Stimulus-Stimulus Pairing to Increase Vocalizations in Children with Language Delays: a Review

M. Alice Shillingsburg[1,2] **· Diane L. Hollander**[3] **·
Rachel N. Yosick**[1] **· Crystal Bowen**[1] **·
Lori R. Muskat**[4]

Abstract Stimulus-stimulus pairing (SSP) is a procedure used to increase vocalizations in children with significant language delays. However, results from studies that have examined the effectiveness of SSP have been discrepant. The following review of the literature summarizes the results from 13 experiments published between 1996 and 2014 that used this procedure with children with language delays. Studies were analyzed across various participant and procedural variables, and an effect size estimate (nonoverlap of all pairs) was calculated for a portion of the participants in the studies reviewed. Results indicated an overall moderate intervention effect for SSP of speech sounds. Recommendations are provided for future researchers about information to report and potential avenues for future studies.

Keywords Language delays · Automatic reinforcement · Language development · Stimulus-stimulus pairing · Vocalizations · Effect size · Nonoverlap of all pairs

Children typically learn vocal verbal behavior as they interact with caregivers and others in their environment (Bloom 1979; Dawson 2008; Hart and Risley 1995; Rogers and Dawson 2010; Stone and Yoder 2001). Speech often begins with production of vowel sounds and babbling, which develops into syllable repetition and imitation of the intonation and speech patterns of those around them (Rogers and Dawson 2010). The influence of both contingent and noncontingent parent speech sounds on infant

---

✉ M. Alice Shillingsburg
alice.shillingsburg@choa.org

[1] Marcus Autism Center, 1920 Briarcliff Road NE, Atlanta, GA 30329, USA

[2] Emory University, Atlanta, GA, USA

[3] Georgia Neurobehavioral Associates, Canton, GA, USA

[4] Georgia School of Professional Psychology, Argosy University, Atlanta, GA, USA

vocalizations has been demonstrated in a number of studies (e.g., Bloom and Esposito 1975; Hart and Risley 1995; Pelaez et al. 2011; Poulson 1984; Rheingold et al. 1959; Tamis-LeMonda et al. 2001). A behavior-analytic conceptualization of language development posits that infant babbling develops as a result of both social and automatic reinforcement. Initially, newborns produce vocal sounds through the movement of the tongue, lips, and other organs of speech when engaging in reflexive behaviors such as crying and coughing. The production of these sounds strengthens vocal muscles, which enables the infant to produce more frequent and varied vocalizations (Bijou and Baer 1965; Schlinger 1995). During early infant development, caregivers frequently talk to their babies during daily activities such as holding, feeding, and diaper changing. Caregivers may socially reinforce infant vocalizations during these activities when they deliver reinforcing stimuli (e.g., attention, food) or remove aversive stimuli (e.g., wet diaper) contingent upon them. Automatic reinforcement appears to increase babbling via a two-step process. First, a caregiver's speech sounds are repeatedly paired with reinforcing stimuli. Through this pairing process, the caregiver's speech sounds may be established as conditioned reinforcers. Following pairing, when the infant babbles and the response product is acoustically similar to a sound that has become a conditioned reinforcer, the resulting auditory stimulation may automatically reinforce that specific way of vocalizing (Bijou and Baer 1965; Schlinger 1995). It should be noted that although the processes of social reinforcement, pairing, and automatic reinforcement have been described here as distinct, these processes likely occur concomitantly. For example, if an infant babbles "mama" and her mother picks her up and says, "mama" back to her, the infant's vocalizing "mama" may be reinforced by the mother's attention. In addition, the sound "mama" spoken by the mother is paired with a reinforcing stimulus (i.e., attention) and may be established as a conditioned reinforcer. The infant will be more likely to vocalize, "mama" in the future due to social reinforcement; when she does, the resulting sound may automatically reinforce the vocalization, as well.

In children with autism spectrum disorders, reinforcement, pairing, and stimulus control processes related to the development of social interaction and verbal behavior may be disrupted (Spradlin and Brady 1999). One mechanism by which speech development may be derailed is through diminished orientation to social stimuli. In typically developing infants, orienting to social stimuli develops early in infancy and is thought to be a requisite skill for joint attention, which has been correlated with language development (e.g., Dawson et al. 1998; Dawson et al. 2004; Toth et al. 2006). Research indicates that newborns recognize their mothers' voices (e.g., DeCasper and Fifer 1980) and show preference for looking at faces (e.g., Batki et al. 2000; Turati et al. 2002). For example, Dawson et al. (1998) examined social attention in children with autism, other developmental disabilities, and those who were typically developing. They found that children diagnosed with autism often did not orient to social stimuli, such as the speech sounds of others. Dawson et al. suggested that infants with autism likely encounter the same exposure to social stimuli (e.g., parental holding, talking) as their typically developing peers but do not engage actively in these interactions. Specifically, it is possible that social stimuli are not as salient or reinforcing to children with autism, and therefore, the speech sounds of others do not acquire typical reinforcing and/or discriminative functions. Consequently, targeted intervention to increase speech sounds in children with autism is often necessary.

Early and intensive behavioral intervention (EIBI) has been shown to result in substantial increases in vocal language in children with autism (Dawson 2008; Eldevik et al. 2009; Howard et al. 2005; Rogers and Vismara 2008; Sallows and Graupner 2005). Vocal models are typically used to prompt mands, tacts, and intraverbals (e.g., Barbera and Kubina 2005; Bourret et al. 2004; Watkins et al. 1989). If a child does not demonstrate an echoic (i.e., vocal imitation) repertoire, echoic training can be employed in which a therapist provides vocal models and differentially reinforces the child's successive approximations to the target sound (e.g., Butz and Hasazi 1973; Harris 1975; Hung 1976). However, shaping of any behavior, including echoic behavior, may be difficult when few instances of the behavior or invariable responses occur. For children with significant language deficits, especially those who produce few to no speech sounds, a stimulus-stimulus pairing (SSP) procedure may be used to increase the frequency of speech sounds.

As an intervention, SSP of speech sounds is designed to increase vocalizations through the same pairing process that is thought to lead to increases in infant babbling in typical development. Although there are variations of the SSP procedure, the essential feature is a therapist presenting a vocal sound along with the delivery of a high-preference social event or tangible item (hereafter referred to as preferred item). Through repeated pairings of therapist-emitted sounds with preferred items, the goal of SSP is to establish specific vocal sounds as conditioned reinforcers. Following pairing, when the child produces sounds that are the same or acoustically similar to sounds that have been paired, those vocalizations may be automatically reinforced (Sundberg et al. 1996). An increase in the frequency of specific sounds, even if temporary, would provide opportunities for reinforcement, shaping, discrimination training, and/or mand training.

