Supplement Article

Sound Production Treatment: Synthesis and Quantification of Outcomes

Dallin J. Bailey,[a,b] Kelly Eatchel,[a,b] and Julie Wambaugh[a,b]

Purpose: This investigation was designed to provide a quantification and synthesis of a series of single-case experimental design investigations into the effects of sound production treatment, an articulatory-kinematic treatment for acquired apraxia of speech (AOS). The main purpose was to perform a meta-analysis of aggregated sound production treatment data in order to provide benchmarks to serve as indicators of magnitude of change. Additional analyses explored various factors influencing effect sizes and level of performance.
Method: Effect sizes were calculated for treated and untreated items for 24 participants across 10 investigations. Benchmarks were calculated as the quartiles of the distributions of the effect sizes. Correlational analyses were performed to examine (a) end-of-treatment performance relative to follow-up performance, (b) response of trained items relative to untrained items, and (c) effect size relative to participant variables.
Results: Effect sizes were predominantly large and positive; benchmarks for treated items were larger than those for untreated items. End of treatment and follow-up performance were positively correlated, and response generalization was positively correlated with AOS severity.
Conclusion: These benchmarks may assist in evaluating the effects of interventions for individuals with AOS utilizing similar outcome measures in both clinical and research settings.

I n the general area of treatment of apraxia of speech (AOS), the first methodological attempt to synthesize the extant published data took the form of a systematic review with accompanying treatment guidelines (Wambaugh, Duffy, McNeil, Robin, & Rogers, 2006a, 2006b), which was completed as part of the Academy of Neurological Communication Disorders and Sciences' Treatment Guidelines Project. Due to a lack of sufficient research concerning any one treatment or technique, Wambaugh et al. (2006a, 2006b) categorized treatment reports into four general approaches: (a) articulatory-kinematic approaches, (b) rate and/or rhythm control treatments, (c) augmentative and alternative communication techniques, and (d) intersystemic facilitation and reorganization techniques. In their evaluation of the strength of evidence for these general approaches, Wambaugh et al. (2006b) found the strongest evidence for the general category of articulatory-kinematic approaches. A recent systematic review performed as an update to the initial AOS

treatment guidelines (Ballard et al., 2015) bolstered support for articulatory-kinematic treatments. Their review of the AOS treatment literature published during the 9 years following the 2006 report (2004–2012) found that the vast majority of evidence published since the publication of the 2006 guidelines also supported articulatory-kinematic techniques.

To date, the only attempt to quantify the effects of AOS treatments across investigations was reported by Strom (2008). Strom focused on articulatory-kinematic approaches and performed separate meta-analyses for eight group and 11 single-case experimental design (SCED) investigations. Strom constrained the scope of these meta-analyses to include only studies reporting correct sounds or words as the dependent variable. The primary results of these meta-analyses were that articulatory-kinematic treatments are effective for improving sound production.

Sound Production Treatment

One particular articulatory-kinematic treatment that has received relatively systematic study and for which "efficacy data are more adequate . . . than for any other treatment for AOS" (Duffy, 2013, p. 450) is sound production treatment (SPT; Wambaugh, 2004; Wambaugh & Cort, 1998; Wambaugh, Kalinyak-Fliszar, West, & Doyle, 1998; Wambaugh, Martinez, McNeil, & Rogers, 1999; Wambaugh & Mauszycki, 2010; Wambaugh & Nessler, 2004; Wambaugh,

[a]Researchers at the Aphasia and Apraxia of Speech Research Program, VA Salt Lake City Health Care System, UT
[b]University of Utah, Salt Lake City

Correspondence to Dallin Bailey: dallinbailey@gmail.com

Nessler, Cameron, & Mauszycki, 2013; Wambaugh, Nessler, Wright, & Mauszycki, 2014, in press; Wambaugh, West, & Doyle, 1998). SPT uses modeling repetition, minimal pair contrastive practice, orthographic cuing, integral stimulation, articulatory placement cuing, and repeated practice in a response-contingent hierarchy.

As a unified treatment, SPT has received systematic examination since 1998 (Wambaugh & Cort, 1998; Wambaugh, Kalinyak-Fliszar, et al., 1998; Wambaugh, West, et al., 1998). Although certain techniques utilized in SPT had been included in AOS treatment reports prior to 1998 (LaPointe, 1984; Rosenbek, Lemme, Ahern, Harris, & Wertz, 1973; Wertz, LaPointe, & Rosenbek, 1984), SPT's specific treatment components had not yet been combined into a fully operationalized and replicable form. As such, it was not until the initial SPT report (Wambaugh, Kalinyak-Fliszar, et al., 1998) that a protocol specifying the techniques was sufficiently developed to allow replication in the literature.

SPT has been demonstrated to increase sound production accuracy in treated and untreated words, phrases, or sentences for speakers with AOS with a range of AOS severities from mild (Wambaugh et al., in press) to severe (Wambaugh, Kalinyak-Fliszar, et al., 1998; Wambaugh & Mauszycki, 2010). SPT research has recently been directed toward the examination of the effects of practice schedule and treatment intensity, with participants showing improved sound production as a result of both blocked and random practice schedules (Wambaugh et al., 2014) and in both intensive and traditional (nonintensive) treatment applications (Wambaugh et al., 2013).

Previous meta-analysis and systematic evidence reviews could not specifically support any one individual treatment or technique, citing a lack of sufficient replications (Strom, 2008; Wambaugh et al., 2006a). This is unfortunate because various therapies, including SPT, have shown promise as effective treatments for chronic AOS. However, replications of the effects of SPT since the publication of these reviews have increased its evidence base sufficiently to now warrant a meta-analysis focused on SPT as a specific technique (Wambaugh & Mauszycki, 2010; Wambaugh et al., 2013, 2014, in press). Data are available from 24 participants across 10 research reports.

Meta-Analysis of SCEDs

SCED investigations constitute a substantial portion of AOS treatment evidence and the entirety of SPT evidence (Ballard et al., 2015; Wambaugh et al., 2006a). A predominance of SCED studies is to be expected in a relatively young field of inquiry such as treatment of AOS. These designs can provide persuasive evidence of a treatment's effects for participants included in a given SCED report. However, synthesis of findings across sets of SCED investigations may be necessary to develop generalized conclusions concerning a particular treatment or therapeutic approach that can inform practice and future research (Maggin & Chafouleas, 2013).

Although a long-standing concern directed toward SCEDs relates to a perceived limitation in ability to generalize findings beyond the participants in a given SCED investigation, meta-analysis of findings from related SCED investigations (e.g., SPT) provides a means for enhancing the external validity of treatments whose efficacy has been demonstrated with SCEDs. The development of meta-analytic methods for application with single-subject investigations represents a meshing of nomothetic and idiographic research paradigms that offers potential for aggregating findings from related single-subject treatment investigations (Beeson & Robey, 2006). There are numerous challenges and debates concerning the synthesis of single-subject research, which include determining methodological quality, quantifying treatment effects, and synthesizing findings to draw implications (Evans, Gast, Perdices, & Manolov, 2014; Maggin & Chafouleas, 2013). Although significant attention has recently been devoted to these challenges, there remains a lack of consensus concerning optimal methods to be used. However, there is increasing recognition of the value of SCED meta-analysis and a strong likelihood that continuing advancements will be made in single-case meta-analytic methods (Campbell, 2013).

One outcome of a meta-analysis of single-subject experimental data includes the establishment of effect size benchmarks (Beeson & Robey, 2006). Benchmarks further scientific evidence by providing a reference point against which effects from other treatments and other participants or patients may be compared. A meta-analysis translates the different methods and dependent variables of the primary studies into standard deviation units (effect sizes), commonly on the basis of the participant's baseline variability. The average effect size reflects the mean of the distribution, and the benchmarks are derived from the quartiles of the distribution (Beeson & Robey, 2006). The establishment of SPT benchmarks provides clinicians and researchers with a comparison for gauging the effects of their treatment—whether it is SPT or another AOS treatment—on the sound production accuracy of the speaker with AOS.

Purposes

The present meta-analysis was designed to establish SPT effect size benchmarks as a resource for clinicians and researchers performing treatment for AOS. In addition, factors possibly related to the maintenance and generalization effects of SPT were examined. For most SPT participants, improvements in sound production accuracy achieved during treatment were maintained at posttreatment intervals. However, maintenance effects have varied within and across participants.

An additional purpose of this meta-analysis was to examine the relationship of performance during treatment to performance at posttreatment follow-up. Strom (2008) suggested that in several cases, treatment may have been prematurely terminated in some articulatory-kinematic investigations and that additional gains may have been possible with continued treatment. This may have been the case with some SPT investigations.

Last, given the likelihood that participant characteristics affect outcomes, possible associations of case variables (e.g., AOS severity) with effect sizes were also explored. Although generalization to untrained exemplars of trained items has typically been a robust effect of SPT, there has been some degree of variability within and across participants.

