model. In Stage 3 of the treatment protocol, the process was the same as Stage 2; only the pause-time between productions was increased to 10 seconds. The extended pause-time was used to increase the contextual interference otherwise fulfilled in a sequential presentation when more than five stimuli are trained.

### ***Dependent variables***

Speech productions during baseline, retention, and maintenance measures were initially rated using an 11-point multi-dimensional rating scale as previously described [13,22,23]. Productions were judged on perception of production's overall articulation accuracy, intelligibility, and immediacy (Table 3). A rating was assigned based on a paired articulation and immediacy parameter with the highest number representing the most accurate and immediate. Articulation criteria were based on accuracy, presence of distortions, completeness for intelligibility, and perseverations. Each articulation parameter was further characterized as immediate or delayed. A production was considered delayed if a  $\geq 2$  seconds pause occurred in any element of the production. The highest rating of 11 was assigned to productions free of distortions, intelligible and immediate; a 10 was assigned if a delay was present in any element of the production. If distortions, sound additions, or omissions were present a 9 was assigned if immediate, an 8 if delayed. Productions with missing elements but the general message was unaffected were rated as 7 if immediate and 6 if delayed. This grouping included a parameter to account for co-existing aphasia characterizing omission of morphological inflections and stereotypic utterances that do not affect the intended meaning of the message. Individuals with AOS will often attempt to self-correct; however, the self-corrections are not always successful. To account for the success of the self-correction a rating of 5 was assigned if the self-correction resulted in an accurate production. If the self-correction contained articulation errors, the production was rated accordingly. This assignment was used to account for the underlying impairment in the motor control system. Rating of 4 (immediate) and 3 (delayed) were assigned for utterances that were missing crucial elements including omissions typical of co-existing aphasia, resulting in an unintelligible utterance or the meaning of the message being maintained. Perseverative productions or wrong target including off topic stereotypic utterances were assigned a 2 if immediate and 1 if delayed. The mean of the rating for all stimuli was reported for baseline, retention, and maintenance measure.

A post-hoc rating of baseline, retention, and maintenance measures from videotaped recordings was completed using the hybrid rating scale. The hybrid rating scale provided three separate measures; number of correctly produced words, presence of distortions in correctly produced words, and immediacy of production. Each word in the stimulus was judged for articulation accuracy using criteria adopted from You-

mans and Youmans [19] script training for individuals with AOS. An approximation of the word was considered correct if the sound error (distortion, addition, substitution, or omission) did not change the syllable count or the meaning of the stimulus and if the production was intelligible. Stereotypic productions or substitution of words with a similar meaning were considered errors. For the words judged correct, the presence of distortions was indicated with a binary measure. The third measure evaluated automaticity and motor control of the production. The same criteria for immediacy was used as the 11-point multidimensional rating scale: a production was considered delayed if there was a 2 seconds or longer delay of some element of the production, attempts for self-correction or searching/groping behaviors observed. Unlike the distortion measure, immediacy was independent of articulation accuracy. By doing this, each parameter could be analyzed separately as an indicator for control of stereotypic productions for individuals with co-existing aphasia. Baseline, retention, and follow-up measures were rated according to number of correct word approximations, presence of additions, substitutions, and/or distortions (0 = present; 1 = none), and immediacy (0 = a delay ( $\geq 2$  seconds), restart, or groping behaviors were present in any part of the production; 1 = immediate) of the productions (Table 4). A total score was calculated based on the number of correct word approximations+

distortion score+ immediacy score. A proportion of accuracy in each parameter was calculated for each measure (baseline, retention, maintenance) to represent speech motor changes.

#### Data analysis

Data were analyzed with both visual inspection and effect size calculations. In addition, the mean treatment gain for each treatment cycle was calculated by subtracting the mean of the last three treatment retention measures from the mean of the three baseline measures. The standard mean difference (d) as described in Beeson and Robey [24] is reported as a conservative estimation of effect size. Treatment effect sizes and follow-up effect sizes were calculated using excel according to guidelines provided in Bailey et al [25].

#### Reliability

The treating clinicians scored the utterances on the oral reading tasks during baseline, retention, and maintenance measures. A blind 20% of the recordings were rerated by the treating clinician for the intra-rater reliability. Inter-rater reliability was performed on a blind 20% of the recordings by one of the authors who did not complete the initial rating. All reliability ratings were performed following the completion of the study. Reliability using Krippendorff's alpha on an interval scale was  $\alpha$  at 0.75 for multidimensional rating scale; the hybrid scale

