ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Predicting Treatment Outcomes in Rapid Syllable Transition Treatment: An Individual Participant Data Meta-Analysis

Wei Lin Ng,[a] Patricia McCabe,[a] Rob Heard,[a] Veronica Park,[a] Elizabeth Murray,[a] and Donna Thomas[a]

[a] The University of Sydney, New South Wales, Australia

ARTICLE INFO

Article History:
Received November 22, 2021
Revision received January 17, 2022
Accepted January 22, 2022

Editor-in-Chief: Bharath Chandrasekaran
Editor: Julie D. Anderson

https://doi.org/10.1044/2022_JSLHR-21-00617

ABSTRACT

Purpose: The purpose of this study is to identify predictors of treatment outcomes in Rapid Syllable Transition Treatment (ReST) for childhood apraxia of speech through an individual participant data meta-analysis.
Method: A systematic literature search identified nine ReST studies for inclusion. Individual participant data were obtained, and studies were coded for methodological design, baseline participant characteristics, service delivery factors, and treatment outcomes. Bivariate analyses were conducted to identify potential predictor variables. Multiple linear regressions were then performed to identify predictors of treatment outcomes.
Results: Data for 36 participants from seven studies were included in the statistical analyses. In multivariate modeling, better performance on treated pseudowords posttreatment was predicted by higher baseline expressive language and Goldman-Fristoe Test of Articulation scores, lower speech inconsistency and percentage of vowels correct, and higher pretreatment accuracy on pseudoword targets. Better performance on untreated real words posttreatment was predicted by higher pretreatment accuracy on real words. Gains in performance and retention of gains were not significantly predicted by any individual variable or combination of variables.
Conclusions: Baseline speech and expressive language skills and accuracy on pseudowords and real words were significant predictors of absolute posttreatment performance. Regardless of baseline characteristics, all children were statistically as likely to achieve gains during ReST and retain these gains for up to 4 weeks posttreatment. Large-scale prospective research is required to further examine the effects of dose frequency and co-occurring language impairments on treatment outcomes and the complex co-effects of percentage of vowels correct with other potential predictors.
Supplemental Material: https://doi.org/10.23641/asha.19611714

Childhood apraxia of speech (CAS) is a neurological speech sound disorder characterized by deficits in motor planning and programming of movements required for accurate speech sound production and prosody (American Speech-Language-Hearing Association [ASHA], 2007). CAS may occur as a result of a neurological impairment or genetic variation, in association with other neurobehavioral disorders (e.g., fragile X syndrome, galactosemia), or as an idiopathic speech sound disorder (Morgan & Webster, 2018). The speech of children with CAS is characterized by "(a) inconsistent errors on consonants and vowels in repeated productions of syllables or words; (b) lengthened and disrupted coarticulatory transitions between sounds and syllables; and (c) inappropriate prosody, especially in the realization of lexical or phrasal stress" (ASHA, 2007). This speech disorder can be mild to profound in severity and

Correspondence to Patricia McCabe: tricia.mccabe@sydney.edu.au.
Disclosures: *Portions of this work were presented at the 2021 Speech Pathology Australia National Conference. The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

generally persists across the life span. Children with this phenotype of speech disorder also often present with or are at risk for persistent motor, language, and literacy difficulties (e.g., Stein et al., 2020) and appear to be at an increased risk of psychosocial issues in adolescence and adulthood (e.g., Lewis et al., 2021).

Corresponding to the varying difficulties children with CAS present with a range of treatment methods, including motor-based treatments (e.g., dynamic temporal and tactile cueing [DTTC]; Strand 2020), linguistic approaches (e.g., Integrated Phonological Awareness; McNeill et al., 2009), and augmentative and alternative communication (e.g., Cumley & Swanson, 1999), have been reported (Morgan et al., 2018). In a systematic review of treatment outcomes for children with CAS, three treatments were identified as suitable for clinical use with enough evidence to suggest likely improvements to children's speech (Murray et al., 2014). These were Rapid Syllable Transition Treatment (ReST; e.g., Ballard et al., 2010), DTTC (e.g., E. Maas & Farinella, 2012), and Integrated Phonological Awareness (e.g., McNeill et al., 2009). Since that review, a single randomized controlled trial has reported efficacy of the Nuffield Dyspraxia Program alongside ReST (Murray et al., 2015) and, therefore, four treatments are likely to be effective with this population.

ReST

ReST is based on principles of motor learning (see the work of E. Maas et al., 2008), a set of practice and feedback conditions optimal for learning motor skills, and was designed to directly address the core deficit of dysprosody in CAS (Ballard et al., 2010). It uses two- or three-syllable pseudowords containing varied stress patterns and sounds that model real words (e.g., "dorfa" and "kadefee"), allowing children with CAS to practice motor planning and programming without the influence of existing learned motor patterns (Ballard et al., 2010). Each treatment session comprises a training phase and a practice phase, in which random, high-intensity practice of pseudowords is conducted (Thomas et al., 2014). In the training phase, target words are introduced, and the child is provided with cues and knowledge of performance feedback, where the clinician corrects the child's production to achieve correct productions (McCabe, Thomas, & Murray, 2020). The practice phase is designed around principles of motor learning, involving a high number of target productions, variable practice, random presentation of targets, and low-frequency knowledge of results feedback, where the child is told if the word was said correctly or incorrectly on 50% of practice items (McCabe, Thomas, & Murray, 2020).

ReST has consistently demonstrated positive treatment effects in previous studies, where children showed improved production of treated pseudowords on perceptual measures of phoneme accuracy, lexical stress contrasts, and segmental accuracy during treatment (e.g., McCabe et al., 2014; Murray et al., 2015). Generalization of treatment effects to untreated real words and pseudowords, and maintenance of these effects up to 4 months posttreatment has also been demonstrated (e.g., Ballard et al., 2010; Murray et al., 2015). Additionally, positive treatment effects has been replicated across various service delivery models, including lower dose frequency and telehealth delivery (Thomas et al., 2014, 2016), as well as adapted versions of ReST for Korean and Italian children (Oh & Ha, 2021; Scarcella et al., 2021). Positive treatment effects have also been observed following Treatment for Establishing Motor Program Organization, a treatment similar to ReST (Miller et al., 2021).

Although the efficacy of ReST has been demonstrated by previous studies, variations in individual response to treatment are present within and across studies. For example, in the work of McCabe et al. (2014), while all children showed improved production of treated pseudowords and generalization to untreated words and connected speech, the magnitude of treatment effect, generalization, and retention varied across participants. Little is known about the factors that predict individual response to treatment, with previous studies identifying a pressing need for research in this area (e.g., Murray et al., 2015; Thomas et al., 2018). At present, speech-language pathologists depend on clinical judgment and making comparisons to children who have participated in research to date to gauge the suitability of ReST for a child. Identifying predictors of performance may allow for treatment to be accurately matched to participant characteristics, informing evidence-based practice for treating children with CAS.

Metrics of Treatment Outcomes

Treatment outcomes can be quantified in three ways: (a) absolute performance that reflects performance at a given point in time, (b) gain in performance that is derived from the difference in performance at two points in time, and (c) retention of gains that measures change in performance in the period following treatment. For our study, all three metrics were considered as absolute posttreatment performance does not account for a participant's baseline performance or retention of gains. To illustrate, a participant who improves from 15% accuracy at baseline to 80% accuracy posttreatment (i.e., 65% gain) has probably made more significant improvement than a participant who improves from 45% to 90% (i.e., 45% gain). A higher absolute posttreatment performance is not necessarily synonymous with greater improvement. Consequently, identifying predictors of absolute posttreatment performance is different to identifying predictors of gain in performance.

Possible Predictors of Treatment Outcomes

To date, no studies have examined predictors of treatment outcomes in CAS intervention. Although limited, there have been studies examining predictors of gain in performance following intervention for speech sound disorders. In a group of preschool children with impairments in phonology and morphosyntax, Tyler et al. (2003) found that a highly inconsistent phonological system was associated with greater phonological change from baseline to 2 weeks posttreatment. For school-aged children with speech sound disorders receiving school-based services, speech sound severity was reportedly the only factor that significantly predicted gain, where children with lower percentage of consonants correct (PCC) at baseline made greater gains from baseline to posttreatment (Farquharson et al., 2019). Age and language ability did not significantly predict gain.

Child Factors

Possible reasons for variations in individual response to the treatment have been suggested in the literature. In the work of Ballard et al. (2010), the child with the most severe dysprosody at baseline showed poorer performance during treatment and did not maintain treatment effects 4 weeks posttreatment. Similarly, in the work of Thomas et al. (2014), children with more severe speech difficulties at baseline responded more slowly to treatment and demonstrated weaker generalization of skills to untreated targets. Conversely, children with milder speech difficulties at baseline showed stronger generalization of skills to untreated targets. This was replicated in the work of Thomas et al. (2016), where older participants with milder initial speech difficulties generalized skills to untreated targets and more complex behaviors of using treated pseudowords in carrier phrases (e.g., "I want a keedefa"). Collectively, these findings suggest that age and/or speech severity at baseline are likely to predict treatment outcomes in ReST.