Method

The present study adhered to guidance presented in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009) statement for improving the quality and standardization of systematic reviews and meta-analyses.

Articles Included in the Meta-Analysis

The reports used in the meta-analysis included those identified by the results of the extensive database reviews performed during development of the AOS treatment guidelines and guideline update (Ballard et al., 2015; Wambaugh et al., 2006a). The review authors reported that they had performed extensive searches of a total of 13 major databases: Wambaugh et al. (2006a) reviewed MEDLINE, PsycINFO, and CINAHL; Ballard et al. (2015) also reviewed CINAHL and MEDLINE as well as Cochrane Database of Systematic Reviews, EMBASE, Expanded Academic ASAP, Google Scholar, PubMed, Proquest, Proquest Dissertation and Thesis, ScienceDirect, Scopus, and Web of Science. These reviews revealed a total of seven published articles evaluating SPT treatment. Beyond these results, two additional reports on SPT were known to the authors, including one presented at a conference poster session in 1998 (Wambaugh & Cort, 1998) and one published following the window of articles included in the guidelines update (Wambaugh et al., 2014). These were included in order to include all available reports examining SPT. To be included in the present meta-analysis, reports had to provide retrievable data concerning the efficacy of SPT. Nine published reports were found, and data were available for an additional investigation completed in our Veterans Affairs laboratory and recently accepted for publication (Wambaugh et al., in press). All investigations used multiple-baseline SCED designs.

Before effect sizes were calculated and included in the meta-analysis, the nine published reports and the protocol for the unpublished report were examined for quality and presence of experimental control according to the SCED standards explicated in a report by the What Works Clearinghouse (Kratochwill et al., 2010). The first and second authors independently reviewed the reports according to these standards. These standards rigorously address various aspects of the quality of the experimental design. The design standards (Kratochwill et al., 2010) provide criteria for four separate design characteristics to meet design standards; otherwise, the studies either meet design standards with reservations or do not meet design standards. The criteria

include (a) the systematic manipulation of the independent variable by the investigator, (b) interassessor agreement, (c) number of attempts to demonstrate an intervention effect, and (d) the number of data points within a phase (at least three). Each study was examined for all four criteria. Agreement between the two authors on the ratings was calculated as the percentage of these criteria for which the two raters assigned the same rating for the same study. Agreement between the raters was complete, or 100%.

All of the SPT studies met requirements for meeting design standards for all of the published standards except for the standard proposing minimal thresholds for interassessor agreement. Therefore, in each study, the investigator controlled the manipulation of the independent variable, the number of phases was sufficient to replicate the intervention effect, and at least three data points were used in each phase to determine the number of intervention effect replications. However, the suggested standard for interassessor agreement was not met in each case, with four studies reporting interrater reliability for only 10% of probe data (as opposed to the 20% required to meet evidence standards; Wambaugh & Mauszycki, 2010; Wambaugh et al., 1999, 2013, 2014). In addition, one study did not report any interassessor agreement (Wambaugh & Nessler, 2004). For the purposes of this study, the data from these studies were still included in the meta-analysis. This was justified by the facts that they met a majority of the design standards, an adequate level of experimental control was maintained, and the investigators reported satisfactory levels of agreement in closely related investigations.

Source Data for Effect Sizes

All individual data points (percentage correct target productions in target stimuli repetition probes) were extracted from the 10 separate reports involving a total of 24 participants. Original data points for probe values were available in electronic form for several of the studies. For studies for which original electronic data were not available, probe data were extracted from the published graphs. Values were hand measured by one author from a printed copy of the graphs and confirmed by another author using the online graph digitizer WebPlotDigitizer (Rohatgi, 2014). All discrepancies found were resolved by consensus. All of the studies reported percentage accuracy data for probes of sound production accuracy. In two studies, the entire word was scored as correct or incorrect; all sounds within the word were required to be produced correctly (Wambaugh et al., 2013, in press). In the remaining eight studies, only the target sound within the word was scored.

Source Data for Correlations

The data for the correlational analyses were level of performance data (i.e., percentage accuracy achieved on probes). These comparisons were planned in order to examine how level of acquisition of treated items might predict level of performance with untreated items (i.e., response

generalization) and how it might predict level of performance of both in the follow-up phase. The *acquisition level of performance* was defined as the mean of the last three probe values of a given treatment phase. The *follow-up level of performance* was defined as the first follow-up probe value due to the differences in number of follow-up points and follow-up time intervals between studies. These initial follow-up points represent a range of 1 to 10 weeks ($M = 3.4$ weeks, $SD = 2.64$ weeks).

Participants

Pertinent demographic and speech and language profile information was gathered for all of the participants. Participants were all adults with AOS and Broca's aphasia resulting from a cerebrovascular accident. Other information coded included sex, age, months postonset of stroke (MPO), etiology, Western Aphasia Battery (WAB: Kertesz, 1982; WAB–Revised: Kertesz, 2007) Aphasia Quotient (AQ), and a scaled clinical rating of AOS severity (1 = *mild*, 5 = *severe*). The descriptors for each participant are shown in Table 1.

Three of the 24 participants were found to have completed SPT treatment two times each—that is, the same participants were reported in separate publications, although with unique data on both occasions. Several years separated the different studies for these participants; thus, some of

their descriptors (e.g., age, MPO) had changed. These participants were included twice in the analysis because they had been reported twice and because two unique sets of published data were available for each of them. The participant identifiers used for the current investigation are shown in Table 1. The three participants who received SPT in two investigations were Participants 2/7, 4/8, and 6/9. Their treatment targets differed across studies, with partial overlap in targets for Participant 6/9. However, as reported by Wambaugh and Nessler (2004), this participant had not maintained treatment gains for the overlapping targets through the intervening years. A second exposure could have effects on response to treatment, though at present how this might affect effect sizes is unknown.

Variation in the Independent and Dependent Variables Across Studies

The reports included in the meta-analysis all utilized SPT as the only intervention, with slight variations across reports stemming from the process of treatment development and testing. In all investigations, the following steps of the response-contingent SPT hierarchy were used: (a) verbal model and request for repetition, (b) orthographic cueing of target sound with additional verbal model and request for repetition, (c) integral stimulation, and (d) articulatory placement cueing (see Wambaugh & Mauszycki,

Table 1. Participant characteristics from all sound production treatment studies.

Primary study	Original participant no.	Meta-analysis participant no.	Age (years)	Sex	MPO	Etiology	AOS severity	WAB AQ	Aphasia type
Wambaugh, Kalinyak-Fliszar, et al. (1998)	P1	1	53	M	20	CVA	Moderate	30.0	Broca's
	P2	2	52	M	33	CVA	Severe	29.3	Broca's
	P3	3	63	M	67	CVA	Moderate	31.2	Broca's
Wambaugh, West, et al. (1998)	P1	4	49	F	61	CVA	Mild to moderate	75.5	Broca's
Wambaugh & Cort (1998)	P1	5	58	M	26	CVA	Moderate to severe	34.0	Broca's
Wambaugh et al. (1999)	P1	6	62	M	8	CVA	Moderate	51.0	Broca's
Wambaugh (2004)	P1	7	54	M	70	CVA	Moderate to severe	37.1	Broca's
	P2	8	48	F	50	CVA	Mild to moderate	63.4	Broca's
Wambaugh & Nessler (2004)	P1	9	66	M	48	CVA	Moderate to severe	70.0	Broca's
Wambaugh & Mauszycki (2010)	P1	10	55	M	24	CVA	Severe	14.8	Broca's
Wambaugh et al. (2013)	P1	11	51	M	231	CVA	Mild to moderate	65.1	Broca's
	P2	12	50	F	61	CVA	Moderate	51.8	Broca's
	P3	13	34	M	26	CVA	Moderate to severe	26.7	Broca's
	P4	14	53	M	232	CVA	Moderate	61.0	Broca's
Wambaugh et al. (2014)	P1	15	55	M	64	CVA	Moderate	68.9	Broca's
	P2	16	46	M	87	CVA	Moderate to severe	53.4	Broca's
	P3	17	55	M	86	CVA	Moderate to severe	34.5	Broca's
	P4	18	71	M	58	CVA	Severe	28.7	Broca's
	P5	19	51	F	83	CVA	Moderate	56.4	Broca's
	P6	20	58	M	28	CVA	Moderate to severe	48.2	Broca's
Wambaugh et al. (in press)	P1	21	83	F	48	CVA	Mild to moderate	60.9	Broca's
	P2	22	46	F	17	CVA	Mild	52.9	Broca's
	P3	23	37	M	34	CVA	Mild to moderate	65.4	Broca's
	P4	24	55	M	259	CVA	Mild to moderate	59.9	Broca's

Note. MPO = months postonset of stroke; AOS = apraxia of speech; WAB AQ = Western Aphasia Battery Aphasia Quotient; M = male; CVA = cerebrovascular accident; F = female.

