AJSLP

Review Article

Speech Prosody Interventions for Persons With Autism Spectrum Disorders: A Systematic Review

Sarai Holbrook[a] and Megan Israelsen[a]

Purpose: Persons with autism spectrum disorder (ASD) may demonstrate abnormal prosodic patterns in conversational speech, which can negatively affect social interactions. The purpose of this systematic review was to identify interventions measuring the improvement of expressive speech prosody in persons with ASD in order to support clinician's evidence-based decision making.
Method: We used 13 electronic databases to search for relevant articles using terms related to autism, intervention, and speech prosody. The databases identified a total of nine articles for the title, abstract, and full-text reviews. Five more articles were included after performing descendant and reference searches. One peer-reviewed article was excluded due to insufficient data received from the authors. We coded the resulting 13 articles for report, setting, intervention, outcome, and results characteristics and methodological quality.
Results: Results showed that interventions specifically targeting speech prosody using established and emerging evidence-based practices across more than 1 treatment day resulted in moderate to large improvements in speech prosody in persons with ASD. Interventions that indirectly targeted prosody or were very short resulted in small or nonsignificant effects.
Discussion: The results of this literature review suggest that interventions that directly target speech prosody using established evidence-based practices for ASD may be most effective for increasing typical prosodic patterns during speech for persons with ASD. Further research is needed to establish which interventions are most effective for each age range and context.
Supplemental Material: https://doi.org/10.23641/asha.12735926

Atypical speech prosody has long been a clinical marker for autism spectrum disorder (ASD; Asperger, 1991; Fusaroli et al., 2017; Kanner, 1943; McCann & Peppé, 2003). Speech prosody is defined as a conglomerate of characteristics that exist above the level of words, phrases, and sentences in connected speech (Stevens et al., 1983; Szczepek Reed, 2011). These characteristics may include rate, pitch or intonation, stress, pauses, intensity, and duration. Atypical prosody can adversely affect interactional partners' perceptions of and reactions to each other and thus may have far-reaching effects academically, socially, and vocationally (Gordon et al., 2019; Szczepek Reed, 2011; Wiklund, 2016; Wynn et al., 2018).

An Immediate Impression

Atypical speech prosody has been described as a "bellwether" sign of ASD and may be one of the first things to set persons with ASD apart from peers with neurotypical development (NTD; Bone et al., 2015; de Marchena & Miller, 2017; Diel & Berkovits, 2010; Mesibov, 1992; Paul et al., 2005). Because a deficit in social communication is one of the core features of ASD (American Psychiatric Association, 2013), atypical prosody could exacerbate an already significant problem. Even slight or infrequent prosodic differences can give the feeling of oddness or unattractiveness (Shriberg et al., 2001). This is consistent with what de Marchena and Miller (2017) found in their survey of 146 diagnosing clinicians of persons with ASD. These authors examined the behaviors that contributed to "frank" (rapid or obvious) impressions of ASD. Unusual prosody was one of the seven most reported symptoms associated with the "frank" presentation of ASD, along with poor

[a]Department of Special Education and Rehabilitation, Utah State University, Logan

Correspondence to Sarai Holbrook, who is now at the School of Health Sciences and Wellness-Communication Sciences and Disorders, College of Professional Studies, University of Wisconsin–Stevens Point: sholbroo@uwsp.edu

Editor-in-Chief: Julie Barkmeier-Kraemer
Editor: Erinn H. Finke

Received September 27, 2019
Revision received April 18, 2020
Accepted June 3, 2020
https://doi.org/10.1044/2020_AJSLP-19-00127

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

eye contact, repetitive language and behaviors, odd body mannerisms, reduced social initiations, and atypical social engagement. Although unusual prosody may not be present in the speech of all persons with ASD (Fusaroli et al., 2017; Paul et al., 2005; J. Q. Simmons & Baltaxe, 1975), it has been found to contribute to perceptions of disorder in a significant number of individuals with ASD who do exhibit atypical prosody.

Furthering this idea, Redford et al. (2018) assessed inexperienced listeners' impressions of short, multiword utterances taken from 17 children with ASD and atypical prosody and 17 children with NTD. The listeners consistently designated the speech of children with ASD as "disordered" when compared to the ratings of the group with NTD. A second group of inexperienced listeners rated the children with ASD as less likeable than the children with NTD. A third group of inexperienced listeners rated speech characteristics, such as monotony and fluency, of the children with ASD. The authors examined the relations between the ratings of this third group and the previous two. They found that the ratings of presence or absence of disorder and ratings of likeability were most strongly associated with ratings of monotony and poorer intelligibility. These findings suggest that interaction partners can make negative social judgments about persons with ASD based on a short amount of time listening.

Social Ramifications for a "Substantial Minority"

While not all persons with ASD present with atypical prosody (Fusaroli et al., 2017; Paul et al., 2005; J. Q. Simmons & Baltaxe, 1975), for many persons with ASD, speech prosody differences set them apart from their peers with NTD. J. Q. Simons and Baltaxe (1975) found that four of the seven verbal adolescents with ASD in their sample (57%) demonstrated atypicalities in the suprasegmental aspects of their speech. Similarly, Shriberg et al. (2001) reported that 47% of the participants in their sample of 30 male adolescents and young adults with high-functioning autism or Asperger syndrome scored significantly differently from their peers with typical development on measures of phrasing, stress, loudness, pitch, and resonance. More recently, in a study examining the acoustic and perceptual prosodic characteristics of 11 upper elementary–age children with ASD with IQ scores in the typical range, Dahlgren et al. (2018) found the children's prosodic characteristics differed significantly only in number of words per utterance. However, of interest to this discussion, three of the 11 children (27%) with ASD in the Dahlgren study were correctly judged by speech-language pahtologist raters as having ASD. These three children had significantly poorer scores on a measure of narrative proficiency and were significantly different in their speech rate and fluency when compared with children who were not rated having ASD. It would appear that some characteristics of these children's speech prosody, coupled with narrative proficiency, combined to give the impression of ASD for these expert raters. In support of this idea, de Marchena and Miller (2017) found that experienced diagnosticians of persons with ASD reported that an average of 40% of their patients with ASD presented with "frank" ASD, a construct representing a conglomerate of characteristics that give an immediate impression of ASD. Atypical prosody is one of the characteristics included in the construct of "frank" ASD. Thus, while it is true that atypical prosody is not a universal marker of ASD, for some individuals with ASD, atypical speech prosody is a characteristic that sets them apart from their peers with NTD. This conclusion has not been fully verified with and thus may not be directly generalized to individuals requiring moderate to maximal support because the research just discussed centered on individuals who would likely have a severity level of 1, "requiring support," under the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*; American Psychiatric Association, 2013). Even so, for those individuals with ASD for whom prosody has been shown to be different from individuals with NTD, atypical speech prosody may have undesirable repercussions.

Differences in speech prosody of individuals with ASD can be associated with social functioning and others' perception of them (Fusaroli et al., 2017; Grossman et al., 2013; Jaramillo, 2018; Paul et al., 2005). Paul et al. (2005) examined the associations between the prosodic functioning of persons with ASD and ratings of their communication and socialization abilities. Thirty adolescent and adult males with ASD and atypical speech prosody were assessed using The Prosody-Voice Screening Profile (Shriberg et al., 1990). The Prosody-Voice Screening Profile scores were then compared to the participants' Vineland Adaptive Behavior Scales (Sparrow, et al., 1984) standard scores on socialization and communication and Autism Diagnostic Observation Schedule–Generic (ADOS-G; Lord et al., 2000) raw scores on socialization and communication. Atypical resonance predicted poorer Vineland Adaptive Behavior Scales socialization scores and ADOS-G communication scores. Atypicality in sentential stress demonstrated a small, but significant correlation with and trended toward significance as a predictor for ADOS-G communication ratings. These results suggest that prosody can have a nuanced but important effect on the social communication of persons with ASD.

Grossman et al. (2013) also investigated the influence of the speech prosody of persons with ASD on others' perceptions of them. They examined the emotionally laden narrative retells of 18 children and adolescents with ASD and 11 age-matched peers with NTD. Participants watched a short recording of an actor telling a series of stories with four distinct emotions in each story. They then retold the stories while being video-recorded. Trained, blinded raters coded the resulting retells for emotions displayed, emotional intensity, and how natural or awkward the participants appeared. Acoustic analyses of the stories included pitch and intensity parameters. Perceptually, the trained raters scored the children and adolescents with ASD as more expressive than their peers with NTD with a trend toward being more awkward. Acoustically, children and adolescents with ASD exhibited higher maximum pitches and wider

ranges in pitch and intensity than their peers with NTD. The ratings of awkwardness were moderately and significantly correlated with ADOS (Lord et al., 1999) social communication ratings, suggesting that the children who were rated as more awkward had more severe social communication impairments. These findings are consistent with the findings of Paul et al. (2005), who called for the development of appropriate interventions for the "substantial minority" (p. 867) of persons with ASD who also had prosodic difficulties because of the potential stigmatizing effect of even subtle differences.