Although SSP of speech sounds is used clinically to increase vocalizations in children with language delays, studies have reported this procedure to be effective with some participants and ineffective with others (Miliotis et al. 2012). Authors have suggested a number of participant and procedural variables that may contribute to these discrepant results, including participant age, pre-existing vocal repertoire, type of preferred items used, number of pairing trials, and number of experimenter-emitted sounds per trial (Stock et al. 2008). However, to date, the body of research on using SSP to increase vocalizations has not yet been systematically reviewed to inform best practice. Therefore, the purpose of the current review is to provide a systematic quantitative analysis of studies that used SSP to increase vocalizations in children with language delays. To quantify effectiveness of SSP across studies, we used the nonoverlap of all pairs method (NAP; Parker and Vannest 2009), a nonparametric effect size calculation used in single-case research. Recommendations are provided for future researchers about information to report and potential avenues for future studies.

Method

Inclusion and Exclusion Criteria

A computerized multi-database literature search was conducted using PubMed Central and PsycINFO databases. The terms *stimulus-stimulus pairing*, *vocalizations*,

automatic reinforcement, *autism*, and *language development* were entered into the keyword fields. Searches were limited to articles with publication dates ranging from 1996 to 2014. Studies included in this review were limited to published, peer-reviewed investigations that used SSP to increase vocalizations of children with language delays. As such, studies that investigated the use of SSP to establish books, toys, worksheets, or other items as conditioned reinforcers or evaluated effects on responses other than vocalizations were excluded.

Dimensions and Coding

Data were collected on the journal and number of participants included in each study. Studies were analyzed and coded according to the participant characteristics and procedural variations described below.

Participant Characteristics

Age and Gender Data were collected on the reported age (years, months) and gender of all participants. Further, participants were coded as over 5 years old or 5 years old and younger.

Diagnosis Data were collected on the diagnosis of each participant as reported by the authors and the method by which the diagnosis was determined.

Language Skills Based on the descriptive information available, we characterized participants who were reported to vocally mand/request, tact/label, or respond with intraverbals as displaying functional language skills. Participants who emitted sounds or echoics only were classified as not displaying functional language skills. The method of language assessment was also recorded.

SSP Procedural Variations

Although all studies reviewed used SSP to increase vocalizations in individuals with language delays, the treatment strategies varied along several key characteristics. As such, six procedural variations were analyzed and coded for evaluation.

Target Sound Data were recorded on whether the sounds targeted for pairing were novel or displayed by the participant prior to the study. *Novel sounds* were identified as those sounds not observed during baseline or pre-assessment observations. *Current sounds* were observed during baseline and/or pre-assessment observations. Data were also collected on the method for selecting target sounds.

Number of Experimenter-Emitted Sounds per Pairing Data were recorded on the number of times the experimenter emitted the sound during each trial.

Type of Pairing The timing of preferred item delivery in relation to the presentation of the experimenter-emitted sound was also a variable of interest. Pairing was coded as *simultaneous* (i.e., the preferred item and the sound were delivered at the same time), *delay* (i.e., the preferred item was delivered during the sound), *trace* (i.e., the preferred

item was delivered after the entire sound), or *discrimination training* (i.e., the preferred item is delivered contingent upon an arbitrary response in the presence of the sound but not in its absence).

Number of Pairings per Minute Data were collected on the number of pairings per minute used in each study. If the exact number of pairings per minute were not reported, we estimated the number of pairings per minute based on the information provided by examining the reported intertrial intervals, number of trials per session, duration of sessions, reinforcement intervals, duration of sound production, and delays for adventitious reinforcement. When the information provided indicated a variable number of pairings per minute, we estimated the minimum and maximum amount of time for each trial, resulting in a range of pairings per minute.

Adventitious Reinforcement If the authors stated that the preferred item was delayed or withheld following the emission of the target sound from the participant, the procedure was coded as controlling for adventitious reinforcement.

Type of Preferred Items Data were collected regarding the method by which preferred items were selected and included surveys (i.e., report of preference), informal observations (i.e., unsystematic observations), and stimulus preference assessments. Additionally, the selected preferred items were coded as social interactions, edibles, toys, or a combination of these.

Effect Size To quantify results across studies, nonoverlap of all pairs (NAP) was selected as a measure of intervention effectiveness. NAP is an effect size estimate that has demonstrated utility in summarizing intervention effectiveness across large samples of single-case data (Parker and Vannest 2009). NAP requires few data assumptions, can be efficiently calculated by hand, and is strongly correlated with visual analysis judgments. Unlike more commonly used effect sizes such as Cohen's *d*, NAP does not rely on means or medians, and instead examines the location of the entire score distribution (Parker and Vannest 2009). The NAP probability score is the percentage of data that shows improvement from one phase to another and normally ranges from 0.5 to 1 and (Vannest et al. 2013). For single-case design studies, NAP can be calculated by hand from a graph for each participant and then averaged across participants to calculate an overall effect size for all participants in the sample. Scores are reported as *small*/*weak* (0–0.65), *medium*/*moderate* (0.66–0.92), and *large*/*strong* (0.93–1.0) (Parker and Vannest 2009).

   NAP was calculated for the participants from a portion of the studies included in the review based on the following exclusion criteria. Studies were excluded if the authors reported only cumulative data (Stock et al. 2008; Sundberg et al. 1996; Yoon and Bennett 2000; Yoon and Feliciano 2007) or did not report a baseline phase (Lepper et al. 2013) because analysis of overlap between baseline and treatment conditions could not be conducted. The procedure outlined in Parker and Vannest (2009) for hand calculation of NAP was followed. Within each target sound SSP graph, all baseline data points were compared with all treatment data points, and the total number of possible comparisons was calculated (number of baseline data points multiplied by number of treatment data

points). If a baseline data point value was higher than a treatment data point value, that comparison was marked as an "overlap." If a baseline data point value was equal to a treatment data point value, that comparison was marked as a "tie." The total number of overlaps and ties were counted for each graph. These totals were then used in the following equation to solve for NAP for each target sound:

$$\text{NAP} = \frac{\text{Number of total BL/TX comparisons} - [\text{Number of comparisons that overlap} + (.5)\text{Number of comparisons that tie}]}{\text{Number of total BL/TX comparisons}}$$

The average NAP of the sample ($n = 35$) was then calculated by summing each individual NAP score and dividing by 35. NAP calculations for each graph were calculated by two independent researchers. The intraclass correlation coefficient (ICC) was calculated as a measure of interobserver agreement and showed perfect agreement (ICC $= 1.0$, $p < 0.0001$).