2010, for the most current SPT protocol). The protocol variations primarily were alterations in response to findings from previous reports as well as modifications to tailor the approach to specific participants.

Variations in Hierarchy Ingredients

The protocol in all the studies began with a verbal model of the target sound in a word along with a request for repetition. In addition, Wambaugh, West, et al. (1998) provided a written model of the target (a sentence) at the outset. Protocols varied in their use of minimal pairs: (a) Modifications were made to reduce patient frustration when minimal pair production was difficult (Wambaugh, Kalinyak-Fliszar, et al., 1998; Wambaugh et al., 1999), (b) minimal pairs were used only if the patient response was in error (as opposed to requiring production of both as a mandatory step, which was done in the earlier studies), and (c) the minimal pair step was eliminated for patients with milder cases of AOS, whose multisyllabic target words or sentences precluded real-word minimal pair selection. In all protocols, the written letter was used as a visual cue to focus the participant's attention on the target sound. Integral stimulation and articulatory placement cuing were also used as response-contingent steps in each study. In the first study only (Wambaugh, Kalinyak-Fliszar, et al., 1998), a silent juncture step, consisting of a model that separated the target sound from the rest of the word with a pause, was included in the hierarchy. This step was omitted in all subsequent studies due to a lack of facilitatory effect. Two of the studies included an attempt to elicit production of the target sound in isolation as a final hierarchy step (Wambaugh, 2004; Wambaugh et al., 1999).

Variations in Administration

In the early SPT investigations, one treatment target was submitted to the SPT hierarchy at a time—that is, dependent variables were treated sequentially in order to determine whether treatment effects generalized to unrelated targets. In subsequent investigations, multiple targets were treated simultaneously. Schedule of practice with multiple targets (i.e., blocked vs. random) was also manipulated in recent SPT investigations (Wambaugh et al., 2013, 2014, in press).

The number of productions per session necessarily varied across studies and participants as well as between sessions within participants. As noted by Wambaugh et al. (2013), the number of productions per SPT session typically increases as a participant achieves greater facility in production. In initial sessions, participants require more steps of the hierarchy and more articulatory placement instruction. This results in the completion of fewer trials than in later sessions, where less frequent cueing is needed. In addition, the later SPT studies had protocols that requested more repetitions (five additional productions after a correct production) than the earlier studies (zero or one additional production after a correct response).

The number of treatment sessions completed with each participant also varied across studies and participants.

In some investigations, a predetermined number of treatment sessions was completed per dependent variable. In other investigations, a behavioral criterion (e.g., 80% correct in two consecutive probe sessions) was used to determine number of treatment sessions. In two investigations, previously trained targets received a booster phase of training to counter declining maintenance performance (due to intervening treatment phases with other treatment targets; Wambaugh & Mauszycki, 2010; Wambaugh & Nessler, 2004). The same words were targeted, which may have influenced the follow-up probes.

Variations in Dependent Variables and Scoring

In the majority of the reports, accuracy of production of target sounds in the context of words, phrases, or sentences served as the dependent variable. As noted previously, in two investigations, accuracy of production of all sounds within words was scored.

In all investigations, determination of accuracy was based on transcription of target items produced during probes in which the examiner provided a model and requested a repetition. Broad phonetic transcription was used in the majority of the studies, with narrow transcription (Haley, Bays, & Ohde, 2001) used in the last three investigations. This was primarily a difference in convention rather than methodology because distortions were always scored as incorrect productions of the target in all studies.

The specific items that served as dependent variables were individually selected for each participant in all of the investigations (see Table 2 for targets by participant)—that is, the treatment targets for SPT were tailored to meet individual participants' sound production errors. As a consequence, treatment targets (and corresponding dependent variables) included specific consonants and consonant clusters in words, words in phrases, and sentences.

Acquisition effects were measured in all SPT reports; production of items that were submitted to treatment was measured in probes. In all but one investigation, production of untrained exemplars of trained behaviors also served as a dependent variable (i.e., response generalization). Stimulus generalization effects were also measured in a few SPT investigations (Wambaugh, 2004; Wambaugh, Kalinyak-Fliszar, et al., 1998; Wambaugh & Nessler, 2004). Due to the lack of stimulus generalization data for most SPT investigations, the stimulus generalization data were not included in the current meta-analysis.

Despite the differences across reports, the core approach of SPT as a cohesive treatment method was maintained throughout all studies. Although variation was shown to exist among the treatment methodologies and the treatment targets of the SPT primary studies, the articles were deemed similar enough to be combined in a meta-analysis. The differences were relatively minor protocol modifications, and the probes, from which the data for the dependent variable were collected, were quite similar across studies. With future replications of SPT, fine-grained comparisons of these variations may become possible.

Table 2. Effect sizes (ES) for individual treatment targets.

Primary study	Meta-analysis participant no.	Treatment target	Treated items		Untreated items	
			Treatment phase ES	Follow-up phase ES	Treatment phase ES	Follow-up phase ES
Wambaugh, Kalinyak-Fliszar, et al. (1998)	1	/z/-initial one- and two-syllable words	5.82[a]	8.57[a]	13.57	11.26
	1	/ʃ/-final one- and two-syllable words	4.90[a]	1.22[a]	11.94	1.50
	1	/dʒ/-initial one- and two-syllable words	7.24	9.80	4.44	6.01
	2	/ʃ/-final one- and two-syllable words	5.61	6.49	4.67	5.66
	2	/r/-initial one- and two-syllable words	9.67	7.00	9.10	19.68
	2	/sw/-initial one- and two-syllable words	5.71	4.75	7.54	8.61
	3	/tʃ/-initial one- and two-syllable words	5.96	3.91	3.47	1.21
	3	/ʃ/-final one- and two-syllable words	2.17	0.94	4.46	0.16
	3	/z/-initial one- and two-syllable words	3.37	26.08	4.01	20.79
Wambaugh, West, et al. (1998)	4	Stops, all positions, three- to five-word sentences	6.94	6.83	9.01	8.21
	4	Fricatives, all positions, three- to five-word sentences	6.81	6.60	5.86	8.60
	4	Glides/liquids, all positions, three- to five-word sentences	4.44	2.40	3.34	2.47
Wambaugh & Cort (1998)	5	/b/-initial one-syllable words	12.50	—	11.26	—
	5	/dʒ/-initial one-syllable words	6.33	—	5.49	—
	5	/d/-initial one-syllable words	2.12	—	1.92	—
	5	/g/-initial one-syllable words	2.19	—	2.42	—
Wambaugh et al. (1999)	6	/p/-initial one-syllable words	15.17	16.50	16.50	16.50
	6	/k/-initial one-syllable words	24.06	18.00[a]	17.13	17.50
	6	/ʃ/-initial one-syllable words	14.00	8.00[a]	6.74[a]	4.00[a]
Wambaugh (2004)	7	/v/-initial one- and two-syllable words	15.17	15.50	8.98	9.80
	7	/r/-blends–initial one- and two-syllable words	2.92	7.35	1.89	5.20
	8	/ʃ/-syllable–final one-syllable words in two-word phrases	5.81	5.81	5.48	7.57
	8	/z/-syllable–initial one- and two-syllable words in two-word phrases	4.82	3.35	2.77	–0.77
Wambaugh & Nessler (2004)	9	/s/-initial one-syllable CV(C) words	19.16	16.21	—	—
	9	/p/-initial one-syllable CV(C) words	8.13[a]	5.93[a]	—	—
	9	/v/-initial one-syllable CV(C) words	3.06[a]	0.95[a]	—	—
	9	/k/-initial one-syllable CV(C) words	6.48[a]	5.93[a]	—	—
	9	/ʃ/-initial one-syllable CV(C) words	3.42[a]	7.28[a]	—	—
	9	/dʒ/-initial one-syllable CV(C) words	–0.17	5.93[a]	—	—
	9	/l/-initial one-syllable CV(C) words	8.12	4.03	—	—
	9	/m/-initial one-syllable CV(C) words	1.65	2.78	—	—
	9	/n/-initial one-syllable CV(C) words	6.21	5.41	—	—
Wambaugh & Mauszycki (2010)	10	/b/-initial one-syllable CV(C) words	8.96	11.34	3.82[a]	5.48[a]
	10	/s/-initial one-syllable CV(C) words	2.18[a]	5.79[a]	3.39[a]	2.74[a]
	10	/l/-initial one-syllable CV(C) words	7.75	7.38	3.82[a]	7.30[a]
	10	/m/-initial one-syllable CV(C) words	6.37	2.51	3.31	–0.73
	10	/d/-initial one-syllable CV(C) words	4.36[a]	4.14[a]	2.12[a]	1.83[a]
	10	/f/-initial one-syllable CV(C) words	10.37[a]	11.57[a]	5.94[a]	7.30[a]