**Table 4.** Scoring procedure for hybrid rating scale

Number of correct word approximations	A production of a word was considered correct if it had the same number of syllables and the word was recognizable even if articulation errors were present. Score of 0: <ul style="list-style-type: none"> <li>• If no words were produced correctly (0 for entire production)</li> <li>• For the word if the word was clear but, not the target</li> <li>• If an omission/deletion/addition changed the number of syllables (e.g. -ing)</li> </ul> The score was number of correct word approximations according to the stimulus.
Distortions	Score 0 if: <ul style="list-style-type: none"> <li>• There was an error in any word in the stimuli</li> <li>• Presence of distortion, addition, or substitution in any part of the utterance.</li> <li>• Score of number of correct word approximations was 0.</li> </ul> Score of 1 if: <ul style="list-style-type: none"> <li>• The number of correct words was the maximum score for stimulus and NO distortions present in words produced correctly.</li> </ul>
Immediacy	Score 0 if: <ul style="list-style-type: none"> <li>• Any part of the production was delayed equal to or more than 2 seconds.</li> <li>• Production was attempted more than once (i.e. self-corrected) regardless of accuracy of production.</li> <li>• Evidence of searching or groping behaviors or fillers during utterance</li> </ul> Score of 1 if: <ul style="list-style-type: none"> <li>• All of utterance was produced with less than 2 seconds pause during any part of the production independent of articulation accuracy.</li> </ul>

words correct was  $\alpha$  at 0.97 and total was  $\alpha$  at 0.97, indicating good reliability for the scale measures. Point-to-point agreement for judgment of binary measures for the hybrid scale was calculated by determining the number of agreements between the original scorer and the second and third scorer divided by the total number of stimuli scored in that retention. Good overall agreement was found for distortions at 95% and immediacy at 89%.

## RESULTS

### Multidimensional rating scale

Figure 1 illustrates the speech motor changes during intervention for P1 according to the multidimensional rating scale.

The mean baseline rating for P1 was approximately 8.27 for set 1 and 7.93 for set 2. Productions were described as having incomplete articulation and distortions and delayed elements. Figure 2 illustrates the speech motor changes during intervention for P2 according to the multidimensional rating scale. The mean baseline rating for P2 on set 1 was 5.47 and 6.60 for set 2. Productions were described as incomplete articulation maintaining the general message for some but not all and delays were prevalent for some elements of the production. A trending baseline was observed for both participants; however, a decrease in performance occurred immediately following treatment supports the expectation that changes were secondary to the intervention.