Similar predictors of performance have been suggested in a study investigating the efficacy of DTTC, another motor-based treatment that incorporates principles of motor learning to treat children with severe CAS through the use of a cueing hierarchy, graded shaping of speech movements, and repeated practice of these movements (Strand, 2020). In the work of E. Maas and Farinella (2012), the child who did not respond to treatment was the youngest of four participants and had the most severe CAS. It was suggested that younger children and/or children with more severe CAS may experience cognitive overload when presented with complex tasks such as the random presentation of stimuli (E. Maas & Farinella, 2012), as used in studies of both ReST and DTTC. Thus, age and/or initial speech severity may have a general effect on motor learning and predict treatment outcomes in ReST, where the effectiveness of using pseudowords to improve real-word productions depends on the child's ability to generalize acquired motor planning and programming skills (E. Maas et al., 2014).

While some studies have suggested age and/or initial speech severity as possible predictors of treatment outcomes in ReST, a few studies have indicated otherwise. Although age was significantly correlated with absolute performance on treated pseudowords 1 week posttreatment, Murray et al. (2015) found that age and severity did not correlate with overall change from baseline to 4 months posttreatment. Thomas et al. (2018) reported similar findings, where no correlation was found between effect size and age or between effect size and initial speech severity. Considering the contrasting evidence in the literature, age and speech variables have been examined as potential predictors here.

Dose Frequency Factors

High treatment intensity is critical for motor learning and has shown to result in better treatment outcomes for children with CAS (e.g., Edeal & Gildersleeve-Neumann, 2011). A standardized set of terms for treatment intensity was proposed by Warren et al. (2007). This includes (a) dose: the number of productions by the child in one session, (b) dose form: the activity used to deliver the dose, (c) dose frequency: the number of times a dose is delivered (e.g., 1 hr per day, 4 days per week), (d) total intervention duration: the duration over which intervention is provided (e.g., 3 weeks), and (e) cumulative intervention intensity: a measure of overall intensity calculated by multiplying dose by dose frequency by total intervention duration.

In most ReST studies, participants produced 100 responses per session in the practice phase for four sessions per week across 3 weeks, giving a cumulative intervention intensity of 1,200 practice trials. In the work of Thomas et al. (2014), the effect of a lower dose frequency (i.e., twice-weekly ReST) was explored. Compared to 4 times per week of ReST, positive treatment effects were achieved with similar acquisition and generalization but no ongoing improvement following the end of treatment. However, the small sample size of the study did not allow for interpretation of individual variations in response and consideration of covariates (Thomas et al., 2014). Thus, the association between dose frequency and treatment outcomes in ReST has been examined in this study with a larger sample size alongside participant characteristics that could influence treatment outcomes as well.

Rationale for Individual Participant Data Meta-Analysis

As CAS is a relatively rare disorder with an estimated prevalence rate of one in 1,000 children (Shriberg

et al., 2019), previous ReST treatment studies have reported small participant numbers. The research designs of previous studies (i.e., mostly single-case experimental designs and two small randomized controlled trials) have also limited generalization of results (Iversen, 2013). Contrary to a study-level meta-analysis, an individual participant data (IPD) meta-analysis synthesizes raw data from individual participants in each study (Riley et al., 2010). It improves data quality, allows for inclusion of unpublished studies, and facilitates the use of standardized measures and statistical methods (Stewart et al., 2015). Particularly for this study, an IPD meta-analysis provides greater power to investigate the relationship between participant characteristics and treatment outcomes (Riley et al., 2010).

Aim and Research Questions

The aim of this study was to identify predictors of treatment outcomes that make ReST more effective for some children with CAS than others. Performance on treated pseudowords and untreated real words was examined to look at direct treatment effects and generalization of these effects. Three metrics were considered for each of these outcomes: (a) absolute performance at 1 week posttreatment, (b) change in performance (gain) from baseline to 1 week posttreatment, and (c) change in performance (retention) from 1 week to 4 weeks posttreatment. The following research questions were addressed:

1. Which factors predict absolute performance on treated pseudowords and untreated real words at 1 week posttreatment?
2. Which factors predict gain in performance on treated pseudowords and untreated real words from baseline to 1 week posttreatment?
3. Which factors predict retention of gains in performance on treated pseudowords and untreated real words from 1 week to 4 weeks posttreatment?

Method

This study is an IPD meta-analysis completed following a systematic search for eligible studies. Ethical approval was provided by the University of Sydney Human Research Ethics Committee for post hoc analysis of data obtained in included studies.

Eligibility Criteria

Studies were eligible if at least one participant with a confirmed diagnosis of CAS was treated with ReST and speech outcome measures were reported for at least one treated and one untreated behavior. Studies with participants who had co-occurring developmental or genetic disorders, or hearing, visual, or oral structural impairments were excluded. Detailed inclusion criteria are shown in Table 1.

Study Identification and Selection

A literature search was conducted in June 2020 on the following electronic databases: PsycINFO, Cumulative Index to Nursing and Allied Health Literature, Embase, Education Resources Information Center, Linguistic Language Behavior Abstracts, Medline, ProQuest Dissertations & Theses Global, Scopus, speechBITE, and Web of Science. The following search terms were used: (apraxia OR dyspraxia) AND (child*) AND (speech) AND (treatment or intervention or ReST or Rapid Syllable Transition*). Discussion with the second author indicated that the first ReST study commenced in 2007, and therefore, database searches were restricted to journal articles published and theses or dissertations uploaded between January 2006 and June 2020. A total of 1,808 studies were identified from these searches.

All references were exported to EndNote X8 for removal of duplicates. The remaining references were screened by title and abstract for relevance. Full texts of references that passed the initial screening were retrieved and assessed for inclusion. Eligibility of identified studies was assessed independently by the first author (W.L.N.) and a final-year speech-language pathology student. Interrater reliability for study inclusion was 89% at the title and abstract screening stage and 86% at the full-text stage. Reference lists of studies that met inclusion criteria were manually searched for relevant studies, adding one study for review (Schultz, 2018). Original authors of included studies were asked if they knew of any unpublished studies. The second author (P.M.) identified a manuscript that

Table 1. Inclusion criteria.

Aspect	Criteria
Design	Treatment study
	Conducted between January 2006 and June 2020
	Published or unpublished journal articles, theses, or dissertations
	Conducted and written in English
Intervention	ReST or ReST-like speech motor intervention
Participants	Aged between 4 and 13 years
	Confirmed diagnosis of CAS
	No co-occurring developmental or genetic disorder
	Normal hearing and vision
	Normal oral structure
Outcomes	Speech outcome measures (e.g., PCC) for at least one treated and one untreated behavior

Note. ReST = Rapid Syllable Transition Treatment; CAS = childhood apraxia of speech; PCC = percentage of consonants correct.

has been submitted for publication as well as an unpublished study for inclusion (McCabe, Preston, et al., 2020; Staples et al., 2009).

Data Extraction and Synthesis

Authors of eligible studies were contacted via e-mail to obtain permission to use their data and to request IPD. Where possible, data were extracted from published journal articles. Unreported or unavailable data were obtained from electronic spreadsheets and paper files provided by authors. Extracted data were thoroughly cross-checked for consistency across all available sources. Authors were contacted to resolve any inconsistencies and provide missing data. Data relating to study-level characteristics, including study design, intervention intensity, and mode of delivery, were extracted. IPD were extracted for age, sex, baseline assessment data, and probe data for treated pseudowords and untreated real words at baseline, 1 week, and 4 weeks posttreatment. Later data points (e.g., 4 months posttreatment) were available for some studies but not all and were thus excluded from the analysis. Change scores were derived by calculating the difference in accuracy of treated pseudowords and untreated real words from baseline to 1 week posttreatment and from 1 week to 4 weeks posttreatment.

Across studies, different tools or versions of tools were used to measure participants' baseline skills. Depending on age, participants completed either the Clinical Evaluation of Language Fundamentals (CELF) Preschool–Second Edition–Australian and New Zealand Standardised Edition (Wiig et al., 2006) or the Clinical Evaluation of Language Fundamentals–Fourth Edition–Australian Standardised Edition (Semel et al., 2006). As both have subtests measuring similar language domains, the expressive and receptive index scores obtained from these subtests were deemed comparable. The tools used to measure participants' baseline skills also varied by study. Forward digit span scores were obtained using the Comprehensive Test of Phonological Processing–Second Edition (Wagner et al., 2013) in the work of Thomas et al. (2018) and the Test of Auditory Processing Skills–Third Edition (Martin & Brownell, 2005) in the work of McCabe et al. (2014). Although different tools were used, *z* scores and similarity of tasks facilitated direct comparison of scores. Similarly, percentage of speech inconsistency was consistently calculated on three repetitions of 25 words, allowing for comparison of data.