Results

Of the studies identified in the literature search, one was excluded because all participants were typically developing children without language delays (i.e., Smith et al. 1996). Further, one typically developing participant from Sundberg et al. (1996) was not included in our review, leaving a total of 13 studies and 39 participants. All of the studies were published in the *Journal of Applied Behavior Analysis* or *The Analysis of Verbal Behavior*, and most studies included 2 to 3 participants. See Table 1 for an overview of the characteristics of the studies and participants included in the review.

Participant Characteristics

The studies reviewed included participants that varied with regard to age, gender, diagnosis, and presence of functional language.

Age The average age of participants across studies was 3 years, 7 months (range, 1–8 years). Most studies evaluated SSP with preschool children, with the exception of five studies that included school-aged children (i.e., Esch et al. 2009; Esch et al. 2005; Miguel et al. 2002; Miliotis et al. 2012; Rader et al. 2014) for a total of seven participants over the age of 5 years.

Gender Many studies included both males and females as participants; however, across all studies, there was more than double the number of males than females. Gender was not reported for two participants included in Yoon and Bennett (2000).

Diagnosis The majority of participants included in the studies were diagnosed with autism (69.2 %). Other diagnoses reported included educational delay (ED, 15.4 %),

Table 1 Study and participant characteristics

Number (percentage) of studies		
	Review n = 13	NAP analysis n = 8
Journal		
The Analysis of Verbal Behavior	10 (76.9)	6 (75)
Journal of Applied Behavior Analysis	3 (23.1)	2 (25)
Number of participants		
1	1 (7.7)	1 (12.5)
2	3 (23.1)	3 (37.5)
3	6 (46.2)	4 (50)
4	2 (15.4)	–
6	1 (7.7)	–
Number (percentage) of participants		
	Review n = 39	NAP analysis n = 19
Gender		
Male	25 (64.1)	11 (57.9)
Female	12 (30.8)	8 (42.1)
Not reported	2 (5.1)	–
Age		
≤5 years	32 (82.1)	12 (63.2)
>5 years	7 (17.9)	7 (36.8)
Diagnosis		
Autism	27 (69.2)	19 (100)
Educational delay	6 (15.4)	–
Developmental delay	5 (12.8)	–
Intellectual disability with visual impairment	1 (2.6)	–
Functional language		
Yes	11 (28.2)	4 (21.1)
No	28 (71.8)	15 (78.9)

developmental delay (DD, 12.8 %), and intellectual disability with visual impairment (ID/VI, 2.6 %). Although some studies indicated that the diagnosis was made by a professional, no studies included the type of assessment used to inform the diagnosis or the age at which the diagnosis was made.

Language Skills Although all studies provided a description of each participant's pre-existing language skills, there was not a uniform assessment used across studies. The Behavioral Language Assessment (BLA; Sundberg and Partington 1998), an informant assessment that evaluates basic language and related skills in children with limited verbal abilities, was used to identify the language skills of 20 participants across 8 studies (i.e., Carroll and Klatt 2008; Esch et al. 2005; Esch et al. 2009; Miguel et al. 2002; Miliotis et al. 2012; Normand and Knoll 2006; Rader et al. 2014; Stock et al. 2008). However, there were inconsistencies in how the scores were reported. Some

authors reported an overall BLA level (i.e., Carroll and Klatt 2008; Miguel et al. 2002; Miliotis et al. 2012; Normand and Knoll 2006; Rader et al. 2014), while others reported some of the individual scale levels (i.e., Esch et al. 2005; Stock et al. 2008) reported that the BLA was used and provided descriptions of participant verbal repertoires; however, they did not report levels. In addition to the BLA, Esch et al. (2005, 2009) conducted the Kaufman Speech Praxis Test (KSPT; Kaufman 1995), the Peabody Picture Vocabulary Test-III (PPVT-III; Dunn et al. 1997) and the Receptive-Expressive Language Test, Third Edition (REEL; Bzoch et al. 2003). Lepper et al. (2013), Miliotis et al. (2012), and Rader et al. (2014) reported scores from the Early Echoic Skills Assessment (EESA; Esch 2008). Lepper et al. (2013) used an informal pre-experimental observation to obtain information on pre-intervention vocal skills. Sundberg et al. (1996), Yoon and Bennett (2000), Yoon and Feliciano (2007), and Ward et al. (2007) described participant verbal skills but did not indicate if the description was based on informal observations or an assessment tool.

Based on the description provided for each participant, a total of 28 participants were classified as not having functional language skills, and 11 participants were classified as having functional language skills. It should be noted that the participants in the functional language group had varying degrees of language abilities (i.e., some had a few vocal mands while others could emit hundreds of mands, tacts, and intraverbals). The functional language group was not further differentiated because this could not be practically accomplished based on the information provided in the studies.

SSP Procedural Variations

Results from the coding of target sound, number of experimenter-emitted sounds per pairing, type of pairing procedure, number of pairings per minute, control for adventitious reinforcement, and the type of preferred item paired are presented in Table 2.

Target Sound Of the studies reviewed, a total of 17 participants were exposed to novel sound pairing conditions (Lepper et al. 2013; et al. 1996; Yoon and Bennett 2000; Yoon and Feliciano 2007), and 22 were exposed to current (in-repertoire) sound pairing conditions (Carroll and Klatt 2008; Esch et al. 2005; Esch et al. 2009; Miguel et al. 2002; Miliotis et al. 2012; Normand and Knoll 2006; Rader et al. 2014; Stock et al. 2008; Ward et al. 2007); Carroll and Klatt (2008) used one novel target and one current target. The manner in which sounds were selected was described in nine studies. Three of the studies indicated that they selected sounds with the lowest frequency of those observed during baseline or pre-assessment observation (i.e., Carroll and Klatt 2008; Miguel et al. 2002; Stock et al. 2008). Three studies indicated that target sounds were selected if they occurred during a specified range of intervals during baseline or pre-assessment observations. Specifically, Esch and colleagues (2009) selected sounds that occurred during 10–25 % of intervals or if all sounds occurred in fewer than 10 % of intervals; sounds that occurred at all were selected. Miliotis et al. (2012) selected sounds that occurred during 1–5 % of intervals during baseline or pre-assessment observations, and Rader et al. (2014) selected sounds that occurred in 25 % or fewer of intervals during baseline. Two studies indicated that low