After training started on set 1, P1 demonstrated a stable

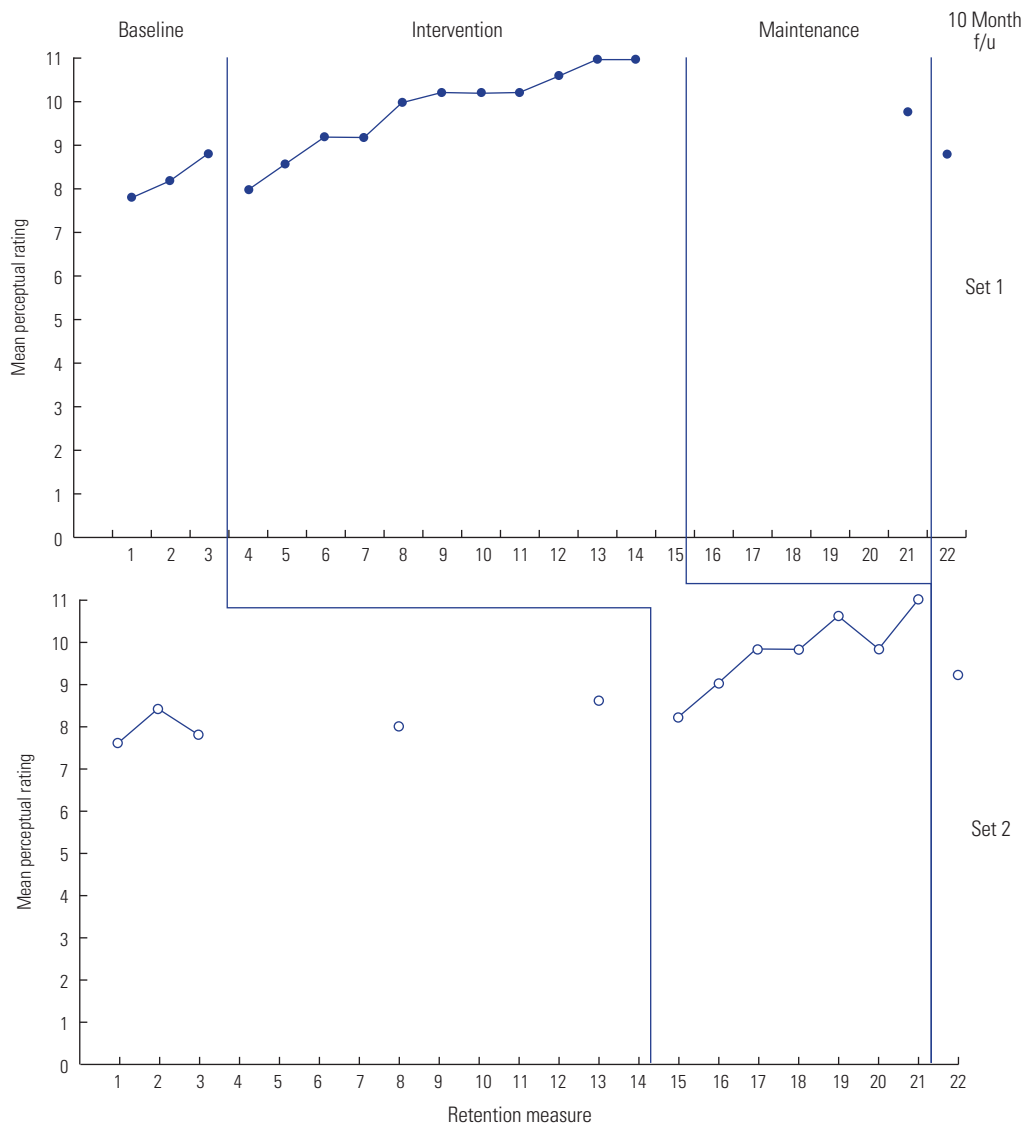


Figure 1. Outcome measures for P1 using the multiple dimension rating scale.



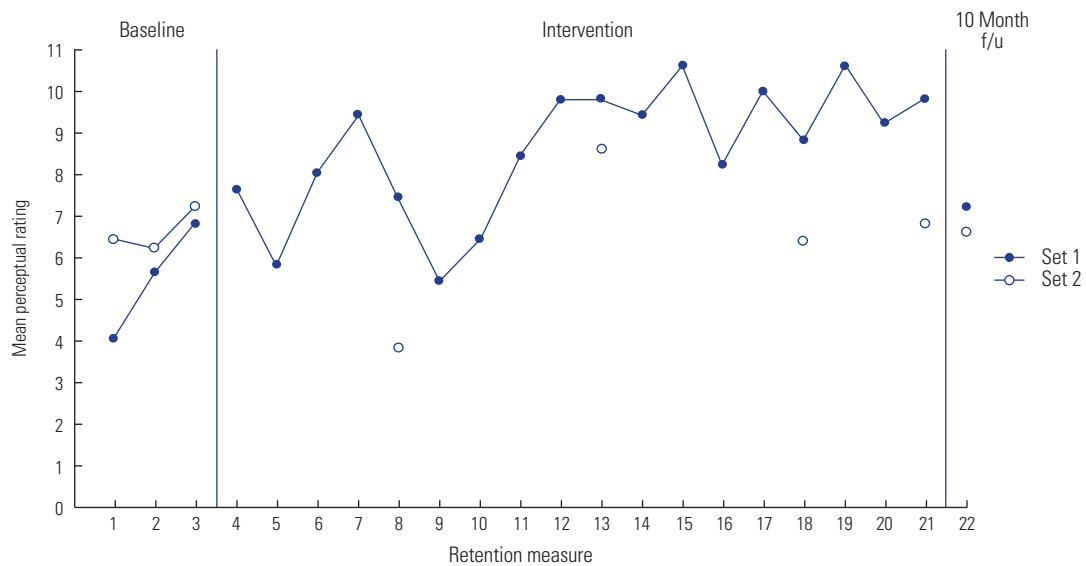


Figure 2. Outcome measures for P2 using the multiple dimension rating scale.

Table 5. Effect sizes (d) for P1 and P2 comparing multidimensional rating scale and hybrid scale

	MDRS	Words correct	Hybrid scale		Total	MDRS	Words correct	Hybrid scale		Total*
			Distortion score	Immediacy score				Distortion score	Immediacy score	
P1			Set 1			Set 2				
Mean baseline	8.27	48.33	0**	0.20*	48.33	7.93	51.33	0**	0.20**	51.33
SD	0.50	3.79	0	0.45	4.16	0.42	2.52	0	0.45	2.52
Treatment										
Mean retention	10.87	54	0.67	4.67	59.33	10.47	53	1.0	4.00	58
Effect size	5.17	1.50	-	<b>9.99</b>	2.91	<b>6.08</b>	0.66	-	<b>8.50</b>	2.65
Control										
Mean retention						8.30	50.50	0	0.50	51
Effect size						0.88	-0.33	-	0.67	-0.13
Maintenance										
Mean retention	8.8	54	0	3	57	9.2	51	1	3	55
Effect size	1.06	1.50	-	<b>6.26</b>	2.29	3.04	-0.13	-	<b>6.26</b>	1.46
P2			Set 1			Set 2				
Mean baseline	5.47	23	0.20**	1.0	23.33	6.60	18.33	0.14	0.33	18.67
SD	1.40	1	0.45	2.0	0.58	0.53	2.08	0.35	0.58	2.52
Treatment										
Mean retention	9.87	28	1	3	32					
Effect size	3.13	5	1.79	4.62	<b>15.01</b>					
Control										
Mean retention						5.12	16.6	0.20	0.40	17.0
Effect size						-2.80	-0.83	0.16	0.12	0.66
Maintenance										
Mean retention	7.20	23	0	0	23	0	23	0	0	23
Effect size	1.23	0	0.45	-0.58	-0.58	-12.47	2.24	-0.40	-0.58	1.72

MDRS, multidimensional rating scale; SD, standard deviation; Mean Retention, last 3 retention measures of treatment cycle.