Where data could not be compared directly, raw data were used to establish a standard metric. For example, unlike the other studies where percentage of vowels correct (PVC) and PCC were obtained from the Single Word Test of Polysyllables (Gozzard et al., 2006), measures reported in the works of McCabe et al. (2014) and McCabe, Preston, et al. (2020) were obtained from connected speech samples. To standardize data, PVC and PCC measures were manually calculated from the Single Word Test of Polysyllables in the paper files of participants from both of these studies. As data for nonword repetition scores and number of words with syllable segregations could not be standardized across studies, these variables were excluded from the analysis. Although oral structure and function scores from the Oral and Speech Motor Control Protocol (Robbins & Klee, 1987) were available for all participants, oral structure scores were excluded due to a lack of variation in scores attained by participants (M = 23, SD = 1). See Supplemental Material S1 for the assessment tools used to measure baseline variables.

Missing data were not imputed as the proportions of missing data for variables varied between 3% and 11%, only sometimes marginally above the suggested upper threshold of 10% where results of statistical analyses may be biased (Bennett, 2001). Additionally, as missing data were assumed to be missing at random, it is likely that results of the analyses based on available data are unbiased (Deeks et al., 2020). Where participant data were missing for any variable, those participants were excluded from the analysis for that variable.

Outcome Measure

The primary outcome measure was the percentage of correct responses for treated pseudowords and untreated real words as it has been used as the primary outcome measure across all included studies, facilitating data standardization. Although minor variations were noted across studies, a response was perceptually judged to be correct if it had the (a) correct sounds, (b) correct stress pattern, and (c) smooth transition between syllables (e.g., Murray et al., 2015; Thomas et al., 2014, 2018). This corresponds to observations of the three consensus-based features of CAS, namely, segments, transitions, and prosody (ASHA, 2007). The percentage of correct responses was obtained by calculating the number of correct productions over the number of attempts multiplied by 100.

Quality Assessment

As IPD quality depends on the quality of original studies (Riley et al., 2010), the methodological quality of individual studies was independently assessed by the first and fourth authors (W.L.N. and V.P.) who are not the authors of any included studies. Any disagreements were resolved through discussion.

Statistical Analyses

Both fixed-effect and random-effects models were considered for this meta-analysis. Although a random-effects

model is preferred when between-studies heterogeneity is present, as observed among the included studies, an estimate of the between-studies variance would likely lack precision due to the small number of studies (Borenstein et al., 2010). Therefore, a fixed-effect analysis was performed.

Preliminary data screening was conducted to ensure that assumptions of multivariate normality, linearity, and homoscedasticity were not violated. Given the large number of potential predictors for the small sample size and paucity of current research to inform variable selection, Pearson correlations were performed between variables and posttreatment scores to identify potential predictors of absolute performance and between variables and change scores to identify potential predictors of performance gain. Although $p < .05$ typically indicates statistical significance in behavioral science research (Cohen et al., 2013), a conservative threshold of $p < .2$ was chosen for this study to ensure that potentially important variables were included in the regression analyses (Hosmer et al., 2013). As the literature has indicated treatment intensity to be of interest in the treatment of children with CAS, dose frequency was included as a control in all regression models. A correlation matrix (see Supplemental Material S2) was used to screen for multicollinearity, as indicated by an absolute correlation coefficient of .8 and above between variables used as predictors in the regression analyses (Shrestha, 2020). PVC, oral function score, and percentage of lexical stress errors were highly correlated with each other. Thus, only one of the three variables was entered into the regression models at any 1 time. As PCC was also highly correlated with several potential predictor variables in initial runs of the regression models, it was subsequently excluded from all regression models.

To examine Research Question 1, multiple regression analyses were conducted using absolute performance on treated pseudowords and untreated real words at 1 week posttreatment. To minimize overfitting, only seven potential predictors could be entered into each regression model to meet a minimum of five events per variable (Tabachnick & Fidell, 1989, as cited in the work of Green, 1991). As more than seven potential predictors were identified in the bivariate correlations, variables were divided into two groups, speech and nonspeech, for entry into regression models. This resulted in two models for absolute performance on treated pseudowords, one for speech and the other for nonspeech variables, and another two corresponding models for real-word performance.

To examine Research Questions 2 and 3, multiple regression analyses were performed using change scores on treated pseudowords and untreated real words from baseline to 1 week posttreatment and from 1 week to 4 weeks posttreatment. As there were seven or less potential predictors identified in the bivariate correlations, variables did not have to be divided into speech and nonspeech. This resulted in two models to predict gain and retention of gains for treated pseudowords and in another two corresponding models for untreated real words. As proposed by Cohen (1988), R^2 values of .02, .13, and .26 in the regression analyses were interpreted as small, medium, and large magnitudes, respectively. Changes in the mean scores of treated pseudowords and untreated real words were analyzed using analysis of variance, with Cohen's d (M_1 - M_2/SD_{pooled}) as an effect size measure. As proposed by Cohen (1998, as cited in the work of Welkowitz et al., 2012), d values of 0.2, 0.5, and 0.8 were interpreted as small, medium, and large, respectively.

Results

Study Selection and IPD Obtained

The Preferred Reporting Items for Systematic Reviews and Meta-Analysis IPD flowchart in Figure 1 outlines the study selection process (Stewart et al., 2015). The literature search identified 1,076 studies for title and abstract screening. After screening, full text of the remaining 63 studies was reviewed and nine eligible studies were identified. IPD were sought and obtained from eight of nine studies, giving a total of 48 participants. Data were unavailable from Miller (2018)[1] as the author was unable to provide IPD within the requested time frame. After IPD were collected, the study of Ballard et al. (2010) was excluded from the study as the paper reported acoustic and perceptual measures that were substantially different to the other included studies and no similar dependent measures could be established from raw data in the paper files.

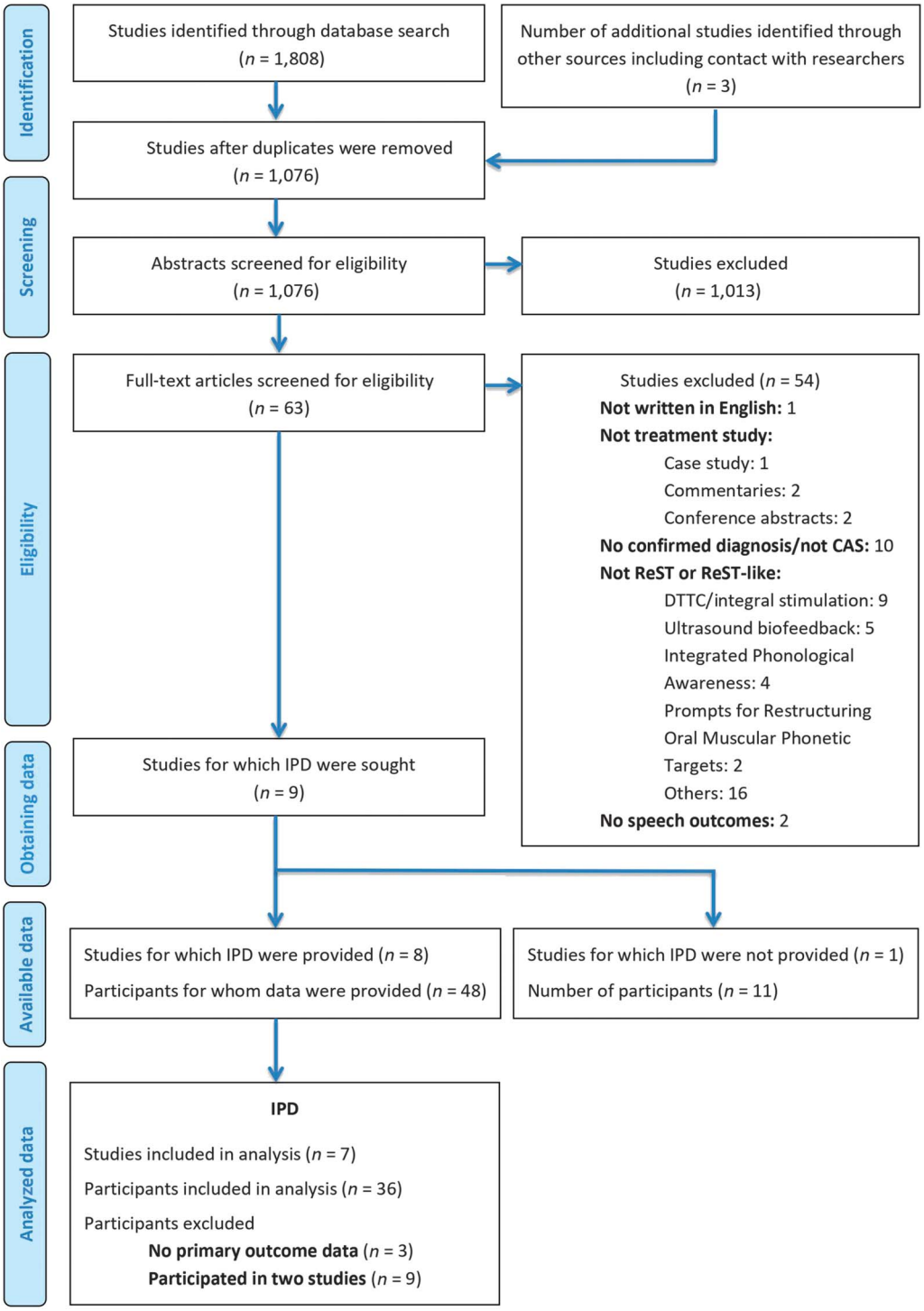
As linear regression analysis requires independence of observations, only data from one study could be included for nine participants who were in two studies. As these participants were part of one published and one unpublished study, data from the published study were retained and data from the unpublished study were excluded. This resulted in a total of 36 participants included in the statistical analyses.