Table 2 SSP procedural variations

Study	Target sound	# of sounds per pairing	# of pairings per min	Type of pairing	Adventitious reinforcement control	Item paired
Sundberg et al. (1996)	Novel	1	2.5	Trace	No	Social
			7.5			
			15			
Yoon and Bennett (2000)	Novel	1	12	Simultaneous	No	Social
Miguel et al. (2002)	Current	5	1.7-4*	Delay	Yes	Edible
Esch et al. (2005)	Current	3	≤ 7.5*	Trace	No	Edible/Toy
Normand and Knoll (2006)	Current	7	1.6-8.6*	Delay	Yes	Not stated
Yoon and Feliciano (2007)	Novel	1	12	Simultaneous	No	Social/Edible/Toy
Ward et al. (2007)	Current	1	10	Delay/Trace	No	Edible/Toy
						Social/Edible/Toy
Stock et al. (2008)	Current	5	3*	Delay	Yes	Edible
Carroll and Klatt (2008)	Current/Novel	5	1.3-2.4*	Delay	Yes	Edible/Toy
Esch et al. (2009)	Current	3	<1-2.3*	Trace	Yes	Edible/Toy
Miliotis et al. (2012)	Current	1 or 3	1.4-10*	Delay	Yes	Edible/Toy
Lepper et al. (2013)	Novel	3	1-2.3*	Delay/ Discrimination Training	Yes	Edible/Toy
Rader et al. (2014)	Current	3	<1-2*	Trace	Yes	Edible/Toy

[a] Pairings per minute were calculated based on reported procedures

occurring sounds were selected but did not provide specific criteria regarding the occurrence prior to selection (i.e., Esch et al. 2005; Normand and Knoll 2006). One study, Ward et al. (2007), chose sounds that occurred with the highest frequency during observation sessions.

Number of Experimenter-Emitted Sounds per Pairing Across studies, there were considerable variations in the number of times the experimenter emitted the target sound per trial. In four studies, the experimenter emitted the target sound once with the delivery of the preferred item (i.e., Sundberg et al. 1996; Ward et al. 2007; Yoon and Bennett 2000; Yoon and Feliciano 2007). In four studies, the target sound was emitted three times by the experimenter (i.e., Esch et al. 2005; Esch et al. 2009; Lepper et al. 2013; Rader et al. 2014). In one study (Miliotis et al. 2012), the target sound was emitted one time per trial for one target sound and three times per trial for a second target sound. In three studies, the target sound was emitted by the experimenter five times during each pairing trial (i.e., Carroll and Klatt 2008; Miguel et al. 2002; Stock et al. 2008). Finally, Normand and Knoll (2006) presented the target sound seven times per trial.

Type of Pairing Procedure Though not specifically labeled as a particular type of pairing in the articles we reviewed, with the exception of Miliotis et al. (2012) and Lepper et al. (2013), the author description of the pairing process allowed us to classify the type of pairing involved in each study. Four studies described the pairing procedure as presenting the sound followed immediately by presentation of the preferred item (i.e., Esch et al. 2005; Esch et al. 2009; Rader et al. 2014; Sundberg et al. 1996) suggesting the use of trace pairing. Miliotis et al. specifically identified the procedure used as delay conditioning and indicated that delivery of the preferred stimulus overlapped the model. Five other studies also described procedures that fit the description of delay conditioning. Lepper et al. (2013) compared two types of pairing procedures. In one, they presented two sounds and then presented the preferred item simultaneously with the third and final sound; the other procedure was discrimination training. Although both procedures were similarly effective, all participants preferred the discrimination training procedure in a treatment preference evaluation. Miguel et al. (2002 and Carroll and Klatt (2008) repeated the sound a total of five times and presented the preferred item after the third but before the last sound was emitted by the experimenter. Stock et al. (2008) also repeated the sound five times, but presented the preferred item between the second and fifth emission. Normand and Knoll (2006) presented the target sound seven times and delivered the preferred item after the fourth emission of the target sound. Two studies included a description of the pairing procedure that fit a description of simultaneous conditioning. Yoon and Bennett (2000) and Yoon and Feliciano (2007) described the pairing procedure as the experimenter emitting the sound and simultaneously presenting the preferred event. One study appears to have included a combination of pairing types. Ward et al. (2007) initially presented pairing trials that fit the description of delay conditioning but changed to trace conditioning. The authors described presenting the sound with the preferred item presented half a second after the sound was initiated. Over time, they inserted a 2-s delay between sound presentation and delivery of the preferred item to allow for direct reinforcement of an echoic response.

Number of Pairings per Minute The rate of experimenter-delivered pairings during treatment sessions varied across studies. Four studies reported the approximated or exact number of pairings per minute implemented (i.e., Sundberg et al. 1996; Ward et al. 2007; Yoon and Bennett 2000; Yoon and Feliciano 2007). The remaining studies did not report the number of pairings per minute but provided the number of trials conducted and some information on the length of intertrial interval, length of consumption interval, number of sounds produced per trial, time between sounds, and/or delay when controlling for adventitious reinforcement. In some studies, a range of session duration was also provided. Across all studies, the rate of pairings ranged from less than 1 per minute to 15 per minute.

Control for Adventitious Reinforcement Of the 39 participants included in the studies reviewed, approximately half (48 %) received procedures that controlled for adventitious reinforcement and approximately half (52 %) did not. Two different types of control procedures were used when the participant emitted a target sound during the control or SSP session. Stock et al. (2008) stated that a preferred item was not delivered contingent upon target vocalizations but did not specify for how long suggesting the next scheduled trial was completed regardless of when the sound was emitted. Miguel et al. (2002), Esch et al. (2009), Carroll and Klatt (2008), Miliotis et al. (2012), and Rader et al. (2014) used a 20-s delay to reinforcement, and Normand and Knoll (2006) used a 30-s delay to reinforcement. In these studies, the experimenter presented the sound, and if the participant emitted the sound, the preferred item was withheld for the specified time interval before the preferred item was delivered or a new trial was initiated.

Type of Preferred Item Paired Preferred items paired with sounds were identified in a variety of ways across studies and included informal observation, surveys, stimulus preference assessment, and a combination of methods. Four studies (i.e., Sundberg et al. 1996; Ward et al. 2007; Yoon and Bennett 2000; Yoon and Feliciano 2007) did not report conducting preference assessments but used items that had previously been observed to function as reinforcers.