This idea is supported by the ability of machine-learning technologies to accurately differentiate between persons with ASD and NTD based on acoustic parameters of their vocal productions. Machine-learning technologies in 15 studies distinguished persons with ASD from persons with NTD 70%–96% of the time based on their acoustic voice characteristics (Fusaroli et al., 2017). Likewise, Jaramillo (2018) found that acoustic analyses of toddlers' vocalizations by machine could reliably distinguish between children with ASD and children with NTD with 80% accuracy, suggesting that differences in the suprasegmental characteristics of vocalizations in persons with ASD can set them apart from a very early age. Although researchers still struggle to identify what features of speech prosody best characterize the speech of persons with ASD, the results of both machine-learning studies suggest that the overall speech prosody of many persons with ASD distinguishes them from their peers with NTD. This distinction can have negative social ramifications (Grossman et al., 2013; Paul et al., 2005; Redford et al., 2018; Shriberg et al., 2001) and is thus a practical concern for interventionists who work with persons with ASD who also exhibit prosodic differences (Nadig & Shaw, 2012; Paul et al., 2005).

The Need for Evidence-Based Prosody Intervention

Three large and high-quality literature reviews of the research on interventions for persons with ASD have been conducted in the last 11 years (National Autism Center, 2009, 2015; Will et al., 2018; Wong et al., 2013). The National Standards Project (NSP) of the National Autism Center published their first review of the evidence for interventions for persons with ASD up to age of 22 years in 2009, and then they published an updated and expanded search in 2015. Combined, these reviews cover the peer-reviewed literature on behavioral interventions for persons with ASD across the life span that may be implemented "in or by school systems, or early intervention, home-, hospital-, vocational- and/or community-based programs or in clinical settings" (National Autism Center, 2009, p. 16). These combined reviews identified 15 interventions with enough evidence to be considered established, 23 interventions that were emerging, and 15 that were unestablished.

Concurrent with the second NSP review, Wong et al. (2013) conducted a systematic review of the literature on general intervention practices in ASD and provided a list of 27 practices that had sufficient support to be evidence-based and 24 interventions that had some evidence but not enough to be considered evidence-based. The classification of "established" in the NSP reviews and "evidence-based" in the Wong et al. (2013) review were similar, with some minor differences such as the requirement that evidence be published by multiple research groups in the Wong et al. (2013) review. We have used both reviews to guide our analysis and interpretation of the results of our review because together they represent the most comprehensive evidence for interventions in ASD.

The purpose of the current review was to summarize the extant literature on speech prosody interventions for persons with ASD and connect it with the evidence-based practices presented by Wong et al. (2013) and the NSP reviews (National Autism Center, 2009, 2015). We thus could provide clinically relevant guidance to speech-language pathologists seeking to implement evidence-based practices in treatment for persons with ASD who also have speech prosody deficits. To further this purpose, we asked the following question framed using a problem/population, intervention, comparison, and outcome, or PICO framework (Huang et al., 2006): Do the speech prosody patterns of persons with ASD improve (i.e., become more typical sounding) after intervention?

Method

Information Sources

We conducted this systematic review using elements of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework outlined by Liberati et al. (2009) and following recommendations made by Cooper (2017). A preliminary title search was conducted in five electronic databases (CINAHL, Education Source, ERIC, MEDLINE, and PsycINFO) using the terms "autis*," "prosody OR intonation," and "intervention OR therapy OR treatment" connected with the Boolean operator "AND" to determine relevant keywords. Keywords were drawn from relevant articles found in this search (i.e., Bellon-Harn et al., 2007; Ferdosi, et al., 2013; Matsuda & Yamamoto, 2013; Miller & Toca, 1979). These key words were used with the thesaurus/subjects/heading function in selected databases to develop our final search terms (see Supplemental Material S1). The final search terms were entered into 13 electronic databases (see Supplemental Material S1) to find relevant articles. Google Scholar was used for descendant searches, and we hand-searched the reference lists of the finalized reports and two review articles (i.e., Fusaroli et al., 2017; McCann & Peppé, 2003) to identify other relevant articles. The references and descendants of any relevant reports found in these searches were reviewed until no more relevant articles emerged. We completed the original searches between January 2016 and December 2017 and performed an updated search in June of 2019.

Inclusion and Exclusion Criteria

The following inclusion and exclusion criteria were used to identify studies for this review. Studies needed to

- be reported in English;
- be reported in a peer-reviewed journal or as a doctoral dissertation;
- include an experimental evaluation of a specific intervention or set of intervention procedures (i.e., intervention package) as the independent variable;
- evaluate a behavioral (not pharmacological) intervention;
- have more than 50% of participants who have a primary diagnosis of ASD;
- have at least one dependent variable measuring one or more aspects of speech prosody production (i.e., rate, pitch, stress/emphasis, pauses, intensity, and duration);
- use an experimental group design, a quasi-experimental group design, or an experimental single-subject design (SSD).

Study Information Gathering

For each study, we created an Excel sheet with all coding questions and explanations loaded into it (see Supplemental Material S2). We coded studies using questions in eight areas of interest following recommendations from Cooper (2017). These areas included coder and coding process, report, setting, participant, intervention, and outcome characteristics, as well as results and quality indicators. A copy of the coding sheet used may be found in Supplemental Material S1. Coder and coding process characteristics included the reviewer's initials and the date coding was completed. Report characteristics comprised the study's citation, publication type, and study design. Setting characteristics detailed the context in which the study was conducted, including the country in which data were collected and where the intervention was delivered (home, clinic, etc.). Participant characteristics included sample size, age, sex, IQ, verbal ability, primary language, race/ethnicity, family history of ASD, socio-economic status, ASD diagnosis source, ASD diagnosis instrument, comorbidities, and participant intervention history. Information collected on intervention characteristics covered determinations of whether speech prosody was the primary goal of the intervention; number, lengths, and frequency of sessions; the duration of the intervention; whether intervention was conducted individually or in a group; the name of the intervention, if available; and a detailed description of the intervention copied and pasted from the article into the coding sheet. Outcome characteristics included if the prosody outcome was measured acoustically or behaviorally, when the prosody outcome was measured, a detailed description of the prosody outcome copied and pasted from the article, inter/intrarater reliability percentages, and fidelity measurements. The results section of the coding sheet involved the exact results (copied and pasted where practical), reported effect sizes (Tau-U, percentage of nonoverlapping data, Cohen's *d,* or other), any missing data information, claims of causality of the intervention on the outcome, if the cause preceded the outcome, if the cause covaried with the outcome, and if there were any other explanations for the outcome (see Shadish et al., 2002).

Quality of Evidence

Methodological quality was rated using a slightly modified version of the system proposed by Reichow et al. (Reichow, 2011; Reichow et al., 2008) for evaluating evidence-based practices in autism. Articles were classified as "strong," "adequate," or "weak" based on ratings of primary and secondary quality indicators. According to Reichow (2011), primary quality indicators are those elements of a study that are essential to establishing the validity of a study, while secondary quality indicators are important but not necessary for this purpose. Primary quality indicators were rated as high-quality, acceptable quality, or unacceptable quality. Primary quality indicators for group studies included information on six areas:

1. participant characteristics,
2. the independent variable,
3. the comparison condition,
4. the dependent variable,
5. the link between research questions and data analysis, and
6. statistical analysis.

Secondary indicators for group studies were rated as either present or absent and included eight areas, namely, random assignment, interobserver agreement, blind raters, fidelity, attrition, generalization or maintenance, effect size, and social validity.

Primary quality indicators for SSD studies were similar, but not identical, to those of group studies. They included information onSecondary quality indicators for SSD studies were as follows: interobserver agreement, kappa, blind raters, fidelity, generalization or maintenance, and social validity.

1. participant characteristics,
2. the independent variable,
3. the baseline condition,
4. the dependent variable or outcome measure,
5. visual analysis, and
6. experimental control.

Group research was deemed "strong" if it "received high-quality ratings on all primary quality indicators and showed evidence of four or more secondary quality indicators" (Reichow et al., 2008, p. 1314). "Adequate" group research showed evidence of four primary indicators with no primary qualities rated unacceptable and contained at least two secondary qualities. "Weak" group research was characterized by the presence of less than four primary quality indicators and/or less than two secondary quality indicators. Single-subject ratings of strong, adequate, and

weak were the same, with the exception that three (not four) or more secondary quality indicators were required for a "strong" rating.

The modification of the system for our review altered the operational definition of "acceptable" quality for participant characteristics. Reported participant characteristics were rated in the following four key areas, including (1) age and gender for participants, (2) participant diagnosis instruments and/or operational definitions for participants' symptoms and behaviors, (3) interventionist characteristics and information on secondary participants (if applicable), and (4) means for getting standardized test scores. In the Reichow system, to receive an "adequate" rating on the participants section, studies had to meet Criteria 1, 3, and 4. In our modification of the system, we required that studies meet any three of the four criteria. This modification was made a priori and ultimately did not change the results.

Strength of Evidence

Cohen's *d* or Hedge's *g* were used to summarize the strength of the evidence for group design studies, depending on whether groups being compared had equal or unequal sample sizes. If the sample sizes were unequal, we used Hedge's *g*. To interpret these statistics, tradition interpretations were deemed appropriate, that is, an effect of 0.2 or less was small, 0.5 was medium, and 0.8 or greater was large, to promote comparison with other studies. If group effect sizes were not reported, means, standard deviations, and sample sizes were entered into an online effect size calculator that produced calculations for both Cohen's *d* and Hedge's *g* (Stangroom, 2018, 2020).

Effects of SSD studies were summarized using Tau-U (Parker et al., 2011; Vannest et al., 2016). If Tau-U was not reported, we manually extracted data from published plots using an online plot digitizing software (Rohatgi, 2018). Next, these data points were place in an online SSD effect size calculator designed specifically for SSD effect size calculation (Vannest et al., 2016). As recommended by the authors, a Tau-U of < .65 signified a weak effect, a Tau-U of .65–.92 indicated a moderate to high effect, and a Tau-U of .93 or greater reflected a strong effect (Rakap, 2015). If the desired effect was a decrease in a behavior, negative Tau-U numbers resulted. The interpretations were similar, but in the opposite direction.