\*Equals the sum of words correct, distortion score and immediacy score; \*\*Indicates extended baseline measure used; Bold indicates significant effect size.



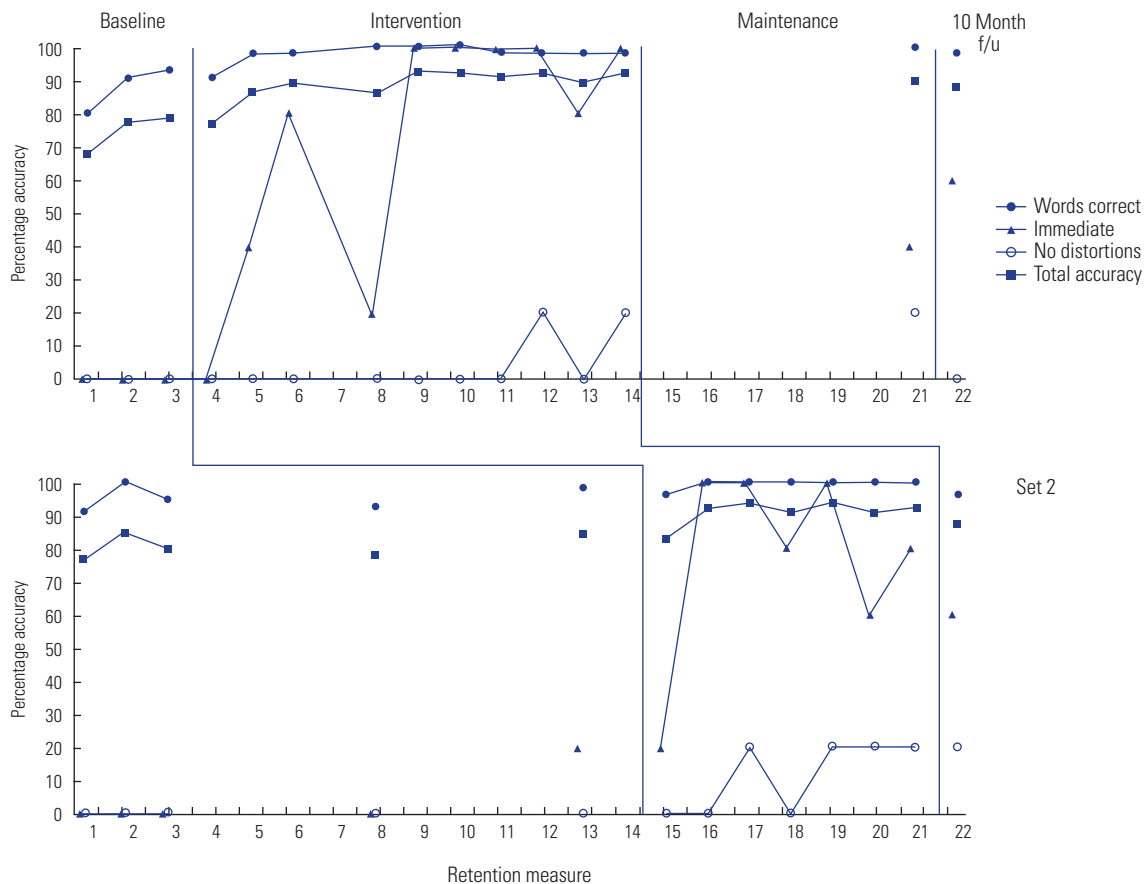
mean retention rating achieving criterion for mastery by training day 11 with a mean of 10.87. The same steady increase occurred after the initiation of training on set 2 reaching a mean of 10.47 before the end of the intervention period. Speech productions were described as accurate, intelligible, and immediate for some but not all elements of the stimuli. For P2, the mean retention ratings on set 1 never stabilized during the intervention. Changes in speech motor learning were evident with a mean retention rating of 9.87 by the end of the intervention, although not meeting criterion for mastery. P2 demonstrated inconsistency in performance throughout the intervention, evidenced by productions containing distortions, omissions and substitutions and delays present in some elements of the productions. The specific contributing factors for the instability in performance cannot be determined from this outcome measure due to the grouping of distortions and immediacy. The mean treatment gain was calculated by subtracting the mean rating for the last 3 retention measures from the mean rating for the 3 baseline measures. The mean treatment gain for P1 was 2.6 and 2.54 for set 1 and 2 respectively.