(*table continues*)

Table 2 *(Continued).*

Primary study	Meta-analysis participant no.	Treatment target	Treated items		Untreated items	
			Treatment phase ES	Follow-up phase ES	Treatment phase ES	Follow-up phase ES
Wambaugh et al. (2013)	11	/s/-initial four-syllable words (IB)	13.06	8.19	3.39	2.99
	11	/ɪ/- and /ɛ/-initial three-syllable words (IR)	4.15	3.13	0.90	3.90
	11	/r/-initial four-syllable words (TR)	12.66	5.78	4.27	4.65[a]
	11	/ɑ/-, /æ/-, /i/-, or /o/-initial three-syllable words (TB)	6.69	5.86	0.86	3.96
	12	/θ/-initial or final two-syllable words (IR)	8.27	4.87	2.88	4.24
	12	/ʃ/-initial or final two-syllable words (IB)	8.81	6.86	3.83	3.32
	12	/z/-initial or final two-syllable words (TB)	8.45	3.82	1.60	−0.32
	12	/dʒ/-initial or final two-syllable words (TR)	4.00	7.10	1.27	1.29
	13	/sw/-initial one-syllable or /θ/-final two-syllable words (TR)	14.68[a]	15.65[a]	0.50[a]	2.98[a]
	13	/sm/-initial one-syllable or /ʃ/-final two-syllable words (TB)	25.50[a]	11.18[a]	4.00[a]	4.47[a]
	13	/bl/-initial one-syllable or /z/-final two-syllable words (IB)	23.18	8.65	2.67	7.45[a]
	13	/fl/-initial one-syllable or /dʒ/-final two-syllable words (IR)	22.45	7.45[a]	7.00[a]	2.98[a]
	14	/θ/- or /br/-initial three-syllable words (TB)	10.71	7.75	1.82	2.24
	14	/z/- or /gl/-initial three-syllable words (TR)	5.07	9.20	2.20	0.38
	14	/st/- or /t/-initial three-syllable words (IR)	7.55	3.02	4.24	3.56
	14	/ʃ/- or /sk/-initial three-syllable words (IB)	5.89	8.35	3.61	4.20
Wambaugh et al. (2014)	15	/pr/-initial three-syllable words	4.14	3.75	3.28	0.30
	15	/ks/-medial three-syllable words	8.03	6.82	2.53	4.02
	15	/gl/-initial three-syllable words	3.05	3.99	−0.42	0.30
	15	/nd/-medial three-syllable words	3.83	6.61	0.40	1.49[a]
	16	/z/-final one-syllable words	21.67	16.38	5.80	5.80
	16	/st/-initial one-syllable words	9.87	5.20	4.78	1.82
	16	/f/-final one-syllable words	7.18	10.47	6.45	8.30
	16	/gl/-initial one-syllable words	9.68[a]	8.20[a]	5.64[a]	5.45[a]
	17	/dʒ/-initial one-syllable words	9.76	9.02	4.02	4.02
	17	/sm/-initial one-syllable words	15.58[a]	12.94[a]	5.68[a]	3.88[a]
	17	/θ/-initial one-syllable words	15.06[a]	6.47[a]	4.54[a]	1.49[a]
	17	/gl/-initial one-syllable words	16.70	10.22	3.18	2.29
	18	/f/-final one-syllable words	11.43	4.49	9.57[a]	2.43[a]
	18	/sm/-initial one-syllable words	13.70	11.07	8.22	8.22
	18	/bl/-initial one-syllable words	10.95[a]	12.02[a]	8.93[a]	4.26[a]
	18	/p/-final one-syllable words	20.01	16.74	8.00	7.00
	19	/gl/-initial two-syllable words	8.72[a]	12.19[a]	8.27[a]	8.57[a]
	19	/kw/-medial two-syllable words	12.17	10.89	3.51	3.51
	19	/θr/-initial two-syllable words	9.77[a]	11.29[a]	13.30	11.97
	19	/st/-medial two-syllable words	6.77	11.32	7.75	4.19
	20	/θr/-initial two-syllable words	19.62	15.30	9.54	7.38
	20	/gl/-initial two-syllable words	8.11	2.18	6.66	3.78
	20	/sn/-initial three-syllable words	6.60[a]	7.91[a]	15.75	4.69[a]
	20	/kw/-initial three-syllable words	4.10	6.66	3.79	10.02

(table continues)

Table 2 *(Continued).*

Primary study	Meta-analysis participant no.	Treatment target	Treated items		Untreated items	
			Treatment phase ES	Follow-up phase ES	Treatment phase ES	Follow-up phase ES
Wambaugh et al. (in press)	21	/θ/-medial three- and four-syllable words (B)	6.57	7.18	1.22	1.98
	21	/s/-blends initial three- and four-syllable words (B)	4.02	4.02	0.00[a]	0.00[a]
	21	/br/-initial or medial three- and four-syllable words (R)	5.58	4.75	1.56	2.56
	21	/l/-medial three- and four-syllable words (R)	3.79	8.50	0.41	1.79
	22	/st/-initial four- and five-syllable words (R)	7.43	10.08	3.35	2.09
	22	/tr/-medial four- and five-syllable words (R)	9.66	8.98	−0.50	−0.07
	22	/θ/-medial four- and five-syllable words (B)	2.47	6.68	2.47	6.68
	22	/fl/-initial and medial four- and five-syllable words (B)	4.86[a]	5.71[a]	1.99[a]	3.12[a]
	23	/st/-initial three- and four-syllable words (B)	7.57[a]	9.05[a]	1.60[a]	1.47[a]
	23	/dʒ/-medial three- and four-syllable words (B)	8.82	7.81	2.36	2.36
	23	/gl/-initial three- and four-syllable words (R)	10.00	8.57	0.16	0.67
	23	/θ/-medial three- and four-syllable words (R)	10.25	14.67	0.77	5.73
	24	/kw/-medial four-syllable words (R)	14.14	2.71	1.53	1.53
	24	/br/-medial four-syllable words (R)	9.26	3.78	0.00[a]	1.89[a]
	24	/fl/-medial four-syllable words (B)	12.22	4.19	1.71[a]	0.00[a]
	24	/sp/-medial four-syllable words (B)	5.35	7.19	1.06	3.80[a]

Note. Em dashes indicate necessary phase data not collected in the original study. CV(C) = consonant–vowel or consonant–vowel–consonant syllable structure; IB = intense blocked; TR = traditional random; IR = intense random; TB = traditional blocked; B = blocked; R = random.

[a]No variance in the original data; used variance from other baselines from the same participant.

Effect Size Calculations for Meta-Analysis

Effect sizes were calculated as standardized mean differences with the delta index formula, which is the difference in means of the two phases (A_1 and A_2) divided by the standard deviation of the baseline phase (A_1; Bloom, Fischer, & Orme, 2003; Busk & Serlin, 1992):

$$\text{Effect size} = \frac{M_{A_2} - M_{A_1}}{SD_{A_1}}.$$

However, this formula is undefined when the baseline phase has zero variance. In these situations, the average of the standard deviations of the available nonzero A_1 phases for the same individual was reported in order to provide a conservative estimate, as suggested by Beeson and Robey (2006). If the zero variance point was for a list of treated items, the average standard deviation of that participant's other treated items lists in the same phase was substituted as a conservative estimate of the effect size; a similar procedure was done for untreated lists with zero variance. In the case of Participant 13's initial baselines, there was no variance in any of the baseline measurements for untreated items and so the variance from the treated items lists was substituted. An effect size was calculated for each target for each participant, yielding multiple effect sizes per participant. Calculation of effect sizes was done in an Excel