Interrater Reliability

Both authors independently performed a full-text review of studies identified in the initial title and abstract search. Interrater reliability for inclusion of studies was 92.6% (25/27 articles). The second author independently double-coded and extracted data for effect size calculation in three (23%) of the 13 included articles after meeting with the first author to discuss definitions on the coding sheet. Interrater reliability for coding was 93.3% on all categories, except effect size extraction and calculation. Interrater reliability for these calculations was 94.6%. All discrepancies were resolved through consensus.

Results

The results of the search of 13 electronic databases are displayed in Figure 1. Seven articles met the selection criteria after initial searching. Descendant and ancestral searching produced five more articles. Two of the 12 articles required more detailed information regarding prosody outcomes (e.g., Argott et al., 2017; Lim, 2010). Additional data were only received from the authors of Argott et al. (2017); therefore, a total of 11 articles were initially included. When the search was replicated in June 2019, two more articles met our inclusion criteria. One of these articles required further information that was requested and received (Parsons et al., 2018). In all, 13 records were included in our review that represented 16 separate analyses.

Study characteristics for all 13 studies can be found in Tables 1 and 2. Results are presented in terms of study design, participant characteristics, independent variables, dependent variables, strength of evidence, and quality of evidence. Our discussion of the independent variables (i.e., the interventions) is organized in terms of the NSP categories of established and emerging interventions. Dependent variables are organized according to how prosody was measured.

Study Design

The literature search yielded 10 studies that utilized SSDs (Argott et al., 2017; Boyd, 2018; Charlop et al., 2010; Daou et al., 2014; Edgerton & Wine, 2017; B. L. Koegel, 2014; R. L. Koegel & Frea, 1993; Nordgren, 2016; Ormand, 2016; Ozdemir, 2008) and three studies that employed group designs (Mayo, 2015; Parsons, et al., 2018; E. S. Simmons et al., 2016).

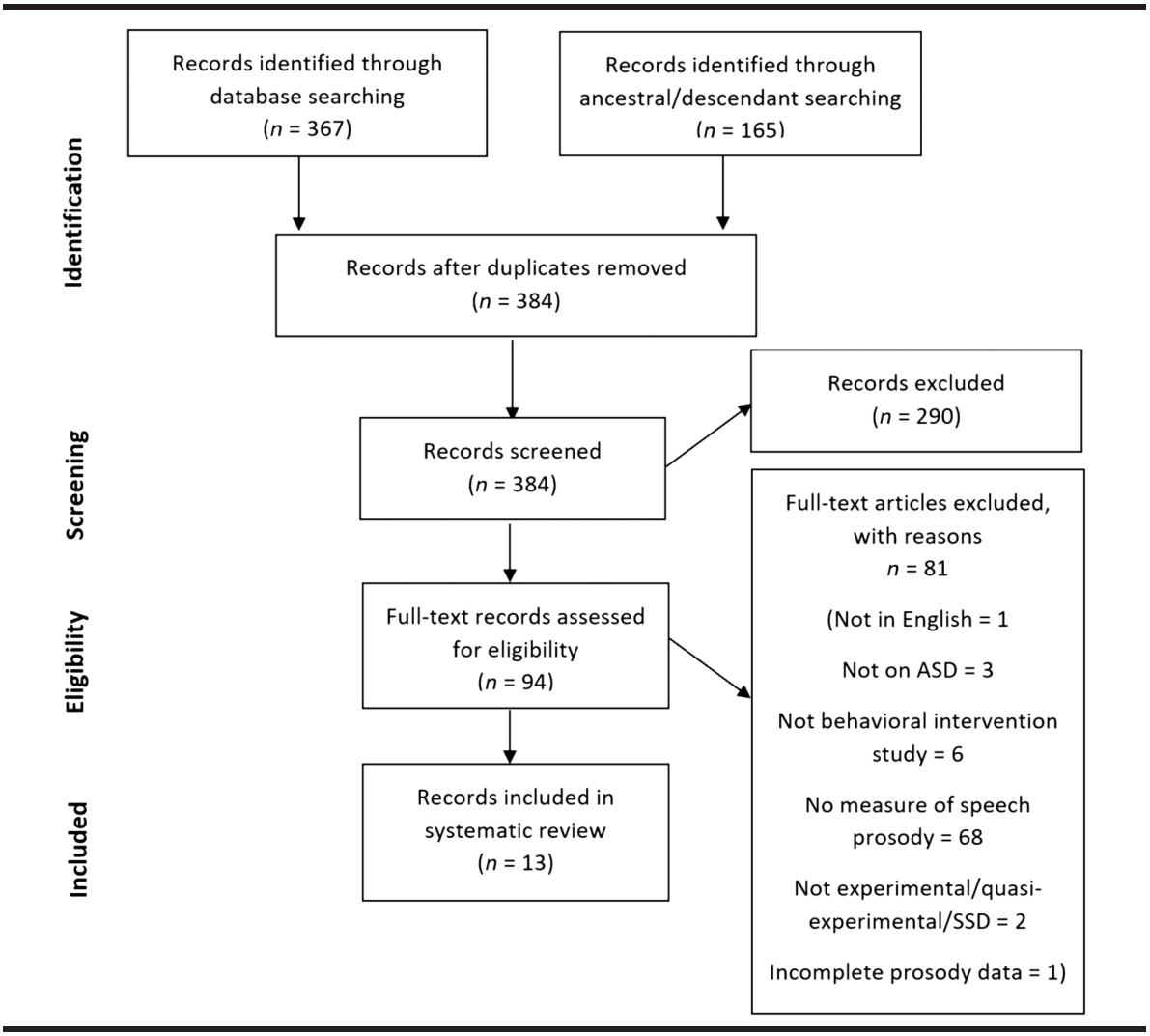
Participant Characteristics

A total of 71 individuals with ASD between the ages of 5;0 and 25;0 (years;months) participated in the studies. The individuals with ASD ranged in language abilities from minimally verbal (i.e., one to two words per utterance) to verbally fluent without any apparent language disorder. Most participants were verbally fluent with or without language difficulties. Only four of the studies reported IQ information for individuals, most of whom had IQs greater than 70. Although most of the studies were published after the publication of the *DSM-5* (American Psychiatric Association, 2013), participants' level of severity was not reported.

Independent Variables: The Interventions

Each of the 13 articles utilized different interventions and intervention packages. Within these, 12 intervention categories emerged, with some studies utilizing elements from more than one category. We used a combination of

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram illustrating the record selection process. These numbers reflect both the initial search and the updated search.



terminology from the NSP literature reviews and the Wong et al. (2013) review. In general, the categories of Wong et al. (2013) were more specific, so we largely utilized their verbiage. We maintained separate categories for the interventions that were part of the NSP definition of "behavioral interventions" but that were mentioned individually in the Wong et al. (2013) literature review (e.g., prompting) to promote clarity and specificity. The resulting 12 categories of intervention in this review were antecedent-based intervention, behavioral interventions, in vivo modeling, parent-implemented instruction, peer-mediated instruction and intervention, pivotal response training, prompting, reinforcement, scripting, social narratives, technology-based instruction and intervention, and video modeling. Below, we detail the interventions and intervention packages used in the included articles. They are organized in two groups based on the NSP designations of established and emerging interventions. There were no unestablished interventions in our review. Established interventions needed to have numerous peer-reviewed studies conducted by multiple researchers and/or research groups demonstrating the efficacy of the intervention being tested (National Autism Center, 2015; Wong et al., 2013). Wong et al.'s criteria for an evidence-based practice were that the practice had to have two or more high-quality experimental or quasi-experimental group design articles that were published by two different authors or groups of authors OR have five or more high-quality SSD articles published by at least three different authors or groups of authors OR a mix of one or more high-quality experimental or quasi-experimental group design articles and three high-quality SSD published by at least two different authors or groups of authors. The NSP criteria (National Autism Center, 2015) for an established intervention was that the intervention had "sufficient evidence of effectiveness," had "produced beneficial effects for individuals involved in the research studies," and "there

Table 1. Individual study and participant characteristics of included studies.

Quality rating	Study	Article type	Study design	No. of persons with ASD and prosody data	Age (in years)	IQ	Verbal ability
Adequate	Charlop et al. (2010)	PRJ	SSD	3	7–11	Not specified	Verbally fluent, not disordered
	Ormand (2016)	D	SSD	3	5–9	Not specified	Verbally fluent
Weak	Argott et al. (2017)	PRJ	SSD	4	10–12	Not specified	Verbally fluent, not disordered
	Boyd (2018) sayWAT Pitch[a]	D	SSD	4	22–25	Not specified	Not specified
	Boyd (2018) sayWAT Volume[a]	D	SSD	4	22–25	Not specified	Not specified
	Boyd (2018) vrSocial Volume[a]	D	SSD	9	7–14	Not specified	Not specified
	Daou et al. (2014)	PRJ	SSD	3	9–13	Not specified	Verbally fluent, not disordered
	Edgerton & Wine (2017)	PRJ	SSD	1	11	Not specified	Verbally fluent, not disordered
	Koegel (2014)	D	SSD	3	22–26	Not specified	Not specified
	Koegel & Frea (1993)	PRJ	SSD	1	13, 16	71 on Stanford–Binet (version not specified)	Not specified
	Mayo (2015)	D	G	15	*M* = 14	> 80	Verbally fluent, not disordered
	Nordgren (2016)	PRJ	SSD	2	5–6	"Severe intellectual disability"	Nonverbal and minimally verbal
	Ozdemir (2008)	PRJ	SSD	1	7–9	62, 86, 74 (WISC-III)	Minimally verbal–verbally fluent
	Parsons et al. (2018)	PRJ	G	10	6–12	Not specified	Average scores on Test of Auditory Comprehension of Language–Fourth Edition
	S. E. Simmons et al. (2016) full	PRJ	G	27	5–19	Not specified	"Speech containing full sentences"
	S. E. Simmons et al. (2016) ASD subset	PRJ	G	12	6–12	Not specified	"Speech containing full sentences"

Note. ASD = autism spectrum disorder; PRJ = peer-reviewed journal; D = dissertation; SSD = single-subject design; G = group design; WISC-III = Wechsler Intelligence Scale for Children–Third Edition.