The mean treatment gains were higher for P2 at 4.40 for set 1.

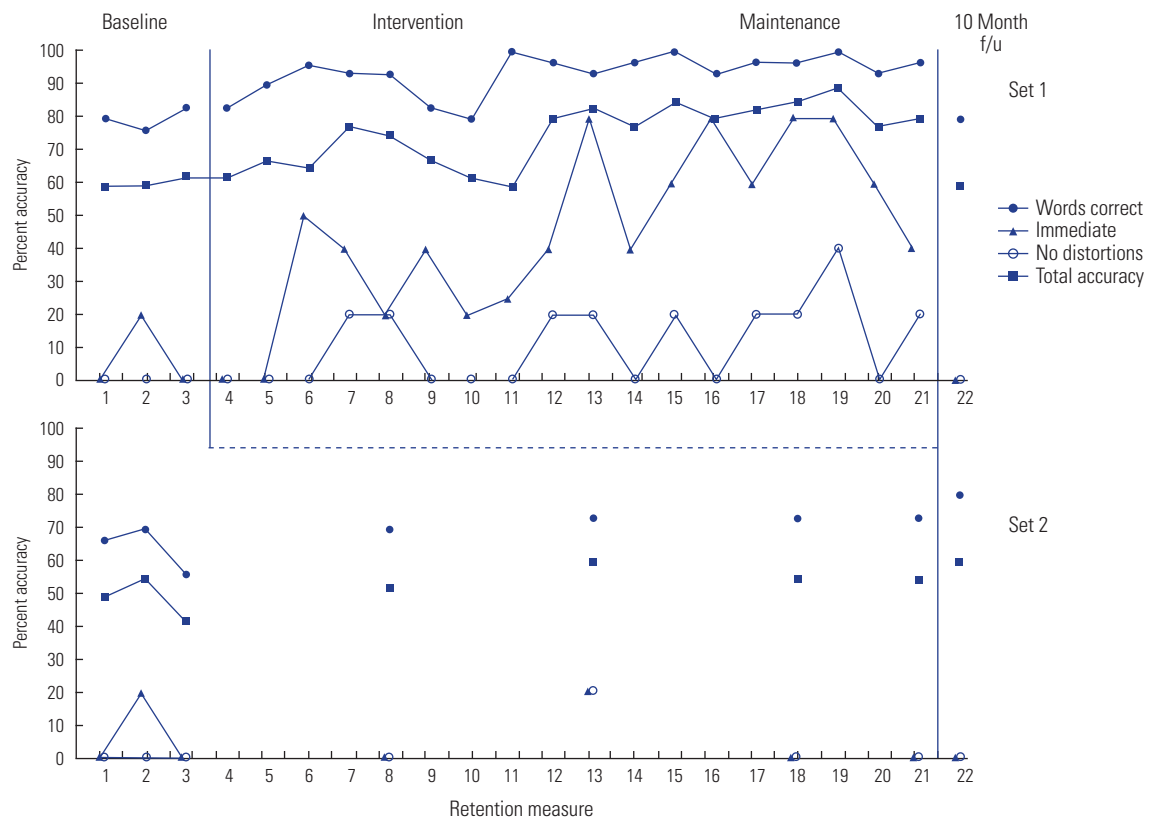
The standard mean difference (d) as described in Beeson and Robey [24] was calculated to provide an estimated treatment effect size (Table 5). According to effect size guidelines from Bailey et al. [25], there was a small treatment effect for P1 on set 2 (d= 6.08). Due to the large variation in baseline measures for P2, the effect size in set 1 (d=3.13) did not meet the effect size benchmarks despite the large mean treatment gain secondary to the degree of variation in baseline performance.

**Hybrid scale**

Each stimulus in the baseline, retention, and maintenance measure was scored for number of accurate word approximations, presence of distortions, and if the criteria for immediacy was met (Table 4). A calculation for a total score used the number of correct word approximations+distortion score+immediacy score to represent the overall rating for each measure. Speech motor changes for P1 (Figure 3) and P2 (Figure 4) using the hybrid scale are represented as a mean percentage of accuracy for each parameter during baseline, retention,



**Figure 3.** Outcome measures for P1 using the hybrid scale.



**Figure 4.** Outcome measures for P2 using the hybrid scale.

and maintenance measures.

Articulation accuracy measured by number of correct word approximations for P1 at baseline was above 80% for both sets of stimuli with a mean for set 1 at 88% and set 2 at 95%. Mean baseline score for both sets of phrases for distortions and immediacy was 0 indicating distortions and delays were present in all productions. The mean baseline total score was 48.33 out of 65 (74%) for set 1 and 51.33 out of 64 (80%) for set 2. The articulation accuracy for P2 was comparably lower with a mean baseline number of correct word approximations for set 1 at 79% and set 2 at 63%. Distortions were present in all productions in all stimuli for both sets of phrases at baseline. P2 demonstrated immediacy in 7% of the productions in both sets of phrases during baseline. The mean baseline total score was 23.33 out of 39 (60%) for set 1 and 18.67 out of 39 (48%) for set 2.