Four studies reported using both a structured survey and a subsequent stimulus preference assessment (i.e., Esch et al. 2005, 2009; Miguel et al. 2002; Stock et al. 2008). The survey used in these studies was the Reinforcement Assessment for Individuals with Severe Disabilities (RAISD; Fisher et al. 1996), a structured interview that can be used with multiple informants to generate a list of preferred stimuli. Following the RAISD, these researchers reported using variations of a multiple-stimulus preference assessment. Two studies (i.e., Miguel et al. 2002; Stock et al. 2008) reported presenting items from the RAISD in a single array to the participants at the beginning of each experimental condition. The first item chosen and consumed by the participant was used during the following condition. Esch et al. (2005, 2009) used a multiple-stimulus without replacement preference assessment (MSWO; DeLeon and Iwata 1996) and presented three separate arrays to the participants to identify the five (Esch et al. 2005) or three (Esch et al. 2009) highest ranked items. Each day, the items that were touched or smiled at the most during a 1-min sampling period were used for sessions that day and were randomly rotated during sessions. Similarly, Rader et al. (2014) used an interview to identify items to include in paired-stimulus preference assessment (Fisher et al. 1992) and presented the three highest ranked items prior to each session. The item selected was used during pairing trials for that session.

Carroll and Klatt (2008) used both engagement- and selection-based stimulus preference assessments. Preferred stimuli were first selected based on a 45-min observation of participants interacting with edibles and toys. The five items with the highest percentage of 1-min intervals of engagement were then presented prior to each session; the first item touched was selected for pairing for that session. Normand and Knoll (2006) and Miliotis et al. (2012) also used an MSWO preference assessment; however, they did not indicate how these items were selected. Before the first session each day, Miliotis et al. identified the three highest ranked items with an MSWO preference assessment and subsequently rotated those items during pairing trials on that day. Lepper et al. (2013) reported that preferred items were selected based on previous preference assessments but did not specify the type.

A variety of types of stimuli were paired with sounds in the 13 studies reviewed. Preferred social interactions alone, consisting of tickles, praise, picking up, pushing on a cart, and hand swinging, were used in two SSP studies (i.e., Sundberg et al. 1996; Yoon and Bennett 2000). Edibles only were used by Miguel et al. (2002), Stock et al. (2008), and Lepper et al. (2013). Edibles and social interactions were used in the Ward et al. (2007) and Yoon and Feliciano (2007) studies. Edibles and toys were used by Ward et al. (2007), Carroll and Klatt (2008), Esch et al. (2005, 2009), Miliotis et al. (2012, and Rader et al. (2014. One participant in Lepper et al. (2013) received only toys during pairings, and Normand and Knoll (2006) did not specify what types of preferred items were used.

Effectiveness of SSP

Eight of the 13 studies (19 participants) in the overall review were included in the calculation of an effect size estimate using NAP. Because more than one sound was targeted for some participants, the NAP was hand-calculated for each target sound for each participant within those studies ($n = 35$). See Table 1 for an overview of the characteristics of the studies and participants included in the NAP analysis. Of the 8 studies included in the effect size calculation, the majority (75 %) was published in *The Analysis of Verbal Behavior* and most (88 %) included 2 to 3 participants. The majority of participants were male (58 %), all were diagnosed with autism, and most (79 %) did not have functional language skills. The average age among the participants included for effect size calculation was 4 years, 6 months (range, 1–8 years).

Table 3 shows effect sizes for the 35 sounds targeted in 8 studies using SSP to increase vocalizations. The average effect size was 0.71 (SD = 0.20; 95 % CI [0.64–0.76]), which corresponds to a moderate treatment effect according to NAP interpretive guidelines (Parker and Vannest 2009). Of the 35 evaluations for which effect sizes were calculated, 12 (34 %) were classified as showing a weak effect, 17 (49 %) were classified as showing a moderate effect, and 6 (17 %) were classified as showing a strong effect of SSP. The percentage of sounds showing strong, moderate, and weak effect sizes across the different variables discussed are shown in Table 4. A higher percentage of children who were younger (5 years or below) showed moderate to strong effects compared to those over age 5. Additionally, a higher percentage of children without functional language showed moderate to strong effects compared to those with functional language. In fact, none of the children classified as having functional language showed a strong effect of SSP. All but one of the sounds included

Table 3 NAP effect size across sounds, participants, and studies

Study	Participant	Sound	Effect size	Interpretation
Miguel et al. (2002)	1	*ee*	1.00	Strong
		uh	0.95	Strong
	2	*bah*	0.90	Moderate
		oo	0.80	Moderate
	3	*dah*	0.61	Weak
		ee	0.66	Moderate
Esch et al. (2005)	1	*ba*	0.50	Weak
		i	0.43	Weak
	2	*si*	0.17	Weak
		bi	0.54	Weak
	3	*i*	0.50	Weak
		a	0.50	Weak
Normand and Knoll (2006)	1	*ah*	0.78	Moderate
		ee	0.44	Weak
Ward et al. (2007)	1	*ahmm*	0.92	Moderate
		ee	0.72	Moderate
	2	*doh*	0.77	Moderate
Carroll and Klatt (2008)	1	*ts*	0.94	Strong
		buh	0.50	Weak
	2	*m*	0.56	Weak
Esch et al. (2009)	1	*beh*	0.69	Moderate
		oo	0.66	Moderate
	2	*aypayk*	0.91	Moderate
		sheba	0.76	Moderate
	3	*reeklo*	0.83	Moderate
		tebba	0.68	Moderate
Miliotis et al. (2012)	1	*ng*	0.97	Strong
		l	0.66	Moderate
	2	*t*	0.99	Strong
		w	0.97	Strong
Rader et al. (2014)	1	*b*	0.81	Moderate
	2	*m*	0.79	Moderate
	3	*m*	0.88	Moderate
Average NAP effect size			0.71	Moderate

in the effect size calculations were current sounds. Thus, an examination of differences in effect size for novel or in-repertoire sounds is not possible. Those who received delay conditioning during the SSP procedure, those for whom edibles only were used, and participants who received procedures to control for adventitious reinforcement showed the highest percentages of moderate to strong effect sizes. The number of experimenter-emitted sounds per pairing did not appear to affect the effect size calculations.