Study and Participant Characteristics

Table 2 presents key features of the included studies. Among the seven studies, there were five published studies, one unpublished thesis, and one manuscript submitted for publication. In most studies ($n = 5$), ReST was conducted at a high dose frequency of 4 times per week

[1]Later published as Miller et al. (2021) subsequent to the completion of this study.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analysis IPD study selection flowchart. IPD = individual participant data.



over 3 weeks. In the remaining two studies, ReST was conducted at a lower dose frequency of twice per week over 6 weeks. Except for Staples et al. (2009), all studies shared similar cumulative intervention intensity of 1,200 trials.

Table 3 summarizes the baseline characteristics of all participants and presents the proportions of data available for analyses. Data were pooled from 36 participants aged between 4 and 13 years, with a larger proportion of boys (n = 28) to girls (n = 8).

Table 2. Summary of included studies.

Study	Design	*n*	Mean age in months (range)	Intervention intensity					Mode of delivery
				Dose (trials)	Dose frequency per week	Total intervention duration	Session duration	Cumulative intervention intensity	
Staples et al. (2009)	QED	3	68 (52–78)	100	3–4 times	3 weeks	60 min	1,000 trials	FTF in clinic by student SLPs
McCabe et al. (2014)	SCED	4	81 (65–102)	100	4 times	3 weeks	60 min	1,200 trials	FTF in clinic by student SLPs
Thomas et al. (2014)	SCED	4	72 (56–96)	100	2 times	6 weeks	60 min	1,200 trials	FTF in clinic by student SLPs
Murray et al. (2015)	RCT	13	73 (48–142)	100	4 times	3 weeks	60 min	1,200 trials	FTF in clinic by student SLPs
Thomas et al. (2016)	SCED	5	97 (63–134)	100	4 times	3 weeks	60 min	1,200 trials	Video conferencing by SLPs and student SLPs
Thomas et al. (2018)	SCED	5	97 (61–139)	100	4 times	3 weeks	60 min	1,200 trials	FTF in clinic by student SLPs and at home by parents
McCabe, Preston, et al. (2020)	RCT	2	117 (77–156)	100	2 times	6 weeks	60 min	1,200 trials	FTF in clinic by student SLPs

Note. QED = quasi-experimental design; FTF = face-to-face; SLPs = speech-language pathologists; SCED = single-case experimental design; RCT = randomized controlled trial.

Change Scores and Effect Sizes

Table 4 presents descriptive statistics for change scores on treated pseudowords and untreated real words from baseline to 1 week posttreatment and from 1 week to 4 weeks posttreatment. Participants improved their performance on treated pseudowords by a mean of 25% (SD = 21) from baseline to 1 week posttreatment, $F(1, 36) = 50.69$, $p < .001$, with a very large effect size, $d = 1.05$. On average, they retained these treatment gains from 1 week to 4 weeks posttreatment, $F(1, 35) = 3.41$, $p = .073$, $d = 0.17$. Participants also improved their performance on untreated real words by a mean of 14% (SD = 13), $F(1, 32) = 38.94$, $p < .001$, with a moderately large effect size, $d = 0.59$. On average, they maintained these generalization gains from 1 week to 4 weeks posttreatment, $F(1, 31) = 0.74$, $p = .397$, $d = 0.06$.

Assessment of Study Quality

All studies were assessed to have high methodological quality (see Supplemental Material S3).

Research Question 1: Predicting Absolute Posttreatment Performance

Bivariate Correlations

The correlations between potential predictors and absolute performance 1 week posttreatment are presented in Table 5. Several speech and nonspeech variables were significantly correlated with absolute performance on treated pseudowords and untreated real words 1 week posttreatment. Unsurprisingly, performance on pseudowords 1 week posttreatment was most strongly correlated with performance on pseudowords at baseline. Similarly, performance on real words 1 week posttreatment was most strongly correlated with performance on real words at baseline. Following baseline performance, the next strongest correlations for performance on pseudowords and real words 1 week posttreatment were with PCC and CELF Expressive Language Index (ELI) scores.

Regression Analyses

Table 6 presents results of the multiple regression for predictors of absolute performance on treated pseudowords and untreated real words 1 week posttreatment. All four regressions were significant with very large R^2 values, given that an R^2 value of .26 was considered large. Several predictors that were significant in the correlations were not significant in the models, indicating that the variance they shared with the outcome variable was also shared with at least one other independent variable. For example, accuracy on untreated real words at baseline can independently predict performance on treated pseudowords 1 week posttreatment, as evidenced by its significant bivariate correlation in Table 5, but adds no predictive ability not provided by other independent variables in the multiple regression. Such shared variance is seen in the multiple regressions reported in Table 6.

Treated pseudowords. Among the nonspeech variables, only baseline ELI scores were strongly associated with absolute posttreatment performance on pseudowords in the model. Children with higher baseline ELI scores

Table 3. Baseline characteristics of included participants.

Characteristic	*n* (%)	*M* (*SD*; range)
Demographic		
Age in months	36 (100)	82 (29; 48–156)
Sex	36 (100)	28 boys, 8 girls
Oral motor skills		
Oral function score	36 (100)	90 (13; 60–109)
Speech		
GFTA-2 SS	32 (89)	65 (22; 39–101)
PCC	36 (100)	61 (22; 16–94)
PVC	36 (100)	70 (19; 39–94)
Language		
CELF Receptive Language Index[a]	35 (97)	97 (14; 70–132)
CELF Expressive Language Index[a]	34 (94)	82 (20; 45–116)
PPVT-4 SS	33 (92)	100 (14; 68–129)
Working memory		
Forward digit span SS[b]	32 (89)	7 (3; 1–13)
CAS-related speech feature		
% inconsistency	36 (100)	56 (18; 20–100)
% lexical stress errors	36 (100)	58 (29; 19–98)
Treatment outcome measures at baseline		
% correct treated pseudowords	36 (100)	16 (18; 0–60)
% correct untreated real words	32 (89)	30 (22; 0–85)

Note. GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000); SS = standard score; PCC = percentage of consonants correct; PVC = percentage of vowels correct; CELF = Clinical Evaluation of Language Fundamentals; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007); CAS = childhood apraxia of speech.

[a]Depending on age, scores were taken from the Clinical Evaluation of Language Fundamentals Preschool–Second Edition–Australian and New Zealand Standardised Edition (Wiig et al., 2006), Clinical Evaluation of Language Fundamentals–Third Edition (Semel et al., 1995), or Clinical Evaluation of Language Fundamentals–Fourth Edition–Australian Standardised Edition (Semel et al., 2006). [b]Depending on study, scores were taken from the Test of Auditory Processing Skills–Third Edition (Martin & Brownell, 2005), Comprehensive Test of Phonological Processing (Wagner et al., 1999), or Comprehensive Test of Phonological Processing–Second Edition (Wagner et al., 2013).

Table 4. Descriptive statistics for change scores.

Outcome	*n* (%)	*M* (*SD*; range)
% correct treated pseudowords		
Baseline to 1 week posttreatment	36 (100%)	25 (21; –11 to 65)
1 week to 4 weeks posttreatment	36 (100%)	5 (16; –28 to 57)
% correct untreated real words		
Baseline to 1 week posttreatment	32 (89%)	14 (13; –5 to 50)
1 week to 4 weeks posttreatment	32 (89%)	2 (11; –27 to 21)

performed better on pseudowords 1 week posttreatment. For speech variables, Goldman-Fristoe Test of Articulation (GFTA) Sounds-In-Words standard scores, PVC, speech inconsistency, and baseline performance on pseudowords were collectively associated with absolute posttreatment performance on pseudowords. Children who started therapy with better consonantal and poorer vowel accuracy, lower speech inconsistency, and higher accuracy on targets to be treated showed higher absolute posttreatment performance on pseudowords.