After training began on set 1, both participants demonstrated improvement in all parameters. By the fifth retention measure, P1 demonstrated 98% accuracy for correct word approximations on set 1 with distortions present in all stimuli for 7 of the 9 retention measures. By the ninth retention measure 93% of the productions were perceived to meet the immedi-

acy criteria. The total score remained >88% after the sixth retention measure. On retention measure 12 and 14, at least one of the productions received a score of 100% indicating the production was free of articulatory errors, no distortions and immediate. By the second treatment session on set 2, articulation accuracy for number of correct word approximations was 100% on all stimuli with distortions prevalent on 80% of the stimuli for 4 of the 7 retention measures. For the majority of the retention measures, 80% of the stimuli met the criteria for immediacy. The total score on set 2 for P1 was  $\geq 90\%$  after the second treatment session (retention 16). On retention measure 17, 19, and 21, at least one of the productions received a score of 100%. In this case, the errorless production was consistent on one of the stimuli.

Correct word approximations for P2 on set 1 reached 90% by retention measure 2; however, a decline in performance occurred on retention 6 and 7 and then returned to the previous level of performance. Distortions were prevalent throughout the intervention; however, at least one stimulus was distortion free for half of the retention measures. Throughout the intervention, at least one of the stimuli did not meet the criteria for immediacy during the retention measures. The total

score was >75% after the ninth retention measure. For P1, productions received a score of 100% (no articulation errors, no distortions, and no delays) on at least one of the stimuli on retention measure 7, 13, 17, 18, and 19. The errorless immediate production was consistent for one of the stimuli in retention measures 17, 18, and 19.

The mean treatment gain was calculated for each parameter in the hybrid rating scale as previously described. Treatment gains for P1 for correct word approximations was more on set 1 at 5.67 compared to 1.67 on set 2. The treatment gain on the distortion score was more on set 2 (1.0) compared to set 1 (0.67). Treatment gain for immediacy was 4.67 for set 1 and 4.0 for set 2. The total treatment gain was greater for set 1 at 11.0 compared to 6.67 for set 2. The treatment gains for P2 for correct word approximations was 5.0, distortion score was 0.80 and immediacy score was 2.80. The total treatment gain was 8.67. These scores indicate the primary contributor to the speech motor changes were the correct word approximations for both participants. However, for P1, immediacy was also a contributor on set 1 and the primary contributor on set 2.

A 10-month retention measure was obtained to identify maintenance of speech motor changes. P1 demonstrated maintenance for correct word approximations for all phrases, distortions remained prevalent, and 60% of the productions were produced without delays. P2 did not demonstrate maintenance of speech motor changes for trained stimuli for any parameters.

The mean standard deviation was calculated as previously described (Table 5). An extended baseline measure was used to calculate the mean baseline when there was no change during the three baseline measures for the distortion and immediacy parameters [24]. The estimated treatment effect size indicates a large effect size for P1 in the immediacy parameter in the intervention phase for set 1 ( $d=9.99$ ) and a medium effect size for set 2 in the immediacy parameter in the intervention phase ( $d=8.50$ ). A small effect size was seen in the maintenance phase for the immediacy parameter in set 1 and set 2 ( $d=6.26$ ). For P2 a large effect size was seen in the total score ( $d=15.01$ ) during the treatment phase.

## DISCUSSION

The results from this pilot study investigation support the use of both a qualitative and quantitative scale to measure speech motor changes during intervention. The speech motor changes identified using the multidimensional rating scale

during MLG treatment for acquired AOS were similar to those from previous MLG investigations [12-14,22,23,26]. The hybrid scale comparably captures the speech motor changes in the quantitative measure of correct word approximations and the calculated total score. The calculated total score shared the same pattern of change as the multidimensional rating scale. The total score and the multidimensional rating represent the individual's overall speech motor changes without the distinction of influential factors such as stereotypic productions from co-existing aphasia.

The qualitative measures of distortions and immediacy did prove informative for specific characteristics of speech motor changes not represented in the multidimensional rating scale. The qualitative measure for the presence of distortions served as the indicator that a production was without articulation error. Because of the structure of the scale, the distortion rating represents the proportion of the stimuli produced in each retention were the intended target and error free including any co-existing aphasia errors. The correct word approximations represent the fluctuations in the articulation accounting for both apraxia and co-existing aphasia errors. According to the multidimensional rating scale, P1's speech appears fairly intelligible; however, when the productions were rated using the hybrid scale it is apparent that there are consistent distortions in his speech affecting his intelligibility at the phrase level. Because P2 had a more severe impairment, the multidimensional rating scale was more representative of the fluctuations; however, the hybrid scale contributed information for both the presence and the number of phrases with distortions. This distinction can guide clinicians to make modifications during intervention. Further, for researchers it contributes information to identify factors that influence change to refine protocols.