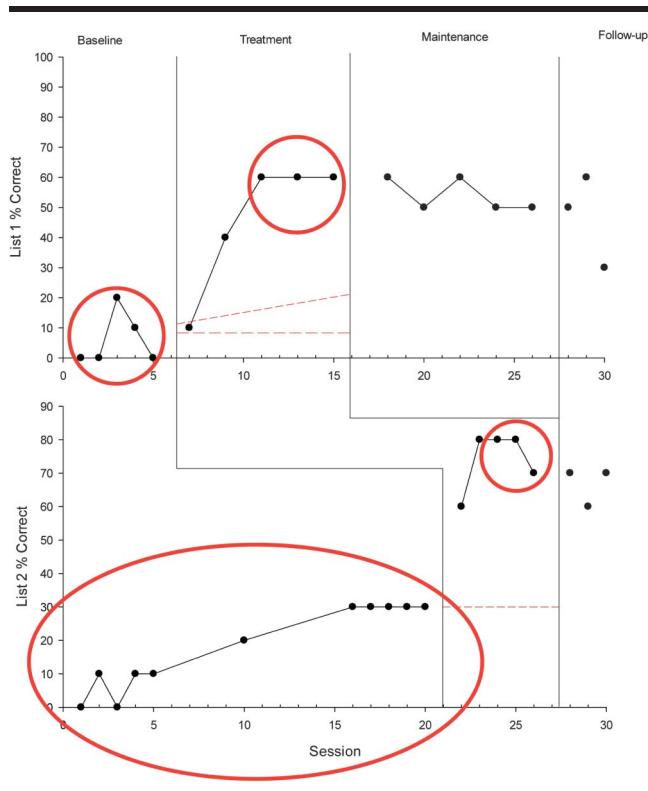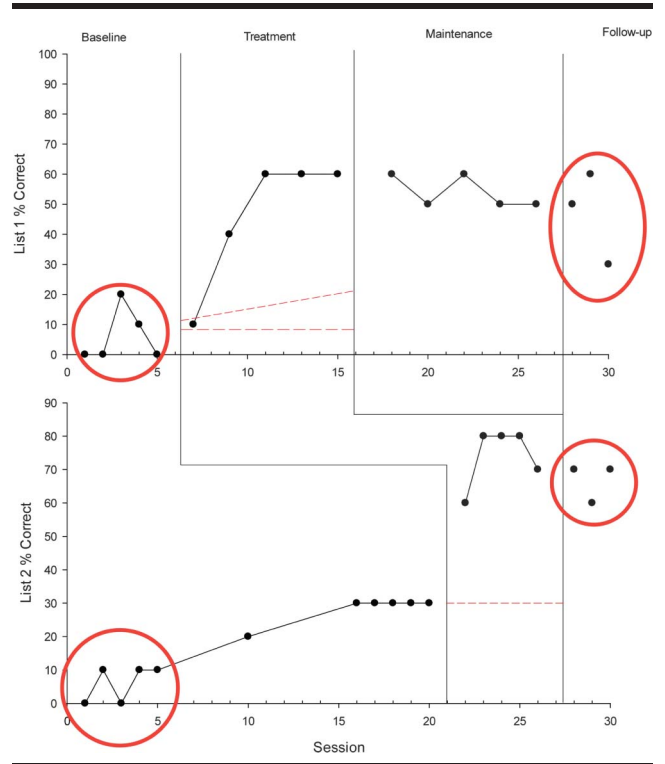Figure 1. Data for treatment effect size.

Figure 2. Data for follow-up effect size.

spreadsheet, with the second author independently verifying the accuracy of the data entry and formulae.

Where possible, the treatment and follow-up effect sizes were calculated for two lists of probe stimuli: the lists of words practiced in the treatment sessions (the treated items) and the list of words for response generalization, which were similar words with the same target sound but were never practiced in treatment (the untreated items). Thus, in most cases, four separate effect sizes per target per participant were calculated:

1.   treatment phase effect sizes for treated items,
2.   treatment phase effect sizes for untreated (but similar) items,
3.   follow-up phase effect sizes for treated items, and
4.   follow-up phase effect sizes for untreated (but similar) items.

The treatment phase effect sizes were calculated using all probe data points from baseline and the last three probe data points from the end of the treatment phase for a given target, as shown in Figure 1. The follow-up effect size was calculated using only the initial baseline phase probe data points (probes before the initiation of treatment of any of the targets for that participant) and all available follow-up points, as shown in Figure 2. The choice to use only the initial baseline probe data was made so that the overall effects of SPT (including any potential across-targets generalization) would be reflected in the follow-up

effect sizes. One study did not report follow-up data (Wambaugh & Cort, 1998), and one study did not measure performance on response generalization lists (Wambaugh & Nessler, 2004). The number of follow-up points ranged from one to three.

The raw effect size data are shown in Table 2. These are unweighted and therefore reflect participant and target effects. The weighted values, as discussed below, are shown in Table 3.

Meta-Analytical Method

In multiple-baseline designs, an effect size is calculated for the treatment's effect on each separate target behavior, weighted by the number of observations in the phases, and then averaged with the other effect sizes for a given participant (Beeson & Robey, 2006). This yields one effect size per participant, reducing error from target effects in the final meta-analysis. The recommendations for weighting described by Beeson and Robey (2006) were followed for the current analysis, yielding one weighted and averaged effect size per participant. This was performed in the same Excel spreadsheet, with the second author again verifying.

As called for by Beeson and Robey (2006), the first, second, and third quartiles of the distribution of effect sizes in each category were computed to determine benchmarks for small, medium, and large effect sizes for SPT. The second quartile is the median of all the data in a category, and the first and third quartiles are the medians of the lower and upper halves of the data, respectively.

Correlation Calculations

The weighted effect sizes were correlated with several of the participant descriptive variables, including age, MPO, WAB AQ, and AOS severity, to explore associations of effect size with participant variables. AOS severity was represented as ordinal scale values standing in for the qualitative severity judgments found in the primary studies, which were based on performance on the Apraxia Battery for Adults–Second Edition (Dabul, 2000) and clinical judgment. These correlations were calculated as Spearman's rho.

Data were also compiled for correlating performance levels between phases in order to calculate (a) the degree to which level of probe performance on treated items predicts level of probe performance on untreated items at the end of treatment and (b) how level of probe performance at the end of the treatment phase predicts level of probe performance at follow-up.

Likewise, performance level data were used to examine the relationship between performance with treated items (acquisition effects) and performance with untreated items (generalization effects). The data used for the correlations are shown in Table 4. These performance levels were also correlated using Spearman's rho correlations.

Table 3. Effect sizes (ES) weighted by participant.

| Primary study | Meta-analysis participant no. | Treated items | | Untreated items | |
		Treatment phase ES	Follow-up phase ES	Treatment phase ES	Follow-up phase ES
Wambaugh, Kalinyak-Fliszar, et al. (1998)	1	6.30	6.53	8.00	6.25
Wambaugh, Kalinyak-Fliszar, et al. (1998)	2	7.01	6.08	7.51	11.32
Wambaugh, Kalinyak-Fliszar, et al. (1998)	3	3.24	10.31	4.11	7.39
Wambaugh, West, et al. (1998)	4	5.79	5.27	5.44	6.43
Wambaugh & Cort (1998)	5	3.90	—	3.62	—
Wambaugh et al. (1999)	6	17.44	14.17	10.96	12.67
Wambaugh (2004)	7	5.30	11.42	3.27	7.50
Wambaugh (2004)	8	5.17	4.58	3.72	3.40
Wambaugh & Nessler (2004)	9	5.07	6.05	—	—
Wambaugh & Mauszycki (2010)	10	6.76	7.12	3.75	3.99
Wambaugh et al. (2013)	11	8.77	5.74	2.24	3.87
Wambaugh et al. (2013)	12	7.19	5.66	2.25	2.13
Wambaugh et al. (2013)	13	22.00	10.73	3.89	4.47
Wambaugh et al. (2013)	14	6.95	7.08	3.03	2.59
Wambaugh et al. (2014)	15	4.36	5.29	1.01	1.53
Wambaugh et al. (2014)	16	10.72	10.06	5.81	5.34
Wambaugh et al. (2014)	17	14.76	9.66	4.21	2.92
Wambaugh et al. (2014)	18	14.47	11.08	8.61	5.48
Wambaugh et al. (2014)	19	9.02	11.42	8.93	7.06
Wambaugh et al. (2014)	20	8.19	8.01	9.21	6.47
Wambaugh et al. (in press)	21	4.91	6.11	0.85	1.58
Wambaugh et al. (in press)	22	5.54	7.86	1.92	2.95
Wambaugh et al. (in press)	23	9.38	10.02	1.05	2.56
Wambaugh et al. (in press)	24	9.93	4.47	1.14	1.80

Note. All values are given in standard deviation units. Em dashes indicate necessary phase data not collected in the original study.

Table 4. Levels of performance in baseline and at the end of treatment phases.