Untreated real words. No significant relationship remained between any nonspeech variable and absolute posttreatment performance on real words. Only baseline performance on real words remained strongly associated with absolute posttreatment performance on real words. Children with higher baseline performance on real words performed better on real words posttreatment.

Research Questions 2 and 3: Predicting Gain in Performance and Retention of Gains

Bivariate Correlations

Table 7 presents the correlations between potential predictor variables and change in performance on treated pseudowords and untreated real words. Only CELF ELI and GFTA scores correlated significantly with treatment gains, and these correlations were not strong. There were no significant correlations between any potential predictor variables and generalization gains, or retention of treatment or generalization gains.

Regression Analyses

None of the multiple regressions shown in Table 8 were significant, indicating that baseline measures provide at best only very weak prediction of treatment change from pre- to posttreatment in treated pseudowords and no ability to predict generalization change in untreated real words, or retention of change in either treated pseudowords or untreated real words.

Discussion

This IPD meta-analysis presents the first synthesis of data from previous studies to identify predictors of treatment outcomes in ReST. This study sought to identify predictors of (a) absolute posttreatment performance, (b) treatment and generalization gains, and (c) retention of gains on measures of treated pseudowords and untreated real words. Absolute posttreatment performance on treated pseudowords was predicted by higher baseline expressive language scores and a combination of lower PVC, lower speech inconsistency, higher GFTA scores, and higher accuracy on pseudowords. Absolute posttreatment performance on untreated real words

Table 5. Correlation matrix of variables and absolute performance 1 week posttreatment.

Potential predictors	Absolute performance on treated pseudowords			Absolute performance on untreated real words		
	Correlation	*p*	*N*	Correlation	*p*	*N*
Dose frequency	–.009	.96	36	–.212	.245	32
Nonspeech variables						
Age in months	.312	.064*	36	.338	.059*	32
Sex	.104	.545	36	.089	.627	32
Oral function score	.430	.009**	36	.341	.056*	32
CELF Receptive Language Index	.104	.552	35	.326	.074*	31
CELF Expressive Language Index	.434	.01**	34	.474	.008**	30
PPVT-4 SS	.152	.4	33	.216	.259	29
Forward digit span SS	.375	.034**	32	.304	.116*	28
Speech variables						
GFTA-2 SS	.394	.026**	32	.264	.145*	32
% inconsistency	–.381	.022**	36	–.436	.013**	32
PVC	.277	.102*	36	.264	.145*	32
PCC	.498	.002**	36	.534	.002**	32
% lexical stress errors	–.292	.084*	36	–.199	.274	32
% correct treated pseudowords at BL	.684	< .001**	36	.434	.013**	32
% correct untreated real words at BL	.416	.018**	32	.864	< .001**	32

Note. CELF = Clinical Evaluation of Language Fundamentals (Semel et al., 1995, 2006; Wiig et al., 2006); PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007); SS = standard score; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000); PVC = percentage of vowels correct; PCC = percentage of consonants correct; BL = baseline.

*$p < .2$. **$p < .05$.

Table 6. Summary of multiple regression for predicting absolute performance 1 week posttreatment.

Absolute performance on treated pseudowords				Absolute performance on untreated real words			
Variable	*b*	*SE*	*p*	Variable	*b*	*SE*	*p*
Nonspeech variables				Nonspeech variables			
Dose frequency	11.372	12.265	.363	Dose frequency	–11.329	12.617	.380
Forward digit span SS	–0.123	1.747	.944	Forward digit span SS	–1.480	–0.193	.369
Age in months	0.176	0.185	.352	Age in months	0.287	0.168	.104
Oral function score	0.767	0.404	.070	Oral function score	0.184	0.383	.635
CELF Expressive Language Index	0.689	0.276	.020*	CELF Expressive Language Index	0.507	0.331	.142
				CELF Receptive Language Index	0.494	0.375	.203
$F(5, 24) = 5.385$, $p = .002$	$R^2 = .529$			$F(6, 19) = 3.488$, $p = .017$	$R^2 = .524$		
Speech variables				Speech variables			
Dose frequency	12.669	10.054	.219	Dose frequency	–4.562	6.695	.502
GFTA-2 SS	0.477	0.183	.015*	GFTA-2 SS	0.056	0.122	.651
PVC	–0.690	0.280	.021*	PVC	–0.015	0.186	.936
% inconsistency	–0.688	0.297	.029*	% inconsistency	–0.209	0.198	.301
% correct treated pseudowords at BL	1.180	0.272	< .001*	% correct treated pseudowords at BL	–0.159	0.181	.388
% correct untreated real words at BL	–0.078	0.198	.698	% correct untreated real words at BL	0.967	0.132	< .001*
$F(6, 25) = 7.917$, $p < .001$	$R^2 = .655$			$F(6, 25) = 14.543$, $p < .001$	$R^2 = .777$		

Note. *b* = unstandardized regression coefficient; *SE* = standard error; SS = standard score; CELF = Clinical Evaluation of Language Fundamentals (Semel et al., 1995, 2006; Wiig et al., 2006); *F* = *F* distribution; R^2 = multiple correlation squared; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000); PVC = percentage of vowels correct; BL = baseline.

*$p < .05$.

Table 7. Correlation matrix of variables and treatment outcomes.

	Treatment gains and its retention						Generalization gains and its retention					
	Baseline to 1 week posttreatment			1 week to 4 weeks posttreatment			Baseline to 1 week posttreatment			1 week to 4 weeks posttreatment		
Potential predictors	**Correlation**	***p***	***N***	**Correlation**	***p***	***N***	**Correlation**	***p***	***N***	**Correlation**	***p***	***N***
Dose frequency	–.002	.992	36	–.045	.792	36	–.2	.274	32	–.029	.877	32
Nonspeech variables												
Age in months	–.057	.74	36	–.063	.715	36	–.102	.578	32	.298	.098*	32
Sex	.154	.37	36	–.129	.454	36	.05	.784	32	–.130	.477	32
Oral function score	.247	.147*	36	–.132	.441	36	.176	.335	32	.068	.712	32
CELF Receptive Language Index	.155	.375	35	.009	.960	35	.24	.193*	31	–.304	.097*	31
CELF Expressive Language Index	.353	.041**	34	–.167	.346	34	.272	.145*	30	–.028	.882	30
PPVT-4 SS	.12	.507	33	–.141	.435	33	.171	.375	29	–.110	.570	29
Forward digit span SS	.315	.079*	32	–.092	.615	32	–.027	.892	28	.2	.308	28
Speech variables												
GFTA-2 SS	.384	.03**	32	–.335	.061*	32	.115	.531	32	.063	.733	32
% inconsistency	–.282	.096*	36	–.019	.912	36	–.189	.301	32	–.115	.530	32
PVC	–.001	.998	36	–.129	.452	36	–.007	.968	32	.158	.387	32
PCC	.238	.162*	36	–.170	.323	36	.137	.453	32	.163	.372	32
% lexical stress errors	–.052	.765	36	.210	.219	36	.029	.874	32	–.134	.466	32
% correct treated pseudowords at BL	.096	.577	36	–.101	.559	36	–.129	.481	32	.172	.346	32
% correct untreated real words at BL	.103	.574	32	–.175	.338	32	–.054	.771	32	.091	.619	32

Note. CELF = Clinical Evaluation of Language Fundamentals (Semel et al., 1995, 2006; Wiig et al., 2006); PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007); SS = standard score; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000); PVC = percentage of vowels correct; PCC = percentage of consonants correct; BL = baseline.

*$p < .2$. **$p < .05$.

Table 8. Summary of multiple regression for predicting treatment and generalization gains and their retention.