Since the immediacy rating is independent of articulation accuracy, we can distinguish the influence on the overall production. Based on our sample, the independent measure for immediacy proved to be a strong influence for speech motor changes in these two cases. According the effect size guidelines [25], immediacy was identified as the main contributor to speech motor changes on both sets of stimuli with a large effect size on set 1 and medium effect size on set 2 for P1. This change was not identified using the multidimensional rating scale and would not have been included in binary measures as reported in other treatment studies. The distinguishing characteristics for AOS refer to the influence of rate on distortions. Rate has shown to be highly variable across speakers [19]; while immediacy as defined in this scale may be a good

representation of speech changes in AOS. The pattern of ratings for distortion and immediacy using the hybrid scale identified the impaired speaker's trade-off between activation of the motor plan and accuracy.

The trade-off measured in this study was immediacy of the production and the presence of distortions. These parameters were chosen based on the diagnostic characteristics that differentiate AOS from aphasia. Both participants demonstrated changes in the immediacy parameter superior to the improvements in articulatory precision as measured by the presence of distortions. The pattern of the trade-off varied greatly between these two individuals during the intervention. One of the participants demonstrated improvements in the immediacy of the productions although, distortions were consistently present throughout the intervention. This is in contrast to the second participant who demonstrated a pattern of fewer distortions with inconsistency in immediacy. There are many potential contributing factors to explain the behavioral differences such as type and severity of co-existing aphasia, familiarity and content of stimuli, cognitive processes for attention, inhibition, working memory, as well as motivation [28-30]. Further investigation on a larger more diverse sample of various severity and time post-onset would contribute to gaining a better understanding the influence these factors have for rehabilitation.

The behavioral differences could also be explained under the motor control and learning theoretical framework [30]. The MLG protocol is based on the theoretical framework of the principles of motor learning [30,31]. The protocol incorporates deliberate delays between productions to allow time for the individual with AOS to identify the accuracy of the attempt and determine any adjustments prior to the next practice attempt. In addition, the clinician provided knowledge of results feedback reinforces the accuracy of the outcome while omitting specific instructions related to the movement. Under the motor control and learning theoretical framework, it is hypothesized that these procedures facilitate the impaired speaker's intrinsic feedback system. Therefore, it would be reasonable to hypothesize that the outcomes in this study may reflect changes to the intrinsic feedback system, which in turn are facilitating speech motor control. The immediacy parameter complemented by the measure for the presence of distortions may serve as an indicator of speech motor control. Studies to specifically test this hypothesis would have practical clinical implications to refine treatment protocols. Further, the outcomes from this small pilot study supports the use of a

quantitative and qualitative outcome measure to capture the transition of speech motor changes during rehabilitation.

While the small sample size limits any specific clinical or theoretical conclusions, the data do support using the hybrid scale as an AOS outcome measure. Additionally, these results support the hypothesis that changes in one speech behavior do not always influence a change in another speech behavior. These preliminary outcomes suggest that immediacy may influence articulation accuracy while articulation accuracy indicates the automaticity of motor programming and planning. Therefore, using an outcome measure that groups speech parameters like the multidimensional rating scale can be as equally misleading as a simple binary measure of articulation accuracy for speech motor changes in AOS.

## REFERENCES

1. Haley KL, Jacks A, de Riesthal M, Abou-Khalil R, Roth HL. Toward a quantitative basis for assessment and diagnosis of apraxia of speech. *Journal of Speech Language and Hearing Research*. 2012;55(5):S1502.
2. Strand EA, Duffy JR, Clark HM, Josephs K. The Apraxia of Speech Rating Scale: a tool for diagnosis and description of apraxia of speech. *J Commun Disord*. 2014;51:43-50.
3. Wambaugh JL, Nessler C, Wright S, Mauszycki SC, DeLong C. Sound production treatment for acquired apraxia of speech: effects of blocked and random practice on multisyllabic word production. *Int J Speech Lang Pathol*. 2016;18(5):450-464.
4. Wambaugh JL, Nessler C, Wright S, Mauszycki SC. Sound production treatment: effects of blocked and random practice. *Am J Speech Lang Pathol*. 2014;23(2):S225-245.
5. Wambaugh JL, Nessler C, Cameron R, Mauszycki SC. Treatment for acquired apraxia of speech: examination of treatment intensity and practice schedule. *Am J Speech Lang Pathol*. 2013;22(1):84-102.
6. Knock TR, Ballard KJ, Robin DA, Schmidt RA. Influence of order of stimulus presentation on speech motor learning: a principled approach to treatment for apraxia of speech. *Aphasiology*. 2000;14(5-6):653-668.
7. Austermann Hula SN, Robin DA, Maas E, Ballard KJ, Schmidt RA. Effects of feedback frequency and timing on acquisition, retention, and transfer of speech skills in acquired apraxia of speech. *J Speech Lang Hear Res*. 2008;51(5):1088-1113.
8. Katz WF, McNeil MR, Garst DM. Treating apraxia of speech (AOS) with EMA-supplied visual augmented feedback. *Aphasiology*. 2010;24(6-8):826-837.
9. Maas E, Barlow J, Robin D, Shapiro L. Treatment of sound errors in aphasia and apraxia of speech: effects of phonological complexity. *Aphasiology*. 2002;16(4-6):609-622.
10. Strand EA, Stoeckel MA, Baas MA. Treatment of severe childhood