Table 4 Percentage of sounds in each category with strong, moderate, and weak NAP effect sizes

	Percentage (number) of sounds		
	Strong effect	Moderate effect	Weak effect
Age			
≤5 years (n = 21)	14.3 (3)	66.7 (14)	19.0 (4)
>5 years (n = 14)	21.4 (3)	21.4 (3)	57.1 (8)
Functional language			
Yes (n = 7)	–	57.1 (4)	42.9 (3)
No (n = 28)	21.4 (6)	46.4 (13)	32.1 (9)
Adventitious reinforcement control			
Yes (n = 24)	25.0 (6)	58.3 (14)	16.7 (4)
No (n = 11)	–	27.3 (3)	72.7 (8)
Type of reinforcer[a]			
Edibles only (n = 6)	33.3 (2)	50.0 (3)	16.7 (1)
Edibles plus toys/social (n = 27)	14.8 (4)	48.1 (13)	37.0 (10)
Number of sounds per pairing			
1–3 (n = 24)	12.5 (3)	54.2 (13)	33.3 (8)
4–7 (n = 11)	27.3 (3)	36.4 (4)	36.4 (4)
Number of pairings per minute			
1–4 (n = 18)	16.7 (3)	66.6 (12)	16.7 (3)
5 and above (n = 11)	–	27.3 (3)	72.7 (8)
Variable[b] (n = 6)	50 (3)	33.3 (2)	16.7 (1)
Type of pairing			
Delay (n = 15)	40 (6)	33.3 (5)	26.7 (4)
Trace (n = 17)	–	52.9 (9)	47.1 (8)
Delay/trace (n = 3)	–	100 (3)	–

[a] Unspecified sounds are not included

[b] The estimated range of pairings ranged from <1 to >5 per minute

Interestingly, a higher percentage of weak effects were observed when the number of pairings per minute was 5 and above.

Discussion

The present article reviewed and summarized the current literature evaluating SSP to increase vocalizations in children with language delays. To quantify treatment outcomes, an effect size estimate was calculated for a portion of the studies included in the review. Overall, the research indicates a moderate intervention effect for individuals with language delays. However, it should be noted that approximately one third of participants showed only weak effects. Participants who were younger, received delay pairing, and had controls for adventitious reinforcement tended to have higher

proportions of positive treatment results. However, firm conclusions are difficult to draw because numerous procedural and participant variables overlapped, making it difficult to discern which variable(s) produced the results. Results of this review indicate that there is currently not a strong research base to guide clinicians in making decisions about specific procedures because of an overall lack of studies and differences in participants included, information reported, and procedures employed across studies. A summary of the findings on participant and procedural variables follows, with recommendations for researchers about information to report in published research and potential research questions that remain to be answered.

Most of the studies in our literature review included participants who were male, aged 5 and under, diagnosed with autism, and did not demonstrate pre-existing functional language skills. The finding that the majority of studies included toddlers and preschoolers suggests that researchers may deem the procedure more effective with younger children. However, it also highlights a gap in our understanding of the use of the procedure. Thus, future research on the efficacy of SSP with children over age 5 is warranted. Our analysis also indicated that older children did not appear to benefit as much from the SSP procedure as younger children. It is not surprising that younger children may potentially respond better to SSP given what we know about the benefits of early intervention in general (Warren et al. 2011); however, it should be noted that 43 % of participants over 5 years old still achieved moderate to strong intervention effects. Future research should consider investigating SSP with larger samples of both younger and older children with significant language delays.

Studies varied widely in terms of reporting diagnosis, level of functioning, and language skills of participants. In future studies, we recommend that authors include a more comprehensive characterization of participants and include high-quality measures for assessing and diagnosing participants. A diagnostic assessment battery should entail an evaluation of social behavior, language and communication, adaptive skills, and stereotypical behaviors. Assessment tools such as the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2012) should be considered to confirm or rule out a diagnosis of autism, whereas assessments of cognitive development, adaptive skills, and language skills, may shed light on the characteristics of children that are associated with better response to SSP.

Participant pre-existing functional language was a variable of interest because it has been suggested that SSP might be differentially effective with children with severely impaired vocal behavior (e.g., Yoon and Feliciano 2007). In the current review, only 7 of the 35 target sounds included in the NAP calculation were with participants classified as displaying functional language. Therefore, it was difficult to draw conclusions about the effectiveness of SSP in relation to this characteristic. However, the preliminary data from this review indicate that SSP did not result in a strong effect for any of these participants, whereas 21 % of the targeted sounds in participants classified as not having functional vocal language showed strong effects of SSP. These results should be interpreted with caution but point to a potentially interesting area of future research and may shed additional light on which children may benefit most from the procedure. It is clear from the review that the results of SSP vary, and participant characteristics may play an important role in the efficacy of the procedure. For researchers to be able to refine the procedure and for clinicians to effectively

implement the procedure, research should begin to identify specific characteristics of research participants.

Most of the studies targeted sounds that were current sounds for participants but only occurred as babbling or stereotypy, possibly maintained by automatic reinforcement. Some studies further specified how frequently the target sounds were heard during pre-assessment or baseline observations. All of the sounds except one included in the NAP evaluation were current, limiting the discussion on novel versus in-repertoire sounds. No direct comparisons of the procedure with in-repertoire versus novel sounds or low frequency versus high-frequency target sounds have been conducted. These specific evaluations could be conducted in future research.

The majority of studies included procedures in which one to three sounds per pairing were emitted by the experimenter, with relatively fewer having investigated four or more sounds per pairing. Only one study, Miliotis et al. (2012), included a direct comparison of the number of sounds per pairing. These authors found that increases in vocalizations were greater when the target sounds were presented one time per pairing than when presented three times per pairing. Stock et al. (2008) suggested that pairing ratios that include many presentations of the target sound per delivery of the preferred item essentially result in the target sound being unpaired more often than paired. However, in our NAP analysis, there did not seem to be a relationship between the number of experimenter-emitted sounds per pairing and strength of effect. Future research should examine if this variable affects the results of SSP by investigating additional direct comparisons.

From the descriptions of procedures for each participant, we attempted to classify the type of pairing used. Though delay, trace, simultaneous pairing, and discrimination training procedures have all been used, most studies have used delay pairing, and stronger effects were found when delay pairing was employed. The only study to compare pairing procedures was Lepper et al. (2013), who found discrimination training and delay pairing to be similarly effective, but the former more preferred. Potential advantages of discrimination training could be that it requires the participant to demonstrate observation of relevant features of the neutral stimulus and provides a guide for the number of pairing trials to conduct (i.e., after differential responding to the S^D and s-delta is demonstrated). A third variation of pairing is response-stimulus pairing, in which the neutral stimulus and preferred item are delivered together following a response. Recently, Dozier et al. (2012) found response-stimulus pairing to be more effective than SSP in establishing praise as a reinforcer with adults with developmental disabilities. Although not reported to have been evaluated in the SSP of speech sounds literature, studies that have included observing responses (e.g., looking) prior to pairing might be described as employing response-stimulus pairing. Given that different pairing methods may achieve different results, it is recommended that future studies specify the type of pairing procedure used, specify the rationale for its use, and consider conducting comparisons of pairing type.