[a]Boyd (2018) presented two assistive technologies relevant to this review, namely, sayWAT and vrSocial. sayWAT measured pitch and volume and vrSocial measured volume.

[was] no evidence of harmful effects" (National Autism Center, 2015, p. 77).

Established Evidence-Based Practices

Eleven established practices were represented in the 13 studies in this review, namely, antecedent-based intervention, behavioral interventions, in vivo modeling, pivotal response training, parent-implemented instruction, peer-mediated instruction and intervention, prompting, reinforcement, scripting, social narratives, and video modeling.

Antecedent-based intervention. Antecedent-based intervention is identified in the first NSP review and the Wong et al. review and is defined as intentionally arranging events or environments directly preceding a target behavior to either "increase the likelihood of success or reduce the likelihood of problems occurring" (National Autism Center, 2009, p. 44; Wong et al., 2013). Three articles in our review utilized this strategy, two directly (Ormand, 2016; Ozdemir, 2008) and one indirectly (Parsons et al., 2018), to target speech prosody. Children (ages 5–9 years) in the study of Ormand (2016) were reminded to use a quiet voice prior to starting their intervention session. Similarly, a social narrative was read to the child in the study of Ozdemir (2008) whose behavioral target was also decreased loud speech prior to times when he usually exhibited loud speech. Prosody was indirectly targeted in the study of Parsons et al. (2018) in part as therapists reviewed video footage of the children playing prior to their next structured play interaction with a peer.

Behavioral interventions. Behavioral interventions, as defined in the second NSP review, combined antecedent packages and behavioral packages into one larger category. This category of intervention covers a wide array of strategies that occur prior to or as a consequence of targeted behavior in order to either increase, decrease, or modify that behavior (National Autism Center, 2015). Two studies utilized these types of interventions (as defined here) to address prosodic targets with participants (Daou et al.,

Table 2. Prosody intervention characteristics of included studies.

Study	Tx description	Direct/indirect prosody Tx	Behavioral/acoustic measures	Prosody traits addressed	Tx session length (in min)	No. of times prosody measured
Charlop et al. (2010)[a]	Video modeling, scripting, reinforcement	D	B	GI	10–15	16–24
Ormand (2016)[a]	Antecedent modification, prompting, reinforcement, parent-implemented	D	A/B	Intensity	5	24–26
Argott et al. (2017)	Video modeling, in vivo modeling, prompting, scripting, reinforcement	D	B	GI	21–30	45–51
Boyd (2018) sayWAT Pitch[b]	Biofeedback, technology-aided instruction and intervention—augmented reality	D	A/B	Pitch	5	3
Boyd (2018) sayWAT Volume[b]	Biofeedback, technology-aided instruction and intervention—augmented reality	D	A/B	Intensity	5	3
Boyd (2018) vrSocial Volume[b]	Biofeedback, technology-aided instruction and intervention—virtual reality	D	A/B	Intensity	1	5
Daou et al. (2014)	Scripting, in vivo modeling, prompting, behavioral interventions (i.e., shaping, fading), reinforcement	D	B	GI	15–45	83–123
Edgerton & Wine (2017)	Biofeedback, in vivo modeling, prompting, reinforcement	D	A/B	Intensity	Unclear	33
Koegel (2014)	Pivotal response training, reinforcement	D	B	GI, pauses, intensity, rate	Unclear	13–22
Koegel & Frea (1993)	Pivotal response training, reinforcement	I	B	Intensity	Unclear	90
Mayo (2015)	Technology-aided instruction and intervention, practice	D	A	Contrastive stress	Unclear	4
Nordgren (2016)	In vivo modeling, prompting	I	A	Pitch	5–15	23, 34
Ozdemir (2008)	Social narratives, antecedent-based intervention	D	B	Intensity	~20	51
Parsons et al. (2018)	Video modeling, peer-mediated instruction and intervention, antecedent-based instruction, behavioral interventions (i.e., feedback), prompting, parent-implemented intervention	I	B	Stress, affective intonation	60	3
S. E. Simmons et al. (2016) full	Biofeedback, technology-aided instruction and intervention, prompting	D	B	GI, rate, stress, intensity	5–90; *M* (*SD*) = 21.25 (11.82)	2
S. E. Simmons et al. (2016) ASD subset	Biofeedback, technology-aided intervention, prompting	D	B	GI, rate, stress, intensity	10–30; *M* (*SD*) = 25.99 (6.25)	2

Note. Tx = intervention; D = direct; B = behavioral; GI = global intonation; A = acoustic; I = indirect.

[a]Articles of adequate instead of weak quality. [a]Boyd (2018) presented two assistive technologies relevant to this review, namely, sayWAT and vrSocial. sayWAT measured pitch and volume and vrSocial measured volume.

2014; Parsons et al., 2018). Daou et al. (2014) directly utilized shaping and fading procedures as a part of their intervention package to improve the global intonation patterns of the three participants (ages 9–13 years) in their study. Parsons et al. (2018) indirectly used feedback procedures to improve participants' stress and affective intonation.

In vivo modeling. In vivo modeling, often referred to as simply modeling, is defined as demonstrations of a desired behavior or a set of behaviors that takes place in person (as opposed to via video; National Autism Center, 2015; Wong et al., 2013). Three studies explicitly used in vivo modeling procedures to teach prosodic targets to the children in their studies. Two studies employed in vivo modeling to teach desired global intonation patterns (Argott et al., 2017; Daou et al., 2014) and one to affect intensity (Edgerton & Wine, 2017). One study used in vivo modeling to indirectly modify fundamental frequency measures (pitch; Nordgren, 2016).

Parent-implemented instruction. In this type of intervention, parents received focused instruction and training in how to deliver intervention to help their children improve in behavioral targets (National Autism Center, 2009, 2015; Wong et al., 2013). Two studies in this review utilized parent-implemented instruction to address prosody targets (Ormand, 2016; Parsons et al., 2018). Researchers in the study of Ormand (2016) instructed the parent of one of the four children in the study to deliver direct intervention for loud speech. Parents in the study of Parsons et al. (2018) were involved in indirect techniques (e.g., discussion, feedback) to improve prosody.

Peer-mediated instruction and intervention. Only one study in this review (Parsons et al., 2018) indirectly employed peer-mediated instruction and intervention to improve children's stress and affective intonation. This type of intervention involves peers with NTD interacting with children with ASD to increase naturalistic opportunities for the children with ASD to practice and observe target behaviors (National Autism Center, 2015; Wong et al., 2013). In the study of Parsons et al. (2018), prosody was indirectly targeted by having children with ASD and peers with NTD participate in a 30-min child-led play session after video modeling and feedback with the interventionist.

Pivotal response training. Two studies in this review employed this system, one directly (B. L. Koegel, 2014) and one indirectly (R. L. Koegel & Frea, 1993), to improve global intonation, pausing, rate (B. L. Koegel, 2014), and intensity (B. L. Koegel, 2014; R. L. Koegel & Frea, 1993) in young adults and children, respectively. Pivotal response training leverages key behaviors (e.g., motivation, self-management) to effect change in related behaviors in naturalistic environments (National Autism Center, 2015; Wong et al., 2013).

Prompting. Prompting is a widely used intervention defined as "verbal, gestural, or physical assistance" given to a person before or during their acquisition process for a given behavior (Wong et al., 2013, p. 21) and for the purposes of this review included verbal instruction. Seven of the 13 articles in this review employed prompting as a part of their prosody intervention packages (Argott et al., 2017; Daou et al., 2014; Edgerton & Wine, 2017; Nordgren, 2016; Ormand, 2016; Parsons et al., 2018; E. S. Simmons et al., 2016) to address a variety of prosody targets, including global intonation, intensity, pitch, stress, affective intonation, and rate. Two of the studies indirectly targeted prosody (Nordgren, 2016; Parsons et al., 2018), and the other five directly targeted prosodic function (Argott et al., 2017; Daou et al., 2014; Edgerton & Wine, 2017; Ormand, 2016; E. S. Simmons et al., 2016).

Reinforcement. Reinforcement is a technique that employs a contingent response after a desired behavior with the goal of increasing that behavior in the future (Wong et al., 2013). Seven articles in this review utilized reinforcement to directly (Argott et al., 2017; Charlop et al., 2010; Daou et al., 2014; Edgerton & Wine, 2017; R. L. Koegel & Frea, 1993; Ormand, 2016) and indirectly (R. L. Koegel & Frea, 1993) affect prosody the traits of global intonation, intensity, pausing, and rate.