Primary study	Meta-analysis participant no.	Treatment target	Treated items		Untreated items	
			Mean of last three treatments	First follow-up value	Mean of last three treatments	First follow-up value
Wambaugh, Kalinyak-Fliszar, et al. (1998)	1	/z/-initial one- and two-syllable words	63.33	70	83.33	70
	1	/ʃ/-final one- and two-syllable words	53.33	10	50.00	10
	1	/dʒ/-initial one- and two-syllable words	93.33	90	73.33	70
	2	/ʃ/-final one- and two-syllable words	90.00	100	86.67	100
	2	/r/-initial one- and two-syllable words	83.33	80	80.00	90
	2	/sw/-initial one- and two-syllable words	83.33	70	86.67	90
	3	/tʃ/-initial one- and two-syllable words	83.33	60	80.00	40
	3	/ʃ/-final one- and two-syllable words	50.00	20	50.00	10
	3	/z/-initial one- and two-syllable words	76.67	100	93.33	80
Wambaugh, West, et al. (1998)	4	Stops, all positions, three- to five-word sentences	94.67	94	82.00	80
	4	Fricatives, all positions, three- to five-word sentences	90.33	94	82.00	87
	4	Glides/liquids, all positions, three- to five-word sentences	96.33	92	87.00	89
Wambaugh & Cort (1998)	5	/b/-initial one-syllable words	90.00	—	80.00	—
	5	/dʒ/-initial one-syllable words	80.00	—	93.33	—
	5	/d/-initial one-syllable words	83.33	—	83.33	—
	5	/g/-initial one-syllable words	76.67	—	83.33	—
Wambaugh et al. (1999)	6	/p/-initial one-syllable words	93.33	100	90.00	90
	6	/k/-initial one-syllable words	46.67	90	33.33	90
	6	/ʃ/-initial one-syllable words	20.00	40	23.33	20
Wambaugh (2004)	7	/v/-initial one- and two-syllable words	93.33	100	93.33	100
	7	/r/-blends–initial one- and two-syllable words	93.33	80	53.33	50
	8	/ʃ/-syllable–final one-syllable words in two-word phrases	100.00	100	80.00	100
	8	/z/-syllable–initial one- and two-syllable words in two-word phrases	85.33	89	70.00	36
Wambaugh & Nessler (2004)	9	/s/-initial one-syllable CV(C) words	96.00	100	—	—
	9	/p/-initial one-syllable CV(C) words	100.00	75	—	—
	9	/v/-initial one-syllable CV(C) words	37.67	13	—	—
	9	/k/-initial one-syllable CV(C) words	79.67	100	—	—
	9	/ʃ/-initial one-syllable CV(C) words	42.00	100	—	—
	9	/dʒ/-initial one-syllable CV(C) words	0.00	88	—	—
	9	/l/-initial one-syllable CV(C) words	92.00	100	—	—
	9	/m/-initial one-syllable CV(C) words	96.00	63	—	—
	9	/n/-initial one-syllable CV(C) words	100.00	100	—	—
Wambaugh & Mauszycki (2010)	10	/b/-initial one-syllable CV(C) words	54.67	100	60.00	100
	10	/s/-initial one-syllable CV(C) words	21.00	75	53.33	60
	10	/l/-initial one-syllable CV(C) words	91.67	100	60.00	80
	10	/m/-initial one-syllable CV(C) words	87.67	50	66.67	0
	10	/d/-initial one-syllable CV(C) words	42.00	38	33.33	0
	10	/f/-initial one-syllable CV(C) words	100.00	88	93.33	100

(*table continues*)

Table 4 *(Continued).*

Primary study	Meta-analysis participant no.	Treatment target	Treated items		Untreated items	
			Mean of last three treatments	First follow-up value	Mean of last three treatments	First follow-up value
Wambaugh et al. (2013)	11	/s/-initial four-syllable words (IB)	53.33	50	33.33	50
	11	/ɪ/- and /ɛ/-initial three-syllable words (IR)	71.67	95	33.33	60
	11	/r/-initial four-syllable words (TR)	56.67	60	20.00	70
	11	/ɑ/-, /æ/-, /i/-, or /o/-initial three-syllable words (TB)	81.67	85	36.67	50
	12	/θ/-initial or final two-syllable words (IR)	75.00	60	40.00	70
	12	/ʃ/-initial or final two-syllable words (IB)	85.00	70	33.33	30
	12	/z/-initial or final two-syllable words (TB)	83.33	55	23.33	20
	12	/dʒ/-initial or final two-syllable words (TR)	71.67	100	36.67	10
	13	/sw/-initial one-syllable or /θ/-final two-syllable words (TR)	31.67	15	1.67	0
	13	/sm/-initial one-syllable or /ʃ/-final two-syllable words (TB)	55.00	10	13.33	0
	13	/bl/-initial one-syllable or /z/-final two-syllable words (IB)	55.00	25	10.00	20
	13	/fl/-initial one-syllable or /dʒ/-final two-syllable words (IR)	48.33	10	23.33	10
	14	/θ/- or /br/-initial three-syllable words (TB)	45.00	25	20.00	20
	14	/z/- or /gl/-initial three-syllable words (TR)	58.33	25	30.00	20
	14	/st/- or /t/-initial three-syllable words (IR)	33.33	15	26.67	20
	14	/ʃ/- or /sk/-initial three-syllable words (IB)	56.67	35	30.00	30
Wambaugh et al. (2014)	15	/pr/-initial three-syllable words	46.67	30	33.33	0
	15	/ks/-medial three-syllable words	50.00	50	26.67	40
	15	/gl/-initial three-syllable words	66.67	40	13.33	0
	15	/nd/-medial three-syllable words	63.33	50	20.00	40
	16	/z/-final one-syllable words	83.33	70	46.67	20
	16	/st/-initial one-syllable words	83.33	0	86.67	0
	16	/f/-final one-syllable words	93.33	70	86.67	80
	16	/gl/-initial one-syllable words	76.67	90	66.67	60
	17	/dʒ/-initial one-syllable words	93.33	80	80.00	80
	17	/sm/-initial one-syllable words	100.00	100	100.00	60
	17	/θ/-initial one-syllable words	96.67	70	80.00	20
	17	/gl/-initial one-syllable words	66.67	50	60.00	80
	18	/f/-final one-syllable words	96.67	60	100.00	40
	18	/sm/-initial one-syllable words	96.67	90	100.00	100
	18	/bl/-initial one-syllable words	70.00	90	93.33	40
	18	/p/-final one-syllable words	96.67	100	86.67	80
	19	/gl/-initial two-syllable words	83.33	70	73.33	80
	19	/kw/-medial two-syllable words	100.00	90	40.00	40
	19	/θr/-initial two-syllable words	93.33	100	93.33	100
	19	/st/-medial two-syllable words	96.67	100	100.00	60
	20	/θr/-initial two-syllable words	93.33	70	93.33	60
	20	/gl/-initial two-syllable words	96.67	40	66.67	40
	20	/sn/-initial three-syllable words	63.33	50	73.33	40
	20	/kw/-initial three-syllable words	70.00	30	80.00	100

(table continues)

Table 4 *(Continued).*

Primary study	Meta-analysis participant no.	Treatment target	Treated items Mean of last three treatments	Treated items First follow-up value	Untreated items Mean of last three treatments	Untreated items First follow-up value
Wambaugh et al. (in press)	21	/θ/-medial three- and four-syllable words (B)	80.00	80	44.33	100
	21	/s/-blends initial three- and four-syllable words (B)	40.00	60	0.00	0
	21	/br/-initial or medial three- and four-syllable words (R)	66.67	80	33.00	33
	21	/l/-medial three- and four-syllable words (R)	93.33	80	22.00	33
	22	/st/-initial four- and five-syllable words (R)	73.33	100	78.00	33
	22	/tr/-medial four- and five-syllable words (R)	100.00	100	11.00	33
	22	/θ/-medial four- and five-syllable words (B)	86.67	80	86.67	80
	22	/fl/-initial and medial four- and five-syllable words (B)	66.67	80	44.33	67
	23	/st/-initial three- and four-syllable words (B)	73.33	80	33.33	33
	23	/dʒ/-medial three- and four-syllable words (B)	53.33	20	55.67	67
	23	/gl/-initial three- and four-syllable words (R)	73.33	80	22.00	0
	23	/θ/-medial three- and four-syllable words (R)	93.33	100	33.33	33
	24	/kw/-medial four-syllable words (R)	93.33	60	22.00	33
	24	/br/-medial four-syllable words (R)	100.00	20	0.00	33
	24	/fl/-medial four-syllable words (B)	93.33	100	22.00	0
	24	/sp/-medial four-syllable words (B)	86.67	40	22.00	33

Note. Em dashes indicate necessary phase data not collected in the original study. CV(C) = consonant–vowel or consonant–vowel–consonant syllable structure; IB = intense blocked; TR = traditional random; IR = intense random; TB = traditional blocked; B = blocked; R = random.

Results

Studies

The search for SPT studies yielded 10 reports. SCEDs were used in all of the reports. All of the SCEDs were multiple baselines across behaviors designs, and five of the studies included replications across participants.

Effect Size Descriptive Statistics

Descriptive statistics for all of the effect sizes and for the effect sizes weighted within participants are reported in Table 5. Compared with the unweighted values, the weighted values have similar means but reduced variance, range, and number. Treatment phase and follow-up phase mean effect sizes for treated items were significantly larger than those for untreated items, as indicated by significant paired-samples t tests; treatment phase effect size: $t(22) = 4.233$, $p < .001$; follow-up phase effect size: $t(21) = 5.210$, $p < .001$.