The number of pairings per minute varied greatly both within and across studies. This variable was the least likely to be directly reported by authors, requiring us to estimate pairings per minute in 9 of the 13 studies reviewed. In our NAP analysis, weaker effects were achieved when more pairings per minute were conducted. However, it is important to note that these figures were estimated based on the

descriptions of procedures provided by the authors for all of the studies included in the NAP calculation. It is recommended that future researchers record and report the number of pairings per minute conducted. Direct comparisons of varying pairings per minute may help identify best practices. Alternatively, the pairings per minute may not be as important as ensuring that a pairing trial is conducted only when there is an establishing operation at strength for the preferred item being paired. In other words, fewer pairings per minute conducted at the precise moment the items are preferred may be more effective than conducting more pairings at times when the item being paired is not preferred.

A somewhat unexpected finding was that a higher percentage of weak effects were shown when procedures to control for adventitious reinforcement were not included. We had hypothesized that studies that did *not* control for adventitious reinforcement might yield stronger treatment effects due to the effects of direct reinforcement of vocalizations and were surprised to find that participants achieved better outcomes when direct reinforcement was prevented. This finding could be at least partially explained by the fact that the majority (75 %) of participants who received control for adventitious reinforcement during their intervention was also younger (5 years or younger). Because the effects of SSP were stronger for younger participants, it could be that age influenced the higher percentage of moderate to strong treatment effects and the fact that these younger participants also happened to receive adventitious reinforcement control was artifactual. Again, this highlights the difficulty of interpreting the results of this review and calls for more research specifically examining and controlling for some of these variables.

All of the studies reported procedures for selecting preferred items; however, the methods and details provided varied greatly. We recommend that future researchers conduct stimulus preference assessments with items indicated by a caregiver survey and conduct a brief preference assessment immediately prior to SSP sessions to increase the effectiveness of pairing. At least one study (i.e. Esch et al. 2005) indicated that during some pairings, the participant shook his head, "no" and pushed the item away, potentially compromising the pairing effect. Future research may consider including procedures within sessions to ensure pairing trials are conducted only when the paired item is preferred. It might be beneficial to look for other behaviors within the session that indicate there is a current establishing operation for the preferred item, such as a gesture, and time the presentation of the pairing trial to correspond with that indication. For example, a pre-session preference assessment might indicate that an edible and a toy are preferred. Those two items would then be included in the session. However, instead of then presenting pairing trials at pre-specified intervals, pairing trials could be presented every time the child reaches for or gazes at one of the items. Researchers might also consider conducting reinforcer assessments to demonstrate reinforcing effectiveness of stimuli. Another variable that should be considered is whether the paired preferred item competes with or is incompatible with sound production from the participant. For example, preferred toys that make noise (e.g., music toys or movies) or consumption of edibles may actually impede vocalizations. However, our NAP analysis indicated that stronger effects were found when edibles alone were used during SSP. Social interactions such as tickles may not impede vocalizations, but social interactions such as the therapist singing potentially could. This variable was not examined in the current review because the relevant information

was typically not reported in studies. However, future research might consider these issues when selecting items for pairing or specifically seek to examine their effects.

Other procedures that have been used in only a small number of studies also warrant discussion. For example, some studies interspersed nontarget, unpaired (i.e., S−) sounds with target, paired (S+) sounds within sessions (i.e., Esch et al. 2009; Miliotis et al. 2012; Radar et al. 2014). Lepper et al. (2013) included a nontarget, unpaired sound and a control for presentation of preferred items (i.e., preferred items were delivered but delayed 20 s after adult sound emission). Interspersal of S− trials may provide benefits of control for exposure to a sound, as well as making the features of the S+ more salient to the learner. Future research should continue to employ these controls as well as rigorous experimental designs to enhance the conclusions that can be drawn. Last, the manner in which the sounds are emitted by the experimenter may be an important variable. Studies have shown that infants attend to motherese (i.e., sing-song speech) more than neutral or monotone vocalizations from others (Cooper and Aslin 1990). Our review of SSP found that only four studies reported details on the quality or nature of sound production. Specifically, three studies reported presenting sounds with different pitches and intonations (i.e., Sundberg et al. 1996) or exaggerated prosodic patterns (i.e., motherese; Esch et al. 2009; Miliotis et al. 2012) and one described sound production as monotone (Stock et al. 2008). Future research should include information regarding the quality of sound production during pairing trials.

Given the results of our review, it is difficult to determine for whom and under what conditions SSP is most effective. Our analysis was hampered at times by limited details provided by previous studies, as well as by the overall dearth of research that has been conducted on this procedure. More research is needed to clarify for which participants the procedure is most effective. Further, future research should examine variations of SSP (or other types of pairing procedures) to make recommendations about how pairing might be optimally conducted. Though clinicians currently use SSP, there are few, if any, recommendations that can be provided from the literature base to those wishing to employ it. To date, there have been no randomized, controlled trials of the SSP procedure to examine its effectiveness in a larger, better characterized sample of participants. This sort of investigation is warranted given the difficulty in deciphering which specific procedures are likely to be effective with which populations. Additionally, all studies included in the literature review were published in the *Journal of Applied Behavior Analysis* and *The Analysis of Verbal Behavior*, premier journals in the field of applied behavior analysis. However, because of the potential implications for widespread knowledge of SSP, future researchers should consider additional outlets for their studies to bring more attention to this clinical procedure and potentially spur additional interest in other researchers. Finally, it is important to note that, to date, studies investigating SSP to increase vocalizations have not demonstrated that the increases observed can be attributed to their functioning as conditioned reinforcers. To demonstrate this, following pairing trials, vocalizations could be presented contingent upon a response to demonstrate a reinforcer effect. Petursdottir et al. (2011) examined the effects of SSP on the preferences for paired and unpaired speech sounds via button presses that produced auditory stimuli. They found that SSP did not reliably result in differential selection of the paired speech sounds, suggesting future research is needed to facilitate both optimal procedures and a solid conceptual foundation for this intervention.

References

Barbera, M. L., & Kubina, R. M. (2005). Using transfer procedures to teach tacts to a child with autism. *Analysis Verbal Behav, 21*, 155–161.

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior and Development, 23*, 223–229. doi:10.1016/S0163-6383(01)00037-6.

Bijou, S. W., & Baer, D. M. (1965). *Child development II: universal stage of infancy.* New York: Appleton-Century-Crofts.

Bloom, K. (1979). Evaluation of infant conditioning. *Journal of Experimental Child Psychology, 27*, 60–70. doi:10.1016/0022-0965(79)90059-6.

Bloom, K., & Esposito, A. (1975). Social conditioning and its proper control procedures. *Journal of Experimental Child Psychology, 19*, 209–222. doi:10.1016/0022-0965(75)90085-5.