Treatment gains and its retention							
Baseline to 1 week posttreatment				**1 week to 4 weeks posttreatment**			
Variable	***b***	***SE***	***p***	**Variable**	***b***	***SE***	***p***
Dose frequency	7.748	11.786	.519	Dose frequency	–2.864	6.651	.670
GFTA-2 SS	0.421	0.263	.127	GFTA-2 SS	–0.227	0.118	.064
Oral function score	–0.257	0.510	.620				
CELF Expressive Language Index	0.351	0.299	.256				
% inconsistency	–0.535	0.490	.288				
Forward digit span SS	–1.825	2.003	.374				
$F(6, 19) = 1.819$, $p = .149$	$R^2 = .365$			$F(2, 29) = 1.938$, $p = .162$	$R^2 = .118$		
Generalization gains and its retention							
Baseline to 1 week posttreatment				**1 week to 4 weeks posttreatment**			
Variable	***b***	***SE***	***p***	**Variable**	***b***	***SE***	***p***
Dose frequency	–1.436	7.395	.848	Dose frequency	.003	5.364	1.000
CELF Receptive Language Index	.154	0.209	.467	CELF Receptive Language Index	–.191	0.141	.186
CELF Expressive Language Index	.099	0.149	.515	Age in months	.086	0.063	.185
$F(3, 26) = 0.896$, $p = .456$	$R^2 = .094$			$F(3, 27) = 1.607$, $p = .211$	$R^2 = .152$		

Note. *b* = unstandardized regression coefficient; *SE* = standard error; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000); SS = standard score; CELF = Clinical Evaluation of Language Fundamentals (Semel et al., 1995, 2006; Wiig et al., 2006).

was predicted by higher accuracy on real words at baseline. Only CELF ELI and GFTA scores were found to be weakly related to treatment gains. None of the measures were related to generalization gains or retention of gains.

Previously, Thomas et al. (2014) suggested that higher dose frequency of ReST (i.e., 4 times per week) may offer a slight advantage over lower dose frequency (i.e., twice-weekly). However, when the effect of dose frequency was examined again in this study with the addition of two participants from McCabe, Preston, et al. (2020) to the lower dose frequency group, dose frequency was not strongly correlated with treatment outcomes and was not a significant predictor in the regression analyses. Based on our findings, it is possible that dose frequency at the levels included in the analysis is not related to treatment outcomes in ReST. Previous findings in Thomas et al. (2014) may have been an artifact of the small sample. Furthermore, as Thomas et al. (2014) was a single-case experimental design study, a direct comparison of dose frequencies was not possible at that time. It is also possible that variations in response observed then were washed out with the increased sample size here. Nevertheless, although the sample in the present meta-analysis was larger overall, there were only six cases in the lower dose-frequency condition. Further evidence would be beneficial.

Although studies of treatment intensity in the CAS population have suggested that a higher dose frequency favors better treatment outcomes, dose frequency effects have not been examined while keeping cumulative intervention intensity constant (e.g., Namasivayam et al., 2015). As suggested by Thomas et al. (2014), it is possible that, similar to findings in acquired motor speech disorders, dose frequency does not impact treatment efficacy when comparing frequencies of two or more sessions per week. Instead, differences in treatment outcomes may be noted with larger differences in dose frequencies. Further research involving wider variations in dose frequency (e.g., once vs. 4 times per week) while keeping cumulative intervention intensity constant is required to determine how dose frequency impacts treatment efficacy in ReST. Establishing the optimal intervention intensity will allow clinicians to make informed decisions to maximize the efficacy of treatment provided.

In this study, older children were aged between 7 and 13 years (last quartile) while younger children were aged between 4 and 5 years (first quartile). We expected that older children may have better treatment outcomes due to the general effect of age on cognition and motor learning. However, our results indicate that age was not significantly correlated with absolute posttreatment performance, gain in performance, or retention of gains. Neither was age found to be a significant predictor of these outcomes in the multiple regressions. Thus, contrary to previous suggestions that ReST would be more appropriate for use with older children (E. Maas et al., 2014), our findings suggest younger children are equally as appropriate. The youngest children included in the existing studies were 4 years old, and clinicians can be comfortable that if the children in their caseload are similar to those included in the studies, then 4-year-olds are likely to benefit from ReST therapy.

Apart from examining the effect of dose frequency and age on treatment outcomes, this study also identified predictors of treatment outcomes that have not yet been suggested in the literature. Firstly, higher baseline expressive language scores predicted better absolute posttreatment performance on treated pseudowords. Among the 34 participants whose CELF scores were available, 19 had co-occurring expressive and/or receptive language impairments. There are indications in the literature that children with CAS and comorbid language impairments have different speech and speech processing profiles to those without comorbid language impairments (Iuzzini-Seigel et al., 2017; Zuk et al., 2018). These different profiles may be related to differences in ReST treatment outcomes. However, it is also important to highlight that most children in this study had normal receptive language and receptive vocabulary due to the inclusion criteria in several original studies. Future research should examine the efficacy of ReST in a broader CAS population including those with co-occurring language impairments.

It is interesting to note that higher expressive language scores correlated with better performance on measures of consonant articulation (i.e., GFTA scores and PCC). These findings suggest that the ELI scores may be related to or are a function of the children's speech difficulties. It seems plausible then that better expressive language skills may be associated with better speech abilities, which predict better performance on treated pseudowords. This idea is supported by the finding that both higher GFTA scores and lower speech inconsistency also predicted better absolute posttreatment performance on pseudowords.

Secondly, our findings indicate that lower PVC at baseline was one of the few predictors of better absolute performance on treated pseudowords at 1 week posttreatment. However, it should be noted that PVC at baseline has positive but nonsignificant bivariate correlations with both treated pseudowords and untreated real words 1 week posttreatment. Its significant coefficient in the multiple regression thereafter indicates its role as a suppressor variable in the relationship between the speech predictors and absolute performance on pseudowords 1 week posttreatment. In other words, it suppresses the irrelevant variance of other predictors, thereby improving the model. This suggests that the role of PVC in absolute performance on pseudowords in ReST is complex and warrants specific study, possibly in conjunction with the other variables

noted above, such as dose frequency and pretreatment expressive language skills.

Lastly, accuracy on treated pseudowords at baseline predicted absolute performance on the same items 1 week posttreatment. It seems intuitive that a child who can accurately produce treatment targets pretreatment would perform well posttreatment. Similarly, accuracy on untreated real words at baseline predicted accuracy on the same items 1 week posttreatment.

Changes in mean scores of treated pseudowords and untreated real words were analyzed. While the change in performance on treated pseudowords and untreated real words from baseline to 1 week posttreatment was significant, there was no significant change in performance on these measures from 1 week to 4 weeks posttreatment, indicating maintenance of gains rather than ongoing improvement. This is contrary to findings in Murray et al. (2015), where participants not only maintained but also improved their performance on treated pseudowords up to 4 months posttreatment. Future research can explore if there are identifiable differences between children who retain their gains and those who continue to make further gains between 1 week and 4 weeks posttreatment.

When all the regression models are considered concurrently, the results indicate that the factors that predict absolute posttreatment performance do not predict gain in performance from baseline to 1 week posttreatment or retention of gains from 1 week to 4 weeks posttreatment. Participants were as likely to improve their performance and retain these improvements following ReST regardless of differences in individual participant characteristics such as age or baseline performance on measures of speech, language, pseudowords to be treated, or real words.

Strengths and Limitations

While the use of IPD meta-analysis is not common in speech-language pathology research, it is considered by some to be the gold standard of meta-analyses (Sutton & Higgins, 2008). The use of IPD facilitated more flexible and robust analyses and helped to overcome biases associated with aggregate data (Stewart et al., 2015). As a result of compiling the IPD across studies, an overall effect size was calculated for gain in performance of treated pseudowords and for generalization to real words. These effect sizes were very large (d = 1.05) and moderately large (d = 0.59), respectively. This meta-analysis can replace the d and d_2 scores reported in the previous ReST literature and can be seen as more representative due to the larger sample size (n = 36) collated across the seven included studies. However, this study was not without limitations.

Given that existing ReST literature mostly comprises of single-case experimental designs with small participant numbers, our study was limited by a relatively small sample size despite obtaining IPD from most studies. Moreover, participant numbers were further reduced by the need for independence of observations in regression analyses. While a multilevel model could potentially address the violation of this assumption, this study did not have a sufficiently large sample size to obtain reliable results from such analysis (C. J. M. Maas & Hox, 2004). The small sample size had several implications. Firstly, although IPD meta-analysis is recognized for increasing the power and precision of statistical analyses, the pool of available ReST studies is small enough that Type II errors may explain some findings, such as the nonsignificant gain in performance on treated pseudowords from 1 week to 4 weeks posttreatment (p = .073). Secondly, the small sample size and paucity of literature to guide variable selection made statistical selection criteria necessary to minimize the risk of overfitting (Hawkins, 2004). Selecting variables using bivariate correlations is not ideal as it only reflects the relationship between a single independent variable and the outcome measure without considering more complicated relationships between predictors themselves (Kraha et al., 2012). Lastly, while statistical criteria were used to ensure a sufficient sample size for investigating potential predictors, many variables were identified and had to be divided into speech and nonspeech variables for entry into the regression models. Consequently, the influence of these variables on absolute performance in ReST could not be considered concurrently.