Scripting. Three articles in our review employed scripting to directly target global intonation patterns in prescribed utterances (Argott et al., 2017; Charlop et al., 2010; Daou et al., 2014). Scripting involves a verbal or written model of the language that is to be used in targeted situations. Scripts are often practiced repeatedly prior to using them in the targeted context (National Autism Center, 2009, 2015; Wong et al., 2013).

Social narratives. Social narratives are short stories that describe social situations, while highlighting critical social cues and expected behavior that are individualized for the learner (National Autism Center, 2015; Wong et al., 2013). Social narratives were used to directly target loud speech for one child his school setting in the study of Ozdemir (2008).

Video modeling. Video modeling is when a desired behavior is performed by a peer, adult, or the individual and is filmed and made into a short film that the participant views as a part treatment. Video modeling was used to treat global intonation directly in two studies in this review (Argott et al., 2017; Charlop et al., 2010) and to indirectly treat stress and affective intonation in another study in the review (Parsons et al., 2018).

Emerging Evidence-Based Practices

Emerging evidence-based practices were those that had a less robust evidence base. They had fewer studies demonstrating their effectiveness and may not have been performed by multiple research groups, so interventionists desiring to use them should exercise more caution and more carefully evaluate their results with individual clients (Wong et al., 2013).

Technology-aided instruction and intervention. These interventions were ones that featured technology and technological equipment as a primary tool for bringing about change in a targeted behavior and included measures of biofeedback (Wong et al., 2013). These kinds of interventions were listed as confirmed evidence-based practices by (Wong et al., 2013), but not by the second NSP review (National Autism Center, 2015), in which they were classified as emerging.

We elected to go with the more conservative NSP classification. Technology-aided instruction and interventions were used in three of the articles in this review to directly target global intonation, pitch, intensity, stress, and rate (Boyd, 2018; Mayo, 2015; E. S. Simmons et al., 2016).

Dependent Variables

Dependent variables related to prosody fell into two broad categories: perceptual and acoustic analysis of prosody. Ten of the 13 articles employed perceptual ratings (Argott et al., 2017; Charlop et al., 2010; Daou et al., 2014; Edgerton & Wine, 2017; B. L. Koegel, 2014; R. L. Koegel & Frea, 1993; Ormand, 2016; Ozdemir, 2008; Parsons et al., 2018; E. S. Simmons et al., 2016), and three reports employed acoustic analysis of prosody (Boyd, 2018; Mayo, 2015; Nordgren, 2016).

Perceptual Ratings

General prosodic functioning. Overall prosodic functioning was rated in five studies (Argott et al., 2017; Charlop et al., 2010; Daou et al., 2014; B. L. Koegel, 2014; E. S. Simmons et al., 2016). Interventionists in the studies of Argott et al. (2017), Charlop et al. (2010), and Daou et al. (2014) coded prosody in utterances as correct or incorrect prosody based on predetermined descriptions. Speech-language pahtologists in the study of E. S. Simmons et al. (2016) rated their students' overall intonation as "typical," "mildly atypical," or "clearly atypical." Participants in the study of B. L. Koegel (2014) used an interval coding system to rate their own prosody as appropriate or not according to prespecified criteria.

Specific prosodic features. Six studies included ratings of specific features of prosody as a dependent variable (Edgerton & Wine, 2017; R. L. Koegel & Frea, 1993; Ormand, 2016; Ozdemir, 2008; Parsons et al., 2018; E. S. Simmons et al., 2016). In five of the six, appropriate vocal intensity was a primary target (Edgerton & Wine, 2017; R. L. Koegel & Frea, 1993; Ormand, 2016; Ozdemir, 2008; E. S. Simmons et al., 2016). The expressive prosody-related dependent variables in the study of Parsons et al. (2018) were the Focus and Affect subtests of the Profiling Elements of Prosody in Speech-Communication test (Peppé & McCann, 2003), which assess tone (or pitch) and stress, respectively.

Acoustic analysis. Boyd (2018), Mayo (2015), and Nordgren (2016) provided acoustic-based analysis of prosody. Nordgren (2016) and one substudy by Boyd (2018) targeted aspects of pitch (i.e. the pitch aspect of the assistive technology named sayWAT). The other two substudies by Boyd (2018) addressed appropriate vocal intensity (i.e. the volume aspect of the assistive technologies named sayWAT and vrSocial). Durational measures were taken in the studies of Mayo (2015) and Nordgren (2016), along with Swedish Accent 2 in the latter.

Study Quality

As can be seen in Table 1, only two articles met the criteria for "adequate" study quality (Charlop et al., 2010; Ormand, 2016). The other 11 articles were weak in quality.

Details for this decision for according to the Reichow and colleagues' quality rating system described previously (Reichow, 2011; Reichow et al., 2008) are displayed for SSD studies in Table 3 and for group studies in Table 4. Of particular note, only seven of the 16 separate analyses achieved enough evidence for social validity, a critical feature to incorporate when evaluating speech prosody interventions due to the potentially stigmatizing effect atypical prosody may have (Bone et al., 2015; de Marchena & Miller, 2017; Redford et al., 2018).

Strength of Evidence

Effect sizes. A summary of the studies' effects is presented in Table 5. The only studies that reported effect sizes were those of Mayo (2015) and E. S. Simmons et al. (2016). All other effect sizes were hand-calculated. We only extracted or calculated effects sizes with data directly measuring elements of prosody for this review. This changed the designs of R. L. Koegel and Frea (1993) and Ozdemir (2008) to AB designs consisting of one baseline phase and one intervention phase only.

Effects of studies with adequate evidence quality. The intervention effects of Charlop et al. (2010) resulted in Tau-U scores of 1 and 0.963 for baseline versus treatment and baseline versus follow-up, respectively. These scores indicated that the intervention was highly effective for improving prosody in prescribed social interactions for the duration of treatment and that these effects were generalized to different interventionists who provided different social stimuli in another intervention room. The intervention used in the study of Ormand (2016) was highly effective at reducing inappropriately loud vocal intensity during 5-min treatment sessions with a familiar adult as evidenced by a Tau-U score of −1.15. No generalization or follow-up data were reported.

Effects of studies with weak evidence quality. All other studies in this review were weak in quality, so their effects are not well established at this time, even if they achieved large effect sizes and/or used established interventions. Participants in three of the SSD design studies (Argott et al., 2017; B. L. Koegel, 2014; Ozdemir, 2008) and the subset of participants with ASD in the study of E. S. Simmons et al. (2016) showed large effect sizes as an immediate effect of treatment. Three studies demonstrated moderate effect sizes as an immediate result of treatment (Daou et al., 2014; Edgerton & Wine, 2017; E. S. Simmons et al., 2016). Effect sizes for the three substudies in Boyd (2018) were small or in the undesired direction, even with direct treatment of prosody. The effects of the weak studies that indirectly targeted prosody varied. R. L. Koegel and Frea (1993) showed a weak but significant effect of their intervention on a participant's vocal intensity. The effect sizes for the indirect intervention in the study of Nordgren (2016) and were small and nonsignificant. None of the pretest versus posttest comparison in the study of Parsons et al. (2018)

Table 3. Quality indicators for single-subject design studies.

Study	Primary quality indicators						Secondary quality indicators						Rating
	PART	IV	BSLN	DV	VIS ANAL	EXP CON	IOA	KAP	BR	FID	G/M	SV	
Argott et al. (2017)	A	H	U	H	A	H	Yes	No	No	Yes	Yes	No	Weak
Boyd (2018) sayWAT Pitch[a]	A	H	H	U	U	U	No	No	No	No	No	No	Weak
Boyd (2018) sayWAT Volume[a]	A	H	H	A	U	U	No	No	No	No	No	No	Weak
Boyd (2018) vrSocial Volume[a]	A	H	H	U	U	U	No	No	No	No	No	No	Weak
Charlop et al. (2010)	A	H	H	H	H	H	Yes	No	No	Yes	Yes	No	Adequate
Daou et al. (2014)	H	H	A	H	A	A	Yes	No	No	Yes	Yes	No	Weak
Edgerton & Wine (2017)	U	A	U	A	A	U	Yes	No	No	No	Yes	No	Weak
Koegel (2014)	A	H	A	H	A	A	Yes	Yes	No	Yes	Yes	Yes	Weak
Koegel & Frea (1993)	A	A	U	H	U	U	Yes	No	No	No	No	Yes	Weak
Nordgren (2016)	H	H	U	A	U	U	No	No	No	No	No	Yes	Weak
Ormand (2016)	H	A	A	H	H	H	Yes	No	No	No	No	Yes	Adequate
Ozdemir (2008)	H	H	U	A	A	U	Yes	No	No	Yes	Yes	Yes	Weak

Note. Quality ratings are based on study data specific to prosody goals. For ratings of overall studies, please contact the first author. PART = participant characteristics; IV = independent variable; BSLN = baseline condition; DV = dependent variable; VIS ANAL = visual analysis; EXP CON = experimental control; IOA = interobserver agreement; KAP = kappa; BR = blind raters; FID = fidelity; G/M = generalization and/or maintenance; SV = social validity; A = acceptable; H = high; U = unacceptable.

[a]Boyd (2018) presented two assistive technologies relevant to this review, namely, sayWAT and vrSocial. sayWAT measured pitch and volume and vrSocial measured volume.

reached statistical significance, and the expressive prosody effects sizes were moderately small.