Benchmarks for Acquisition and Follow-Up Effects of SPT

The quartiles of the distribution of mean effect sizes, which represent the benchmarks, are shown in Table 6. There are four sets of benchmarks: treatment phase and follow-up phase effect sizes for both treated and untreated items. The first, second, and third quartiles in each set represent benchmarks for small, medium, and large effect sizes, respectively. For treated items, the treatment phase effect size benchmarks are 5.23, 6.98, and 9.65 SD units, respectively, and the follow-up phase effect size benchmarks are 5.90, 7.12, and 10.19 SD units, respectively. For untreated items, the treatment phase effect size benchmarks are 2.25, 3.75, and 6.66 SD units, respectively, and the follow-up phase effect size benchmarks are 2.59, 4.23, and 6.47 SD units, respectively. Note that these benchmarks differ substantially from Cohen's (1988) original benchmarks: The benchmarks introduced here were specifically developed for SPT and related AOS treatments, when effect sizes are calculated as described above.

Correlations Between Levels of Performance for Treated and Untreated Items and Different Study Phases

The results of the correlations between the levels of performance in different study phases and for treated and untreated items are shown in Table 7. The level of performance on treated items achieved by the last three treatment phase probes was found to have a highly significant, though moderately sized, association with the level of performance on untreated items during the same probes ($r_s = .487$, $p < .001$). In addition, performance on treated items during the last follow-up probe had a highly significant, moderately sized association with the performance on untreated items during the same probe ($r_s = .543$, $p < .001$).

The level of performance on treated items achieved by the last three treatment phase probes was found to be significantly associated with the magnitude of the first follow-up value ($r_s = .498$, $p < .001$). The same was found to be true for untreated items ($r_s = .640$, $p < .001$).

Correlations With Participant Characteristics

Spearman correlation coefficients between mean effect size and MPO, AOS severity, WAB, AQ, and age are reported in Table 8. With significance set at $p < .01$ to reduce chances of Type I error due to the number of correlations performed, the only significant correlation was a moderate positive relationship between AOS severity and the treatment phase effect size for untreated response generalization items ($r_s = .528$, $p = .01$). Although strictly nonsignificant, four other correlations approached significance, having p values less than .05. AOS severity had a positive correlation with follow-up phase effect size for untreated items ($r_s = .438$, $p = .041$), and WAB AQ had negative correlations with follow-up phase effect size for treated items ($r_s = -.517$, $p = .012$), treatment phase effect size for untreated items ($r_s = -.482$, $p = .020$), and follow-up phase effect size for untreated items ($r_s = -.466$, $p = .029$).

Table 5. Descriptive statistics of effect sizes (ES).

Target	Variable	Treated items		Untreated items	
		Treatment phase ES	Follow-up phase ES	Treatment phase ES	Follow-up phase ES
All targets	N	94	94	94	94
	M	8.82	7.88	4.69	4.70
	SD	5.45	4.38	3.88	4.33
	Minimum	−0.17	0.94	−0.50	−0.77
	Maximum	25.50	26.08	17.13	20.79
Targets weighted within participants	N	24	23	23	22
	M	8.42	8.03	4.54	4.99
	SD	4.60	2.69	2.99	3.00
	Minimum	3.24	4.47	0.85	1.53
	Maximum	22.00	14.17	10.96	12.67

Table 6. Quartiles of the distribution of weighted effect sizes (ES; sound production treatment ES benchmarks).

	Treated items		Untreated items	
Quartile	Treatment phase ES	Follow-up phase ES	Treatment phase ES	Follow-up phase ES
1	5.23	5.90	2.25	2.59
2	6.98	7.12	3.75	4.23
3	9.65	10.19	6.66	6.47

Discussion

Although effect sizes resulting from SPT vary by participant and articulatory target, they are predominantly large and positive. This is consistent with Strom's (2008) meta-analytical finding that the general category of articulatory-kinematic treatments is efficacious for treatment of AOS. On the basis of the present meta-analysis, SPT may be expected to lead to improvements in sound production of both treated target and untreated exemplars for speakers with AOS representing broad ranges of chronicity, AOS and aphasia severity, and age.

The results of the meta-analysis provide several important benefits. First, the effect size quartiles found may be considered benchmarks for quantifying the magnitude of change observed in sound production of persons receiving treatment for AOS. These benchmarks may serve as the basis for evaluating the effects of AOS interventions utilizing similar outcome measures in both clinical and research settings. For example, clinicians can use the benchmarks to measure the progress of individual patients. Researchers can use the benchmarks for comparing their own SPT investigations or comparing other AOS treatments to SPT. This is an important step toward establishing optimal treatment protocols and furthering our understanding of the relative effects of AOS interventions.

Second, the correlations with participant descriptors provide preliminary clues about what type of participants may be considered suitable candidates for SPT. The evidence presented here suggests that increases in AOS severity are associated with increases in effect sizes—at least for untreated exemplars of targets. In other words, the effect of SPT on untreated response generalization items may be greater for those with greater AOS severity, and as AOS severity lessens, the magnitude of change associated with SPT

may also decrease. The observation that aphasia severity approaches a significant negative associative relationship with SPT effect sizes, with greater aphasia severity (lower WAB AQ) being associated with larger effect sizes, supports this hypothesis. Further research examining the relationship between AOS severity and aphasia may clarify the relative contributions of linguistic and motor speech deficits to response to SPT intervention.

Also, the set of correlations examining associations between levels of performance on treated and untreated items provides clinically useful information. The correlation of level of performance on treated items at the end of treatment gives a highly significant, albeit moderate, estimate of the participant's performance on untreated items both at the end of treatment and at follow-up. Because response generalization is a key factor in successful treatment, this significant finding is encouraging. Clinicians and researchers may reasonably expect generalization to untreated exemplars of the targets they are treating with SPT, provided their protocol and participant characteristics are similar to those examined here.

Limitations and Further Research

There are many possible avenues for further SPT research. In particular, replications are needed across research laboratories. In their design standards, Kratochwill et al. (2010) provided suggestions for minimum requirements for combining studies, including the need for studies to represent the work of multiple independent laboratories. However, it is worth noting that the SPT articles included in this meta-analysis were published by a variety of collaborators and in multiple journals, which adds robustness to the findings.

Table 7. Spearman correlations (*p* values) between levels of performance in different phases.

		Treated items		Untreated items	
Probe set	Variable	Mean of last three treatment phase probes	First follow-up value	Mean of last three treatment phase probes	First follow-up value
Treated items	Mean of last three treatment phase probes	1.000	.498 (.000)**	.487 (.000)**	.375 (.001)**
	First follow-up value		1.000	.468 (.000)**	.543 (.000)**
Untreated items	Mean of last three treatment phase probes			1.000	.640 (.000)**
	First follow-up value				1.000

**Correlation is significant at the .01 level (two tailed).

Table 8. Spearman correlations (*p* values) between participant characteristics and effect sizes (ES).

Variable	Treated items		Untreated items	
	Treatment phase ES	Follow-up phase ES	Treatment phase ES	Follow-up phase ES
MPO	.094 (.662)	−.181 (.407)	−.234 (.282)	−.262 (.240)
AOS severity	.186 (.384)	.410 (.052)	.528 (.010)**	.438 (.041)*
WAB AQ	−.226 (.228)	−.517 (.012)*	−.482 (.020)*	−.466 (.029)*
Age (years)	−.278 (.189)	.080 (.716)	.140 (.523)	.099 (.661)

Note. Apraxia of speech (AOS) severity is based on a numeric scale ranging from 1 (*mild*) to 5 (*severe*). MPO = months postonset of stroke; WAB AQ = Western Aphasia Battery Aphasia Quotient.

*Correlation is significant at the .05 level (two tailed). **Correlation is significant at the .01 level (two tailed).

The effects of the individual components of SPT are unknown. It is possible that some of the ingredients are unnecessary whereas others are critical. A recent investigation of the effects of repeated practice (no additional treatment ingredients other than limited verbal feedback) with speakers with AOS indicated that repetition alone can result in clinically significant improvements in articulation (Wambaugh, Nessler, Cameron, & Mauszycki, 2012). It is likely that repeated production of the target word is an important ingredient of SPT. Future studies that address the effects of other treatment ingredients, such as integral stimulation, could lead to the development of more efficacious treatments. Likewise, various aspects of treatment delivery such as intensity of treatment are deserving of attention.