- apraxia of speech: a treatment efficacy study. *Journal of Medical Speech-Language Pathology*. 2006;14(4):297-307.
11. Brendel B, Ziegler W. Effectiveness of metrical pacing in the treatment of apraxia of speech. *Aphasiology*. 2008;22(1):77-102.
  12. Friedman IB, Hancock AB, Schulz G, Bamdad MJ. Using principles of motor learning to treat apraxia of speech after traumatic brain injury. *Journal of Medical Speech-Language Pathology*. 2010;18(1):13-34.
  13. Lasker JP, Stierwalt JAG, Hageman CF, LaPointe LL. Using motor learning guided theory and augmentative and alternative communication to improve speech production in profound apraxia: a case example. *Journal of Medical Speech-Language Pathology*. 2008;16(4):225-233.
  14. Lasker JP, Stierwalt JAG, Spence M, Cavin-Root C. Using webCam interactive technology to implement treatment for severe apraxia: a case example. *Journal of Medical Speech-Language Pathology*. 2010;18(4):71-76.
  15. Whiteside SP, Varley RA. Coarticulation in apraxia of speech: an acoustic study of non-words. *Log Phon Vocol*. 1998;23:155-163.
  16. Ballard KJ, Azizi L, Duffy JR, McNeil MR, Halaki M, O'Dwyer N, et al. A predictive model for diagnosing stroke-related apraxia of speech. *Neuropsychologia*. 2016;81:129-139.
  17. Cunningham KT, Haley KL, Jacks A. Speech sound distortions in aphasia and apraxia of speech: reliability and diagnostic significance. *Aphasiology*. 2015;30(4):396-413.
  18. Ballard KJ, Granier JP, Robin DA. Understanding the nature of apraxia of speech: theory, analysis, and treatment. *Aphasiology*. 2000;14(10):969-995.
  19. Youmans G, Youmans SR, Hancock AB. Script training treatment for adults with apraxia of speech. *American Journal of Speech-Language Pathology*. 2011;20(1):23-37.
  20. Dabul BL. *Apraxia battery for adults*. 2nd ed. Austin, TX: ProEd; 2000.
  21. Kertesz A. *Western aphasia battery - revised (WAB-R)*. San Antonio, TX: Pearson; 2006.
  22. Johnson RK, Lasker JP, Stierwalt JAG, MacPherson MK, LaPointe LL. Motor learning guided treatment for acquired apraxia of speech: a case study investigating factors that influence treatment outcomes. *Speech, Language and Hearing*. 2017:1-11.
  23. Johnson RK. Motor learning guided treatment for acquired apraxia of speech. *Speech, Language and Hearing*. 2017:1-11.
  24. Beeson PM, Robey RR. Evaluating single-subject treatment research: lessons learned from the aphasia literature. *Neuropsychol Rev*. 2006;16(4):161-169.
  25. Bailey DJ, Eatchel K, Wambaugh JL. Sound production treatment: synthesis and quantification of outcomes. *Am J Speech Lang Pathol*. 2015;24(4):S798-814.
  26. Kim IS, Seo IH. Treating apraxia of speech (AOS) using the motor learning guided (MLG) approach. *Brain & NeuroRehabilitation*. 2011;4(1):64-68.
  27. Ballard KJ, Tourville JA, Robin DA. Behavioral, computational, and neuroimaging studies of acquired apraxia of speech. *Front Hum Neurosci*. 2014;8:892.
  28. Wulf G, Lewthwaite R. Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychon Bull Rev*. 2016;23(5):1382-1414.
  29. Dajani DR, Uddin LQ. Demystifying cognitive flexibility: implications for clinical and developmental neuroscience. *Trends Neurosci*. 2015;38(9):571-578.
  30. Schmidt RA, Lee T. *Motor control and learning: a behavioral approach*. 5th ed. Champaign, IL: Human Kinetics; 2011.
  31. Bjork RA. Assessing our own competence: Heuristics and illusions. In: Koriat DGA, editor. *Attention and performance XVII: cognitive regulations of performance: Interaction of theory and application*. Cambridge, MA: MIT Press; 1999. p. 435-459.