Several potential sources of bias were present in this study. Although a systematic literature search identified all ReST treatment studies conducted to date, IPD from only seven of nine studies could be accessed and included. There is a risk of availability bias where our findings may differ if IPD from the unavailable studies were included. Additionally, publication bias may have been introduced in the inclusion of published rather than unpublished data for participants who were in two studies. However, data from unpublished studies were not completely excluded from the study. Two participants from McCabe, Preston, et al. (2020) and three participants from Staples et al. (2009) were included.

Clinical Implications

Based on our findings, the clinical presentation of children likely to end treatment with higher or lower absolute performance is shown in Table 9. Although the characteristics listed may indicate a child's likelihood of achieving higher absolute posttreatment performance, these factors are unable to account for treatment and generalization gains, or retention of gains attained by participants. In the sample studied here, younger children with varying levels of speech and language skills and performance on pseudowords and real words at baseline

Table 9. Likely clinical presentation of children with higher or lower absolute posttreatment performance.

Measure	Absolute performance on treated pseudowords		Absolute performance on untreated real words	
	Higher	Lower	Higher	Lower
Age	—	—	—	—
CELF Expressive Language Index	Higher scores (97–116)	Lower scores (45–69)	—	—
GFTA-2 SS	Higher scores (89–101)	Lower scores (89–101)	—	—
PVC	Lower (39%–51%)	Higher (88%–94%)	—	—
% inconsistency	Lower (20%–43%)	Higher (68%–100%)	—	—
% correct treated pseudowords at BL	Higher (26%–60%)	Lower (0%–3%)	—	—
% correct untreated real words at BL	—	—	Higher (43%–85%)	Lower (0%–15%)

Note. Higher performance indicates participants in this study who performed at the 75th percentile and above; lower performance indicates those who performed at the 25th percentile and below of values for each variable. Indicative scores from the population in this study are reported. Em dashes indicate no measure differentiated higher or lower performance. CELF = Clinical Evaluation of Language Fundamentals (Semel et al., 1995, 2006; Wiig et al., 2006); GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000); SS = standard score; PVC = percentage of vowels correct; BL = baseline.

benefited from ReST as did older children with differing performance on these measures.

Given the limitations of the current data, clinicians should use Table 9 as a guide for discussion with parents rather than as definitive indicators of ReST suitability. While the factors that influence the extent of improvement and maintenance of gains in ReST remain unknown, our findings provide support that ReST can benefit children with varying pretreatment characteristics, at least within the range of variation of participant characteristics indicated in this study. These findings have limited generalizability to the wider population of children with CAS beyond the subset studied here (i.e., aged between 4 and 13 years; normal oral structure, vision, and hearing; no co-occurring developmental or genetic diagnoses).

Future Directions

Large-scale prospective research is required to further examine moderators of treatment outcomes in ReST specifically and CAS treatment more generally. Future research should investigate the effects of a wider range of dose frequencies and co-occurring language impairments on treatment outcomes in ReST. The complex role of PVC in relation to absolute posttreatment performance on treated pseudowords should also be examined further in conjunction with the other variables studied here. Finally, a similar meta-analysis should be repeated as more ReST research data become available, to construct larger models that consider the joint effects of speech and nonspeech variables on absolute posttreatment performance in ReST.

Author Contributions

Wei Lin Ng: Data curation (Lead), Formal analysis (Equal), Project administration (Lead), Visualization (Lead), Writing – original draft (Lead), Writing – review & editing (Equal). **Patricia McCabe:** Conceptualization (Lead), Data curation (Supporting), Methodology (Supporting), Project administration (Supporting), Resources (Equal), Supervision (Lead), Writing – review & editing (Equal). **Rob Heard:** Data curation (Supporting), Formal analysis (Equal), Methodology (Supporting), Validation (Lead), Visualization (Supporting), Writing – review & editing (Equal). **Veronica Park:** Data curation (Supporting), Formal analysis (Supporting), Methodology (Supporting), Supervision (Supporting), Validation (Supporting), Writing – review & editing (Supporting). **Elizabeth Murray:** Data curation (Supporting), Methodology (Supporting), Resources (Equal), Writing – review & editing (Supporting). **Donna Thomas:** Conceptualization (Supporting), Data curation (Supporting), Methodology (Supporting), Resources (Equal), Supervision (Equal), Writing – review & editing (Equal).

Acknowledgments

We thank Kirrie Ballard for sharing her de-identified research data and Don Robin and Hilary Miller for their assistance with information about data in the Treatment for Establishing Motor Program Organization papers. We acknowledge Vanessa Wan Lin Tan for her assistance with the interrater reliability for the inclusion of studies. We also thank all the clinicians and children from previous studies, without whom this study would not be possible.

References

American Speech-Language-Hearing Association. (2007). *Childhood apraxia of speech* [Technical report]. https://www.asha.org/policy/tr2007-00278/

Ballard, K. J., Robin, D. A., McCabe, P., & McDonald, J. (2010). A treatment for dysprosody in childhood apraxia of

speech. *Journal of Speech, Language, and Hearing Research, 53*(5), 1227–1245. https://doi.org/10.1044/1092-4388(2010/09-0130)

Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health, 25*(5), 464–469. https://doi.org/10.1111/j.1467-842X.2001.tb00294.x

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Erlbaum. https://doi.org/10.4324/9780203771587

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Routledge. https://doi.org/10.4324/9780203774441

Cumley, G., & Swanson, S. (1999). Augmentative and alternative communication options for children with developmental apraxia of speech: Three case studies. *Augmentative and Alternative Communication, 15*(2), 110–125. https://doi.org/10.1080/07434619912331278615

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2020). Chapter 10: Analysing data and undertaking meta-analyses. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions*. Cochrane. https://training.cochrane.org/handbook/current/chapter-10

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4)*. Pearson.

Edeal, D. M., & Gildersleeve-Neumann, C. E. (2011). The importance of production frequency in therapy for childhood apraxia of speech. *American Journal of Speech-Language Pathology, 20*(2), 95–110. https://doi.org/10.1044/1058-0360(2011/09-0005)

Farquharson, K., Tambyraja, S. R., & Justice, L. M. (2019). Contributions to gain in speech sound production accuracy for children with speech sound disorders: Exploring child and therapy factors. *Language, Speech, and Hearing Services in Schools, 51*(2), 457–468. https://doi.org/10.1044/2019_LSHSS-19-00079

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2)*. AGS.

Gozzard, H., Baker, E., & McCabe, P. (2006). Children's production of polysyllabic words. *ACQuiring Knowledge in Speech, Language and Hearing, 8*(3), 113–116.

Green, S. B. (1991). How many subjects does it take to do a regression analysis. *Multivariate Behavioral Research, 26*(3), 499–510. https://doi.org/10.1207/s15327906mbr2603_7

Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences, 44*(1), 1–12. https://doi.org/10.1021/ci0342472

Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley. https://doi.org/10.1002/9781118548387

Iuzzini-Seigel, J., Hogan, T. P., & Green, J. R. (2017). Speech inconsistency in children with childhood apraxia of speech, language impairment, and speech delay: Depends on the stimuli. *Journal of Speech, Language, and Hearing Research, 60*(5), 1194–1210. https://doi.org/10.1044/2016_JSLHR-S-15-0184

Iversen, I. H. (2013). Single-case research methods: An overview. In *APA handbook of behavior analysis, Vol. 1: Methods and principles* (pp. 3–32). American Psychological Association.

Kraha, A., Turner, H., Nimon, K., Zientek, L. R., & Henson, R. K. (2012). Tools to support interpreting multiple regression in the face of multicollinearity. *Frontiers in Psychology, 3,* 44. https://doi.org/10.3389/fpsyg.2012.00044

Lewis, B. A., Benchek, P., Tag, J., Miller, G., Freebairn, L., Taylor, H. G., Iyengar, S. K., & Stein, C. M. (2021). Psychosocial comorbidities in adolescents with histories of childhood apraxia of speech. *American Journal of Speech-Language Pathology, 30*(6), 2572–2588. https://doi.org/10.1044/2021_AJSLP-21-00035

Maas, C. J. M., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58*(2), 127–137. https://doi.org/10.1046/j.0039-0402.2003.00252.x

Maas, E., & Farinella, K. A. (2012). Random versus blocked practice in treatment for childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 55*(2), 561–578. https://doi.org/10.1044/1092-4388(2011/11-0120)

Maas, E., Gildersleeve-Neumann, C. E., Jakielski, K. J., & Stoeckel, R. (2014). Motor-based intervention protocols in treatment of childhood apraxia of speech (CAS). *Current Developmental Disorders Reports, 1*(3), 197–206. https://doi.org/10.1007/s40474-014-0016-4

Maas, E., Robin, D. A., Hula, S. N. A., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17*(3), 277–298. https://doi.org/10.1044/1058-0360(2008/025)

Martin, N. A., & Brownell, R. (2005). *Test of Auditory Processing Skills–Third Edition*. Academic Therapy Publications.