Discussion

The purpose of this systematic review was to identify and summarize the efficacy of interventions specifically targeting the improvement of speech prosody in persons with ASD. We searched 13 electronic databases using search terms related to autism, intervention, and prosody and found seven articles that fit our inclusion criteria. After descendant and reference searching of the original seven articles and two literature reviews of speech prosody in ASD (Fusaroli et al., 2017; McCann & Peppé, 2003), 13 articles were ultimately included in our review. Information was extracted on coder and coding process, report, setting, participant, intervention, and outcome characteristics, as well as results and quality indicators. Overall, results showed that interventions that specifically targeted prosody over time using established evidence-based practices in ASD resulted in moderate to large improvements in prosody in persons with ASD. The areas of prosody that were most successfully targeted (i.e., had the largest effect sizes) were intensity and global intonation in prescribed contexts. Those interventions that indirectly targeted speech prosody or only targeted prosody across 1 day resulted in small or insignificant effects. These results are cautiously encouraging but tentative at this time, given the overall low quality of study design of the included studies.

Direct Versus Indirect Intervention

The results of this review suggest that the largest gains in prosody are most likely to be made when an intervention specifically targets speech prosody. All the studies that indirectly targeted prosody (R. L. Koegel & Frea, 1993; Nordgren, 2016; Parsons et al., 2018) reported small gains in speech prosody targets. The only studies that achieved moderate or large effect sizes were those in which prosody was directly targeted, suggesting that directly targeting

Table 4. Quality indicators for group design studies.

Study	Primary quality indicators						Secondary quality indicators								Rating
	PART	IV	CC	DV	LRQ	STAT	RA	IOA	BR	FID	ATR	G/M	ES	SV	
Mayo (2015)	U	H	H	H	A	A	No	No	No	No	Yes	No	No	No	Weak
Parsons et al. (2018)	A	A	U	A	A	A	No	No	No	No	Yes	Yes	Yes	Yes	Weak
S. E. Simmons et al. (2016)	H	A	U	H	H	H	No	Yes	No	No	Yes	No	Yes	Yes	Weak

Note. Quality ratings are based on study data specific to prosody goals. For ratings of overall studies, please contact the first author. PART = participant characteristics; IV = independent variable; CC = comparison condition; DV = dependent variable; LRQ = link between research question and data analysis; STAT = use of statistical tests; RA = random assignment; IOA = interobserver agreement; BR = blind raters; FID = fidelity; ATR = attrition; G/M = generalization and/or maintenance; ES = effect size; SV = social validity; U = unacceptable; H = high; A = acceptable.

Table 5. Effect sizes of speech prosody outcomes.

Quality rating	Study	Design type	Baseline vs. Tx Tau-U	*p*	Baseline vs. follow-up Tau-U	*p*	Cohen's *d* or Hedge's *g*	*p*
Adequate	Charlop et al. (2010)	SSD	1	< .0001	0.96	.005	—	—
	Ormand (2016)	SSD	–1.15[a]	< .0001	—	—	—	—
Weak	Argott et al. (2017)	SSD	Trial: 0.99 Generalization: 0.94	< .0001 < .0001	Trial: 1.10 Generalization: 1.17	< .0001 < .0001	—	—
	Boyd (2018) sayWAT Pitch[b]	SSD	–0.11	.66	—	—	—	—
	Boyd (2018) sayWAT Volume[b]	SSD	0.11	.66	—	—	—	—
	Boyd (2018) vrSocial Volume[b]	SSD	–0.22	.08	—	—	—	—
	Daou et al. (2014)	SSD	0.74	< .0001	0.91	< .0001	—	—
	Edgerton & Wine (2017)	SSD	0.84	.005	—	—	—	—
	Koegel (2014)	SSD	1.02	< .0001	1.16	.02	—	—
	Koegel & Frea (1993)	SSD	0.60	< .0001	—	—	—	—
	Mayo (2015) ASD baseline[c]	G	—	—	—	—	S: *d* = 0.64 V: *d* = 0.26 NP: *d* = 0.83 PPh: *d* = 0.62	S: .03 V: .27 NP: .01 PPh: .09
	Mayo (2015) ASD postintervention[c]	G	—	—	—	—	S: *d* = 1.94 V: *d* = 1.02 NP: *d* = 1.99 PPh: *d* = 1.38	S: < .001 V: .02 NP: .001 PPh: .09
	Nordgren (2016)	SSD	F0 Min: –0.32 F0 Mean: –0.38 F0 Max: –0.14	F0 Min: 0.14 F0 Mean: 0.08 F0 Max: 0.51	—	—	—	—
	Ozdemir (2008)	SSD	–1.01[a]	.0001	–1.03[a]	.0002	—	—
	Parsons et al. (2018)[d]	G	—	—	—	—	AO: *g* = .36 FO: *g* = .32	AO: .77 FO: .75
	S. E. Simmons et al. (2016) overall[d]	G	—	—	—	—	Rate: NS WS: *d* = 0.48 SS: *d* = 0.77 I: *d* = 0.77 GI: *d* = 0.71	Rate: .10 WS: .01 SS: .001 I: .001 GI: .001
	S. E. Simmons et al. (2016) ASD subset[d]	G	—	—	—	—	Rate: NS WS: NS SS: *d* = 0.80 I: *d* = 0.90 GI: *d* = 0.81	Rate: .12 WS: .34 SS: .02 I: .02 GI: .02

Note. Tx = intervention; SSD = Single-subject design; ASD = autism spectrum disorder; G = group design; S = sentence; V = verb pause; NP = noun phrase; PPh = prepositional phrase; AO = affect output; WS = stress in words; SS = stress in sentences; I = intensity; GI = global intonation.

[a]In these instances, a negative effect size was desirable. [a]Boyd (2018) presented two assistive technologies relevant to this review, namely, sayWAT and vrSocial. sayWAT measured pitch and volume and vrSocial measured volume. [c]Cohen's *d* and associated *p* values calculated for and between mean durations of sentence types of high functioning autism group, not baseline versus posttest differences. [d]Cohen's *d* calculated between pre- and postintervention group means.

prosody is likely necessary to achieve large improvements in this area. Yet, direct intervention of speech prosody does appear to be sufficient for large gains. The small, nonsignificant effects achieved by Boyd (2018) and the inconsistent effects of Mayo (2015) and E. S. Simmons et al. (2016) provide some evidence for this, although a confound for this explanation of the results of the studies of Mayo (2015) and Boyd (2018) may be that the interventions only took place across 1 day, so it is possible that the small or inconsistent effects were due to the shortness of the intervention rather than the ineffectiveness of the treatment used. Another complicating factor for interpreting the results of Mayo (2015) was that they did not report effect sizes that were calculated across time points, but only within time points. Thus, no direct pre- to postmeasurements were calculated, and the direct effect of treatment cannot be seen.

Evidence-Based Practices for Speech Prosody

Each of the 12 interventions used in this review was classified as established or emerging by the NSP (National Autism Center, 2015) or as an evidence-based practice according to the review by Wong et al. (2013). The widespread use of evidence-based practices in the interventions and intervention packages included in this review allows for some cautious optimism that their use for addressing speech prosody in persons with ASD may be effective. It is possible that interventionists who wish to address speech prosody can use combinations of the 12 evidence-based treatments in this review and expect measurable improvement in speech prosody. However, caution should be used because the general low study quality does not allow for high levels of confidence in the results of this study. Similarly, interventionists should not assume that whole intervention packages will effect prosodic change just because they include evidence-based elements. The elements of these packages were not tested individually, so it is unclear which ones were of critical importance for changing prosody. Interventionists will need to verify the effects of any particular combination of interventions individually. The interventions that were in packages with fewer elements, such as those of Charlop et al. (2010) and Ozdemir (2008), may provide a clearer picture of which interventions were the probable agents of change.

Video modeling, antecedent-based intervention, prompting/instruction, scripting, and reinforcement (five evidence-based practices cited by Wong et al., 2015) were used in studies that had adequate quality and achieved large effects. Based on the evidence from this review, these five interventions are the ones most likely to result in changes to prosody in persons with ASD, but this conclusion must be verified by individual clinicians until higher levels are available and evidence on the complete intervention packages is available.

Measurement method. Two types of analysis of prosody were present in the articles in this review, namely, acoustic analyses of prosody and perceptual analyses of prosody, but only three of the studies employed acoustic ratings. The accuracy and reliability of perceptual prosody ratings has been called into question (Diehl & Paul, 2013). While we did not find that interobserver agreement was unacceptable in any of the studies that used perceptual ratings, only one of the SSD studies that used perceptual ratings calculated kappa, which is a better measure of reliability than percent agreement (Reichow, 2011). To establish the need for and efficacy of speech prosody interventions, it is important to use acoustic ratings in more intervention studies that address speech prosody in ASD (see Dahlgren et al., 2018; Redford et al., 2018). Yet, the fact that larger effect sizes were evident in studies that employed perceptual ratings seems consistent with Dahlgren et al.'s (2018) suggestion that detection of speech prosody differences in persons with ASD may require a combination of acoustic and perceptual measures to adequately capture what actually "got better" in persons with ASD and prosodic difficulties.

Acoustic versus perceptual measurements. Most of the articles included in this review used perceptual measurements of speech prosody, and in general, these studies reported moderate to large gains in prosody. At first glance, results from the articles with acoustic measurements of prosody reflect those of Dahlgren et al. (2018) and Fusaroli et al. (2017), who found that consistent, socially meaningful acoustic differences were not present between persons with ASD and persons with NTD. This conclusion needs to be interpreted with caution, however, as it is possible that the lack of significant, socially meaningful changes in these studies was not due to the measurement method but to other factors. In the studies of Boyd (2018) and Mayo (2015), the lack of meaningful change may have been more attributable to the fact that the treatment only occurred for 1 day. In the study of Nordgren (2016), the nonsignificant results may have been due to the indirect nature of the treatment more than the measurement method. As such, these results do not call into question the usefulness of speech prosody interventions for persons with ASD, as has been suggested by Redford et al. (2018), but rather emphasize the importance of careful assessment and clinical decision making for each individual to determine if speech prosody is a barrier to functioning (de Marchena & Miller, 2017; Paul et al., 2005; Shriberg et al., 2001). Nevertheless, because perceptual ratings are more prone to bias and lack of reliability, it is important, where possible, to incorporate acoustic measurement in the assessment of speech prosody.