One particularly relevant issue that is deserving of further research and relevant to SPT (as well as other AOS treatments) is the size of speech motor programs. Although little is known about the neural and cognitive mechanisms associated with the therapeutic effects of SPT, models of speech production provide frameworks that may explain the mechanism of the effects of SPT for speakers with AOS. For example, put in the framework of the directions into velocities of articulators model (Tourville & Guenther, 2011), SPT may improve speech production in AOS by restoring speakers' access to speech sound maps, repairing these motor programs themselves, and/or supplementing the speaker's feedback tuning system through the provision of auditory models and opportunities for repeated practice (Wambaugh, 2010). In terms of mechanisms of action of SPT, generalization to untrained exemplars may occur because untrained exemplars reflect a speech sound map that is the same as or related to that of trained items. Generalization items in SPT investigations were selected to represent phonetic and syllabic contexts that were similar to trained items. The positive response generalization effects associated with SPT are consistent with Schoor, Aichert, and Ziegler's (2012) findings concerning transfer (generalization) effects with massed practice of syllables. Schoor et al. (2012) reported positive transfer when there was "position-true overlap on syllable constituents of the training and transfer syllables" (p. 892). However, greater understanding and awareness of treatment response classes (Wambaugh, 2006) may inform the further development of AOS treatments such as SPT.

As indicated previously, there has been limited study of the stimulus generalization effects of SPT. This is also true for the majority of AOS treatments (Ballard et al., 2015; Wambaugh et al., 2006a). Research focused on developing and utilizing reliable measures to assess stimulus generalization and ecologically valid outcomes is sorely needed. Such outcome measures would specify changes in the level of functional use of targeted sounds following SPT.

As with all meta-analyses, this meta-analysis is based on the contents of its primary studies, the existence of which could have been subject to publication bias. However, all SPT investigations from our laboratory have been reported or are in preparation. To date, we have recruited only participants who have a minimum ability to repeat monosyllabic words and have excluded participants with AOS who failed to demonstrate a sufficient number of erroneous productions. As a consequence, the current SPT publications likely reflect a participant selection bias. Future research examining any predisposing factors to nonresponsiveness to SPT would make a valuable contribution to the state of AOS treatment. It is hoped that these benchmarks will be used and that usage of these benchmarks will assist in clinical settings for evaluating patient progress as well as in the research setting for comparison between treatments or different types of participants.

References

Ballard, K. J., Wambaugh, J. L., Duffy, J. R., Layfield, C., Maas, E., Mauszycki, S., & McNeil, M. R. (2015). Treatment for acquired apraxia of speech: A systematic review of intervention

research between 2004 and 2012. *American Journal of Speech-Language Pathology, 24,* 316–337.

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review, 16,* 161–169. doi:10.1007/s11065-006-9013-7

Bloom, M., Fischer, J., & Orme, J. G. (2003). *Evaluating practice: Guidelines for the accountable professional* (4th ed.). Boston, MA: Allyn & Bacon/Pearson Education.

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single case research design and analysis: New directions for psychology and education* (pp. 187–212). Hillsdale, NJ: Erlbaum.

Campbell, J. M. (2013). Commentary on PND at 25. *Remedial and Special Education, 34,* 20–25.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dabul, B. L. (2000). *Apraxia Battery for Adults* (2nd ed.). Austin, TX: Pro-Ed.

Duffy, J. R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management* (3rd ed.). St. Louis, MO: Elsevier Mosby.

Evans, J. J., Gast, D. L., Perdices, M., & Manolov, R. (2014). Single case experimental designs: Introduction to a special issue. *Neuropsychological Rehabilitation, 24,* 305–314.

Haley, K. L., Bays, G. L., & Ohde, R. N. (2001). Phonetic properties of aphasic-apraxic speech: A modified narrow transcription analysis. *Aphasiology, 15,* 1125–1142. doi:10.1080/02687040143000537

Kertesz, A. (1982). *The Western Aphasia Battery*. New York, NY: Grune & Stratton.

Kertesz, A. (2007). *The Western Aphasia Battery–Revised*. San Antonio, TX: Pearson.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single case designs technical documentation*. What Works Clearinghouse. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

LaPointe, L. L. (1984). Sequential treatment of split lists: A case report. In J. C. Rosenbek, M. R. McNeil, & A. E. Aronson (Eds.), *Apraxia of speech: Physiology, acoustics, linguistics, management*. San Diego, CA: College-Hill Press.

Maggin, D. M., & Chafouleas, S. M. (2013). Introduction to the special series: Issues and advances of synthesizing single case research. *Remedial and Special Education, 34*(1), 3–8.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Medicine, 6*(7), e1000097. doi:10.1371/journal.pmed.1000097

Rohatgi, A. (2014). WebPlotDigitizer [Computer software]. Retrieved from http://arohatgi.info/WebPlotDigitizer/app/?

Rosenbek, J. C., Lemme, M. L., Ahern, M. B., Harris, E. H., & Wertz, R. T. (1973). A treatment for apraxia of speech in adults. *Journal of Speech and Hearing Disorders, 38,* 462. doi:10.1044/jshd.3804.462

Schoor, A., Aichert, I., & Ziegler, W. (2012). A motor learning perspective on phonetic syllable kinships: How training effects transfer from learned to new syllables in severe apraxia of speech. *Aphasiology, 26,* 880–894. doi:10.1080/02687038.2012.660458

Strom, K. E. (2008). *Treatment efficacy in acquired apraxia of speech: A meta-analysis*. University of Oklahoma, Norman.

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes, 26,* 952–981. doi:10.1080/01690960903498424

Wambaugh, J. L. (2004). Stimulus generalization effects of sound production treatment for apraxia of speech. *Journal of Medical Speech-Language Pathology, 12,* 77–97.

Wambaugh, J. L. (2006). Treatment guidelines for apraxia of speech: Lessons for future research. *Journal of Medical Speech-Language Pathology, 14,* 317–321.

Wambaugh, J. L. (2010). Sound production treatment for acquired apraxia of speech. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders, 20,* 67–72.

Wambaugh, J. L., & Cort, R. (1998, March). *Treatment for AOS: Perceptual and VOT changes in sound production*. Poster presented at the Motor Speech Conference, Tucson, AZ.

Wambaugh, J. L., Duffy, J. R., McNeil, M. R., Robin, D. A., & Rogers, M. A. (2006a). Treatment guidelines for acquired apraxia of speech: A synthesis and evaluation of the evidence. *Journal of Medical Speech-Language Pathology, 14*(2), xv–xxxiii.

Wambaugh, J. L., Duffy, J. R., McNeil, M. R., Robin, D. A., & Rogers, M. A. (2006b). Treatment guidelines for acquired apraxia of speech: Treatment descriptions and recommendations. *Journal of Medical Speech-Language Pathology, 14*(2), xxxv–lxvii.

Wambaugh, J. L., Kalinyak-Fliszar, M. M., West, J. E., & Doyle, P. J. (1998). Effects of treatment for sound errors in apraxia of speech and aphasia. *Journal of Speech, Language, and Hearing Research, 41,* 725–743.

Wambaugh, J. L., Martinez, A. L., McNeil, M. R., & Rogers, M. A. (1999). Sound production treatment for apraxia of speech: Overgeneralization and maintenance effects. *Aphasiology, 13,* 821–837.

Wambaugh, J. L., & Mauszycki, S. C. (2010). Sound production treatment: Application with severe apraxia of speech. *Aphasiology, 24,* 814–825. doi:10.1080/02687030903422494

Wambaugh, J. L., & Nessler, C. (2004). Modification of sound production treatment for apraxia of speech: Acquisition and generalisation effects. *Aphasiology, 18,* 407–427. doi:10.1080/02687030444000165

Wambaugh, J. L., Nessler, C., Cameron, R., & Mauszycki, S. C. (2012). Acquired apraxia of speech: The effects of repeated practice and rate/rhythm control treatments on sound production accuracy. *American Journal of Speech-Language Pathology, 21*(2), S5–S27. doi:10.1044/1058-0360(2011/11-0102)

Wambaugh, J. L., Nessler, C., Cameron, R., & Mauszycki, S. C. (2013). Treatment for acquired apraxia of speech: Examination of treatment intensity and practice schedule. *American Journal of Speech-Language Pathology, 22,* 84–102. doi:10.1044/1058-0360(2012/12-0025)

Wambaugh, J. L., Nessler, C., Wright, S., & Mauszycki, S. C. (2014). Sound production treatment: Effects of blocked and random practice. *American Journal of Speech-Language Pathology, 23,* S225–S245. doi:10.1044/2014_AJSLP-13-0072

Wambaugh, J. L., Nessler, C., Wright, S., & Mauszycki, S. (in press). Sound production treatment for acquired apraxia of speech: Effects of blocked and random practice on multisyllabic word production. *International Journal of Speech-Language Pathology.*

Wambaugh, J. L., West, J. E., & Doyle, P. J. (1998). Treatment for apraxia of speech: Effects of targeting sound groups. *Aphasiology, 12,* 731–743.

Wertz, R. T., LaPointe, L. L., & Rosenbek, J. C. (1984). *Apraxia of speech in adults: The disorder and its management*. Orlando, FL: Grune & Stratton.