McCabe, P., Macdonald-D'Silva, A. G., van Rees, L. J., Ballard, K. J., & Arciuli, J. (2014). Orthographically sensitive treatment for dysprosody in children with childhood apraxia of speech using ReST intervention. *Developmental Neurorehabilitation, 17*(2), 137–145. https://doi.org/10.3109/17518423.2014.906002

McCabe, P., Preston, J. L., Evans, P., & Heard, R. (2020). *A pilot randomized control trial of motor-based treatments for childhood apraxia of speech: ReST and ultrasound biofeedback* [Manuscript submitted for publication]. Discipline of Speech Pathology, University of Sydney.

McCabe, P., Thomas, D. C., & Murray, E. (2020). Rapid syllable transition treatment—A treatment for childhood apraxia of speech and other pediatric motor speech disorders. *Perspectives of the ASHA Special Interest Groups, 5*(4), 821–830. https://doi.org/10.1044/2020_PERSP-19-00165

McNeill, B. C., Gillon, G. T., & Dodd, B. (2009). Effectiveness of an integrated phonological awareness approach for children with childhood apraxia of speech (CAS). *Child Language Teaching and Therapy, 25*(3), 341–366. https://doi.org/10.1177/0265659009339823

Miller, H. (2018). *Improvements in speech of children with apraxia: The efficacy of a Treatment for Establishing Motor Program Organization (TEMPO)* [Master's thesis, University of New Hampshire]. University of New Hampshire Scholar Repository. https://scholars.unh.edu/thesis/1190/

Miller, H. E., Ballard, K. J., Campbell, J., Smith, M., Plante, A. S., Aytur, S. A., & Robin, D. A. (2021). Improvements in speech of children with apraxia: The efficacy of Treatment for Establishing Motor Program Organization (TEMPO^{SM}). *Developmental Neurorehabilitation, 24*(7), 494–509. https://doi.org/10.1080/17518423.2021.1916113

Morgan, A. T., Murray, E., & Liegeois, F. J. (2018). Interventions for childhood apraxia of speech. *Cochrane Database of Systematic Reviews, 2018*(5), CD006278. https://doi.org/10.1002/14651858.CD006278.pub3

Morgan, A. T., & Webster, R. (2018). Aetiology of childhood apraxia of speech: A clinical practice update for paediatricians. *Journal of Paediatrics and Child Health, 54*(10), 1090–1095. https://doi.org/10.1111/jpc.14150

Murray, E., McCabe, P., & Ballard, K. J. (2014). A systematic review of treatment outcomes for children with childhood apraxia of speech. *American Journal of Speech-Language Pathology, 23*(3), 486–504. https://doi.org/10.1044/2014_AJSLP-13-0035

Murray, E., McCabe, P., & Ballard, K. J. (2015). A randomized controlled trial for children with childhood apraxia of speech comparing Rapid Syllable Transition Treatment and the Nuffield Dyspraxia Programme–Third Edition. *Journal of Speech, Language, and Hearing Research, 58*(3), 669–686. https://doi.org/10.1044/2015_JSLHR-S-13-0179

Namasivayam, A. K., Pukonen, M., Goshulak, D., Hard, J., Rudzicz, F., Rietveld, T., Maassen, B., Kroll, R., & van Lieshout, P. (2015). Treatment intensity and childhood apraxia of speech. *International Journal of Language & Communication Disorders, 50*(4), 529–546. https://doi.org/10.1111/1460-6984.12154

Oh, D., & Ha, J. (2021). Development and clinical application of Korean-version nonword intervention to improve speech motor programming. *Phonetics and Speech Sciences, 13*(2), 77–90. https://doi.org/10.13064/KSSS.2021.13.2.077

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *BMJ, 340,* 521–525. https://doi.org/10.1136/bmj.c221

Robbins, J., & Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders, 52*(3), 271–277. https://doi.org/10.1044/jshd.5203.271

Scarcella, I., Michelazzo, L., & McCabe, P. (2021). A pilot single-case experimental design study of Rapid Syllable Transition Treatment for Italian children with childhood apraxia of speech. *American Journal of Speech-Language Pathology, 30*(3S), 1496–1510. https://doi.org/10.1044/2021_AJSLP-20-00133

Schultz, E. A. (2018). *Communicative participation improves following motor speech program treatment in apraxia of speech* [Master's thesis, University of New Hampshire]. University of New Hampshire Scholars Repository. https://scholars.unh.edu/thesis/1185/

Semel, E., Wiig, E. H., & Secord, W. A. (1995). *Clinical Evaluation of Language Fundamentals–Third Edition*. The Psychological Corporation.

Semel, E., Wiig, E. H., & Secord, W. A. (2006). *Clinical Evaluation of Language Fundamentals, Australian Standardised Edition–Fourth Edition*. Pearson.

Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics, 8*(2), 39–42. https://doi.org/10.12691/ajams-8-2-1

Shriberg, L. D., Kwiatkowski, J., & Mabie, H. L. (2019). Estimates of the prevalence of motor speech disorders in children with idiopathic speech delay. *Clinical Linguistics & Phonetics, 33*(8), 679–706. https://doi.org/10.1080/02699206.2019.1595731

Staples, T., McCabe, P., McDonald, J., & Ballard, K. J. (2009). *A polysyllabic non-word treatment for childhood apraxia of speech incorporating key principles of motor learning* [Undergraduate honors thesis, University of Sydney]. ResearchGate. https://www.researchgate.net/publication/284082506_A_Polysyllabic_Non-Word_Treatment_for_Childhood_Apraxia_of_Speech_Incorporating_Key_Principles_of_Motor_Learning

Stein, C. M., Benchek, P., Miller, G., Hall, N. B., Menon, D., Freebairn, L., Tag, J., Vick, J., Taylor, H. G., Lewis, B. A., & Iyengar, S. K. (2020). Feature-driven classification reveals potential comorbid subtypes within childhood apraxia of speech. *BMC Pediatrics, 20*(1), 519–511. https://doi.org/10.1186/s12887-020-02421-1

Stewart, L. A., Clarke, M., Rovers, M., Riley, R. D., Simmonds, M., Stewart, G., & Tierney, J. F. (2015). Preferred Reporting Items for a Systematic Review and Meta-analysis of individual participant data. *JAMA, 313*(16), 1657–1665. https://doi.org/10.1001/jama.2015.3656

Strand, E. A. (2020). Dynamic temporal and tactile cueing: A treatment strategy for childhood apraxia of speech. *American Journal of Speech-Language Pathology, 29*(1), 30–48. https://doi.org/10.1044/2019_AJSLP-19-0005

Sutton, A. J., & Higgins, J. P. T. (2008). Recent developments in meta-analysis. *Statistics in Medicine, 27*(5), 625–650. https://doi.org/10.1002/sim.2934

Thomas, D. C., McCabe, P., & Ballard, K. J. (2014). Rapid Syllable Transitions (ReST) treatment for childhood apraxia of speech: The effect of lower dose-frequency. *Journal of Communication Disorders, 51,* 29–42. https://doi.org/10.1016/j.jcomdis.2014.06.004

Thomas, D. C., McCabe, P., & Ballard, K. J. (2018). Combined clinician–parent delivery of Rapid Syllable Transition (ReST) treatment for childhood apraxia of speech. *International Journal of Speech-Language Pathology, 20*(7), 683–698. https://doi.org/10.1080/17549507.2017.1316423

Thomas, D. C., McCabe, P., Ballard, K. J., & Lincoln, M. (2016). Telehealth delivery of Rapid Syllable Transitions (ReST) treatment for childhood apraxia of speech. *International Journal of Language & Communication Disorders, 51*(6), 654–671. https://doi.org/10.1111/1460-6984.12238

Tyler, A. A., Lewis, K. E., & Welch, C. M. (2003). Predictors of phonological change following intervention. *American Journal of Speech-Language Pathology, 12*(3), 289–298. https://doi.org/10.1044/1058-0360(2003/075)

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing*. Pro-Ed.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing–Second Edition*. Pro-Ed.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews, 13*(1), 70–77. https://doi.org/10.1002/mrdd.20139

Welkowitz, J., Cohen, B. H., & Lea, R. B. (2012). *Introductory statistics for the behavioral sciences* (7th ed.). Wiley.

Wiig, E. H., Secord, W. A., & Semel, E. (2006). *Clinical Evaluation of Language Fundamentals Preschool, Australian and New Zealand Standardised Edition–Second Edition*. Pearson.

Zuk, J., Iuzzini-Seigel, J., Cabbage, K., Green, J. R., & Hogan, T. P. (2018). Poor speech perception is not a core deficit of childhood apraxia of speech: Preliminary findings. *Journal of Speech, Language, and Hearing Research, 61*(3), 583–592. https://doi.org/10.1044/2017_JSLHR-S-16-0106