Social validity considerations. When making evidence-based decisions about interventions for speech prosody in persons with ASD, interventionists need to consider if the results of the intervention will translate to the daily social interactions of their clients (Kennedy, 2002; Luiselli & Reed, 2011). Only seven of the 13 studies in this review achieved a positive rating in social validity. So, it is possible that, while some of the interventions had large effect sizes, the results may not generalize to authentic social interactions. Consequently, interventionists need to carefully evaluate intervention approaches with regard to their clients' needs

and preferences before implementation, which is consistent with the principles of evidence-based practice espoused by the American Speech-Language-Hearing Association (2005).

Intervention approaches. Our findings partially supported the idea discussed by Diehl and Paul (2013) that, in structured, discrete tasks, persons with ASD may learn the prosodic "rules" of the task more readily than in naturalistic tasks, which require online processing and adjustment. Interventions that utilized discrete, exact modeling procedures (e.g., Argott et al., 2017; Charlop et al., 2010) resulted in large effect sizes. Both studies that had adequate quality (Charlop et al., 2010; Ormand, 2016) fell into this category, further lending credence to Diehl and Paul's conclusions. Yet, large effects were associated with interventions conducted in more naturalistic conversational contexts (e.g., B. L. Koegel, 2014; E. S. Simmons et al., 2016). Thus, it is possible that large changes in prosody may result in interventions with varying levels of prescriptiveness. In all, our results tentatively suggest that interventions that are highly structured seem to be effective, but that more naturalistic interventions also hold promise in their efficacy. This conclusion is very preliminary at this time, however, because of the general low study quality of the articles included in this review.

Clinical implications. Because not all persons with ASD present with speech prosody differences (Dahlgren et al., 2018; Fusaroli et al., 2017; Paul et al., 2005) and because of the general weak quality of evidence for the speech prosody interventions in this review, interventionists need to make careful decisions when addressing prosody in intervention with persons with ASD. In service of careful, data-driven clinical decision making, interventionists need to first establish that speech prosody is a barrier to functioning in their clients with ASD. Then, they may consider conducting informal SSD studies to determine the effects of the prosody interventions they choose. This will allow them to individually determine if a speech prosody intervention is effective and socially relevant for their clients.

Limitations and Future Directions

Several limitations limit the interpretability and generalizability of the results of this review. Because one of the goals of this review was to be as comprehensive as possible, we included dissertations as well as peer-reviewed articles. The results from dissertations cannot be accepted as readily as peer-reviewed articles because they have not gone through the refining, winnowing process of peer review. Future reviews may be restricted to peer-reviewed studies only to provide an overall higher level of confidence in the quality of the evidence presented.

Most of the studies employed perceptual ratings of prosody rather than acoustic measurements. Perceptual ratings can be prone to bias and lack of reliability, so future investigations into interventions for speech prosody should incorporate acoustic measurement of speech prosody to provide more objective measurements of prosodic improvement.

This review is also limited because of the lack of representation of all persons with ASD and all aspects of prosody in the included studies. Most of the included studies were with individuals who were verbal. Only two of the articles included participants who were nonverbal or minimally verbal. Similarly, most studies did not report IQ measures, and no studies reported *DSM-5* severity levels. Only three of the articles included adults with ASD, and no studies involved participants over 30 years of age. The prosody characteristics addressed in these studies are not comprehensive. Many of the studies focused on intensity or global intonation in prescribed contexts, but few of the studies addressed more subtle aspects of prosody such as contextually appropriate pausing. The restricted nature of the participant and prosody characteristics limits the clinical utility of the results because, for example, the interventions that were effective for intensity in verbal children may not be effective in addressing pitch in verbal adults over the age of 30 years. Currently, the state of evidence on interventions for speech prosody in persons with ASD is neither reflective of the entire population of individuals with ASD nor the full range of potential prosodic impairment. Future studies in this area should seek to broaden the representation of persons with ASD at more levels of severity and across the life span. Similarly, future investigations would do well to target a broader range of prosody characteristics.

Finally, as noted previously, the overall methodological quality of the studies included in this review was low. No studies achieved a "strong" quality rating, and only two achieved an "adequate" rating. All others were classified as "weak." It is clear that, to draw firm conclusions about the efficacy of interventions for speech prosody in persons with ASD, studies with better methodological quality are needed.

Conclusions

The results of this review present limited evidence for interventions for speech prosody in ASD. Until a larger body of higher quality research is available, perhaps the most that can be said is that using established interventions to directly target prosody over longer periods of time is more likely to result in improvements than other techniques. This conclusion can most readily be applied to children and adolescents with ASD who are verbal given the characteristics of the participants in this review. Because of the limited representation of all levels of ASD and low study quality, persons with ASD, their family members, and their interventionists need to carefully engage in the process of clinical decision making and progress monitoring if prosody is determined to be a barrier to communication in order to most effectively determine what intervention or set of interventions will work best to improve prosodic function.

Acknowledgments

We would like to thank Ron Gillam for his guidance in conducting this review and Sandi Gillam for editing manuscript drafts. We gratefully acknowledge both for their support of our work.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

American Speech-Language-Hearing Association. (2005). *Evidence-based practice in communication disorders* [Position statement]. www.asha.org/policy

Argott, P. J., Townsend, D. B., & Poulson, C. L. (2017). Acquisition and generalization of complex empathetic responses among children with autism. *Behavior Analysis in Practice, 10*(2), 107–117. https://doi.org/10.1007/s40617-016-0171-7

Asperger, H. (1991). "Autistic psychopathy" in childhood. In U. Frith (Ed.), *Autism and Asperger syndrome* (pp. 37–92). Cambridge University Press. https://doi.org/10.1017/CBO9780511526770.002

Bellon-Harn, M. L., Harn, W. E., & Watson, G. D. (2007). Targeting prosody in an eight-year-old child with high-functioning autism during an interactive approach to therapy. *Child Language Teaching and Therapy, 23*(2), 157–179. https://doi.org/10.1177/0265659007076292

Bone, D., Black, M. P., Ramakrishna, A., Grossman, R. B., & Narayanan, S. S. (2015). Acoustic–prosodic correlates of 'awkward' prosody in story retellings from adolescents with autism. *Interspeech 2015,* 1616–1620. https://pdfs.semanticscholar.org/d0a6/3db67024f39d633865bddf5cb5f198f3bbdd.pdf

Boyd, L. E. (2018). *Designing and evaluating alternative channels: Visualizing nonverbal communication through AR and VR systems for people with autism*. ProQuest.

Charlop, M. H., Dennis, B., Carpenter, M. H., & Greenberg, A. L. (2010). Teaching socially expressive behaviors to children with autism through video modeling. *Education & Treatment of Children, 33*(3), 371–393. https://doi.org/10.1353/etc.0.0104

Cooper, H. (2017). *Research synthesis and meta-analysis: A step-by-step approach* (5th ed., Vol. 2). SAGE.

Dahlgren, S., Sandberg, A. D., Strömbergsson, S., Wenhov, L., Råstam, M., & Nettelbladt, U. (2018). Prosodic traits in speech produced by children with autism spectrum disorders—Perceptual and acoustic measurements. *Autism & Developmental Language Impairments, 3,* 239694151876452. https://doi.org/10.1177/2396941518764527

Daou, N., Vener, S. M., & Poulson, C. L. (2014). Analysis of three components of affective behavior in children with autism. *Research in Autism Spectrum Disorders, 8*(5), 480–501. https://doi.org/10.1016/j.rasd.2014.01.005

de Marchena, A., & Miller, J. (2017). "Frank" presentations as a novel research construct and element of diagnostic decision-making in autism spectrum disorder: Frank ASD. *Autism Research, 10*(4), 653–662. https://doi.org/10.1002/aur.1706

Diehl, J. J., & Berkovits, L. (2010). Is prosody a diagnostic and cognitive bellwether of autism spectrum disorders. In A. Harrison (Ed.), *Speech disorders: Causes, treatments, and social effects* (pp. 159–176). Nova Science.

Diehl, J. J., & Paul, R. (2013). Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children by children with autism spectrum disorders. *Applied Psycholinguistics, 34*(1), 135–161. https://doi.org/10.1017/S0142716411000646

Edgerton, L., & Wine, B. (2017). Speak up: Increasing conversational volume in a child with autism spectrum disorder. *Behavior Analysis in Practice, 10*(4), 407–410. https://doi.org/10.1007/s40617-016-0168-2

Ferdosi, N., Ashayeri, H., Modarresi, Y., & Rovshan, B. (2013). The effects of melodic intonation therapy on 7–10 year, Persian, autistic, male children's speech. *Advances in Cognitive Science, 15*(3), 12–23.

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). Is voice a marker for autism spectrum disorder? A systematic review and meta-analysis. *Autism Research, 10*(3), 384–407. https://doi.org/10.1002/aur.1678

Gordon, J. K., Andersen, K., Perez, G., & Finnegan, E. (2019). How old do you think I am? Speech-language predictors of perceived age and communicative competence. *Journal of Speech, Language, and Hearing Research, 62*(7), 2455–2472. https://doi.org/10.1044/2019_JSLHR-L-19-0025

Grossman, R. B., Edelson, L. R., & Tager-Flusberg, H. (2013). Emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism. *Journal of Speech, Language, and Hearing Research, 56*(3), 1035–1044. https://doi.org/10.1044/1092-4388(2012/12-0067)

Huang, X., Lin, J., & Demner-Fushman, D. (2006). Evaluation of PICO as a knowledge representation for clinical questions. *AMIA Annual Symposium Proceedings, 2006,* 359–363.

Jaramillo, S. B. (2018). *Acoustic and prosodic analysis of pre-verbal vocalizations of 18-month old toddlers with autism spectrum disorder (Master's thesis, Université du Québec)*. http://espace.inrs.ca/id/eprint/6655

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2*(3), 217–250.

Kennedy, C. H. (2002). The maintenance of behavior change as an indicator of social validity. *Behavior Modification, 26*(5), 594–604. https://doi.org/10.1177/014544502236652

Koegel, B. L. (2014). *Targeting prosodic atypicalities using self-management for individuals with autism spectrum disorders (2015-99150-320)*. ProQuest Information & Learning.

Koegel, R. L., & Frea, W. D. (1993). Treatment of social behavior in autism through the modification of pivotal social skills. *Journal of Applied Behavior Analysis, 26*(3), 369–377. https://doi.org/10.1901/jaba.1993.26-369

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLOS Medicine, 6*(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100

Lim, H. A. (2010). Effect of "Developmental speech and language training through music" on speech production in children with autism spectrum disorder. *Journal of Music Therapy, 47*(1), 2–26. https://doi.org/10.1093/jmt/47.1.2

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule–Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205–223. https://doi.org/10.1023/A:1005592401947

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule–Generic (ADOS-G)*. Western Psychological Services.

Luiselli, J. K., & Reed, D. D. (2011). Social validity. In S. Goldstein & J. A. Naglieri (Eds.), *Encyclopedia of child behavior and*

development (p. 1406). Springer. https://doi.org/10.1007/978-0-387-79061-9_3168

Matsuda, S., & Yamamoto, J. (2013). Intervention for increasing the comprehension of affective prosody in children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 7*(8), 938–946. https://doi.org/10.1016/j.rasd.2013.04.001

Mayo, J. (2015). *Prosodic phrasing in adolescents with high functioning autism: Production following intervention and under dual load conditions*. 62.

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders, 38*(4), 325–350. https://doi.org/10.1080/1368282031000154204

Mesibov, G. B. (1992). Treatment issues with high-functioning adolescents and adults with autism. In E. Schopler & G. B. Mesibov (Eds.), *High-functioning individuals with autism* (pp. 143–155). Springer.

Miller, S. B., & Toca, J. M. (1979). Adapted melodic intonation therapy: A case study of an experimental language program for an autistic child. *The Journal of Clinical Psychiatry, 40*(4), 201–203.

Nadig, A., & Shaw, H. (2012). Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners. *Journal of Autism & Developmental Disorders, 42*(4), 499–511. https://doi.org/10.1007/s10803-011-1264-3

National Autism Center. (2009). *National standards report: The national standards project—Addressing the need for evidence-based practice guidelines for autism spectrum disorders*.

National Autism Center. (2015). *Findings and conclusions: National standards project, Phase 2*.

Nordgren, P. M. (2016). F0, F0 range and duration of utterances—Longitudinal single-subject studies of prosody in two Swedish children with ASC. *Clinical Linguistics & Phonetics, 30*(1), 29–48. https://doi.org/10.3109/02699206.2015.1098728

Ormand, H. M. (2016). *An evaluation of a multi-component intervention for loud speech in children with autism spectrum disorder*. http://hdl.handle.net/2152/43674

Ozdemir, S. (2008). The effectiveness of social stories on decreasing disruptive behaviors of children with autism: Three case studies. *Journal of Autism and Developmental Disorders, 38*(9), 1689–1696. https://doi.org/10.1007/s10803-008-0551-0

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy, 42*(2), 284–299. https://doi.org/10.1016/j.beth.2010.08.006

Parsons, L., Cordier, R., Munro, N., & Joosten, A. (2018). The feasibility and appropriateness of a peer-to-peer, play-based intervention for improving pragmatic language in children with autism spectrum disorder. *International Journal of Speech-Language Pathology, 21*(4), 412–424. https://doi.org/10.1080/17549507.2018.1492630

Paul, R., Shriberg, L. D., McSweeny, J., Cicchetti, D., Klin, A., & Volkmar, F. (2005). Brief report: Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism & Developmental Disorders, 35*(6), 861–869. https://doi.org/10.1007/s10803-005-0031-8

Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics & Phonetics, 17*(4–5), 345–354. https://doi.org/10.1080/0269920031000079994

Rakap, S. (2015). Effect sizes as result interpretation aids in single-subject experimental research: Description and application of four nonoverlap methods. *British Journal of Special Education, 42*(1), 11–33. https://doi.org/10.1111/1467-8578.12091

Redford, M. A., Kapatsinski, V., & Cornell-Fabiano, J. (2018). Lay listener classification and evaluation of typical and atypical children's speech. *Language and Speech, 61*(2), 277–302. https://doi.org/10.1177/0023830917717758

Reichow, B. (2011). Development, procedures, and application of the evaluative method for determining evidence-based practices in autism. In B. Reichow, P. Doehring, D. V. Cicchetti, & F. R. Volkmar (Eds.), *Evidence-based practices and treatments for children with autism* (pp. 25–39). Springer. https://doi.org/10.1007/978-1-4419-6975-0_2

Reichow, B., Volkmar, F. R., & Cicchetti, D. V. (2008). Development of the evaluative method for evaluating and determining evidence-based practices in autism. *Journal of Autism and Developmental Disorders, 38*(7), 1311–1319. https://doi.org/10.1007/s10803-007-0517-7

Rohatgi, A. (2018). *WebPlotDigitizer* (Version 4.1) [Computer software]. https://automeris.io/WebPlotDigitizer

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Company.

Shriberg, L. D., Kwiatkowski, J., & Rasmussen, C. (1990). *Prosody-Voice Screening Profile (PVSP): Scoring forms and training materials*. Communication Skill Builders.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research, 44*(5), 1097–1115. https://doi.org/10.1044/1092-4388(2001/087)

Simmons, E. S., Paul, R., & Shic, F. (2016). Brief report: A mobile application to treat prosodic deficits in autism spectrum disorder and other communication impairments: A pilot study. *Journal of Autism and Developmental Disorders, 46*(1), 320–327. https://doi.org/10.1007/s10803-015-2573-8

Simmons, J. Q., & Baltaxe, C. (1975). Language patterns of adolescent autistics. *Journal of Autism and Childhood Schizophrenia, 5*(4), 333–351. https://doi.org/10.1007/BF01540680

Sparrow, S., Balla, D., & Cicchetti, D. (1984). *Vineland Adaptive Behavior Scales*. AGS.

Stangroom, J. (2018). *Effect size calculators*. Social Science Statistics. https://www.socscistatistics.com/effectsize/default3.aspx

Stangroom, J. (2020). *Effect size calculator for t-test*. Social Science Statistics. https://www.socscistatistics.com/effectsize/default3.aspx

Stevens, K. N., Nickerson, R. S., & Rollins, A. M. (1983). Suprasegmental and postural aspects of speech production and their effect on articulatory skills and intelligibility. In I. Hochberg, H. Levitt, & M. Osberger (Eds.), *Speech of the hearing impaired: Research, training and personnel preparation* (pp. 35–51). University Park Press.

Szczepek Reed, B. (2011). *Analysing conversation: An introduction to prosody*. Palgrave Macmillan. https://doi.org/10.1007/978-1-137-04514-0

Vannest, K. J., Parker, R. I., Gonen, O., & Adiguzel, T. (2016). *Single Case Research: Web based calculators for SCR analysis* (Version 2.0) [Computer software].

Wiklund, M. (2016). Interactional challenges in conversations with autistic preadolescents: The role of prosody and non-verbal communication in other-initiated repairs. *Journal of Pragmatics, 94,* 76–97. https://doi.org/10.1016/j.pragma.2016.01.008

Will, M. N., Currans, K., Smith, J., Weber, S., Duncan, A., Burton, J., Kroeger-Geoppinger, K., Miller, V., Stone, M.,

Mays, L., Luebrecht, A., Heeman, A., Erickson, C., & Anixt, J. (2018). Evidenced-based interventions for children with autism spectrum disorder. *Current Problems in Pediatric and Adolescent Health Care, 48*(10), 234–249. https://doi.org/10.1016/j.cppeds.2018.08.014

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., Brock, M. E., Plavnick, J. B., Fleury, V. P., & Schultz, T. R. (2013). *Evidence-based practices for children, youth, and young adults with autism spectrum disorder*. The University of North Carolina, Frank Porter Graham Development Institute, Autism Evidence-Based Practice Review Group.

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., Brock, M. E., Plavnick, J. B., Fleury, V. P., & Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disorders, 45*(7), 1951–1966. https://doi.org/10.1007/s10803-014-2351-z

Wynn, C. J., Borrie, S. A., & Sellers, T. P. (2018). Speech rate entrainment in children and adults with and without autism spectrum disorder. *American Journal of Speech-Language Pathology, 27*(3), 965–974. https://doi.org/10.1044/2018_AJSLP-17-0134