Bourret, J., Vollmer, T. R., & Rapp, J. T. (2004). Evaluation of a vocal mand assessment and vocal mand training procedures. *Journal of Applied Behavior Analysis, 37*, 129–144. doi:10.1901/jaba.2004.37-129.

Butz, R., & Hasazi, J. (1973). Developing verbal imitative behavior in a profoundly retarded girl. *Journal of Behavior Therapy and Experimental Psychiatry, 4*, 389–393. doi:10.1016/0005-7916(73)90016-5.

Bzoch, K. R., League, R., & Brown V. L. (2003). *Receptive-expressive emergent language test* (3rd ed.) Austin, TX: Pro-Ed.

Carroll, R. A., & Klatt, K. P. (2008). Using stimulus-stimulus pairing and direct reinforcement to teach vocal verbal behavior to young children with autism. *Analysis Verbal Behav, 24*, 135–146.

Cooper, R. P., & Aslin, R. N. (1990). Preferences for infant-directed speech in the first month after birth. *Child Development, 61*, 1584–1595. doi:10.1111/j.1467-8624.1990.tb02885.x.

Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology, 20*, 775–803. doi:10.1017/S0954579408000370.

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders, 28*, 479–485. doi:10.1023/A:1026043926488.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental Psychology, 40*, 271–283. doi:10.1037/0012-1649.40.2.271.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science, 208*, 1174–1176. doi:10.1126/science.7375928.

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforce preferences. *Journal of Applied Behavior Analysis, 29*, 519–533. doi:10.1901/jaba.1996.29-519.

Dozier, C. L., Iwata, B. A., Thomason-Sassi, J., Worsdell, A. S., & Wilson, D. M. (2012). A comparison of two pairing procedures to establish praise as a reinforcer. *Journal of Applied Behavior Analysis, 45*, 721–735. doi:10.1901/jaba.2012.45-721.

Dunn, L. M., Dunn, L. M., & Dunn, D. M. (1997). Peabody picture vocabulary test (3rd ed.) Circle Pines: American Guidance Service.

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2009). Meta-analysis of early intensive behavioral intervention for children with autism. *Journal of Clinical Child and Adolescent Psychology, 38*, 439–450. doi:10.1080/15374410902851739.

Esch, B. E. (2008). Early echoic skills assessment. In M. L. Sundberg (Ed.), *The verbal behavior milestones assessment and placement program (VB-MAPP): a language and social skills assessment program for children with autism or other developmental disabilities*. Concord: AVB Press.

Esch, B. E., Carr, J. E., & Grow, L. L. (2009). Evaluation of an enhanced stimulus-stimulus pairing procedure to increase early vocalizations of children with autism. *Journal of Applied Behavior Analysis, 42*, 225–241. doi:10.1901/jaba.2009.42-225.

Esch, B. E., Carr, J. E., & Michael, J. (2005). Evaluating stimulus-stimulus pairing and direct reinforcement in the establishment of an echoic repertoire of children diagnosed with autism. *Analysis Verbal Behav*, *21*, 43–58.

Fisher, W. W., Piazza, C. C., Bowman, L. G., & Amari, A. (1996). Integrating caregiver report with a systematic choice assessment. *American Journal on Mental Retardation*, *101*, 15–25.

Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. *Journal of Applied Behavior Analysis*, *25*, 491–498. doi:10.1901/jaba.1992.25-491.

Harris, S. L. (1975). Teaching language to nonverbal children with emphasis on problems of generalization. *Psychological Bulletin*, *82*, 565–580. doi:10.1037/h0076903.

Hung, D. W. (1976). Teaching mute retarded children vocal imitation. *Journal of Behavior Therapy and Experimental Psychiatry*, *7*, 85–88. doi:10.1016/0005-7916(76)90052-5.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore: Paul H. Brookes.

Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. A. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities*, *26*, 359–383. doi:10.1016/j.ridd.2004.09.005.

Kaufman, N. R. (1995). *Kaufman speech praxis test*. Detroit: Wayne State University Press.

Lepper, T. L., Petursdottir, A. I., & Esch, B. E. (2013). Effects of operant discrimination training on the vocalizations of nonverbal children with autism. *Journal of Applied Behavior Analysis*, *46*, 656–661. doi:10.1002/jaba.55.

Lord, C., Rutter, M., DiLavore, C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule* (2nd ed.,). Torrance: Western Psychological Services.

Miguel, C. F., Carr, J. E., & Michael, J. (2002). The effects of a stimulus-stimulus pairing procedure on the vocal behavior of children diagnosed with autism. *Analysis Verbal Behav*, *18*, 3–13.

Miliotis, A., Sidener, T. M., Reeve, K. F., Carbone, V., Sidener, D. W., Rader, L., & Delmolino, L. (2012). An evaluation of the number of presentations of target sounds during stimulus-stimulus pairing trials. *Journal of Applied Behavior Analysis*, *45*, 809–813. doi:10.9101/jaba.2012.45-809.

Normand, M. P., & Knoll, M. L. (2006). The effects of a stimulus-stimulus pairing procedure on the unprompted vocalizations of a young child diagnosed with autism. *Analysis Verbal Behav*, *22*, 81–85.

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: nonoverlap of all pairs. *Behavior Therapy*, *40*, 357–367. doi:10.1016/j.beth.2008.10.006.

Pelaez, M., Ortega, J. V., & Gewirtz, J. L. (2011). Contingent and noncontingent reinforcement with maternal vocal imitation and motherese speech: effects on infant vocalizations. *European Journal of Behavior Analysis*, *12*, 277–287. doi:10.1080/15021149.2011.11434370.

Petursdottir, A. I., Carp, C. L., Matthies, D. W., & Esch, B. E. (2011). Analyzing stimulus-stimulus pairing effects on preferences for speech sounds. *The Analysis of Verbal Behavior, 27,* 45-60.

Poulson, C. L. (1984). Operant theory and methodology in infant vocal conditioning. *Journal of Experimental Child Psychology*, *38*, 103–113. doi:10.1016/0022-0965(84)90021-3.

Rader, L., Sidener, T. M., Reeve, K. F., Sidener, D. W., Delmolino, L., Miliotis, A., & Carbone, V. (2014). Stimulus-stimulus pairing of vocalizations: a systematic replication. *Analysis Verbal Behav*, *30*, 69–74.

Rheingold, H. L., Gewirtz, J. L., & Ross, H. W. (1959). Social conditioning of vocalizations in the infant. *Journal of Comparative and Physiological Psychology*, *52*, 68–73. doi:10.1037/h0040067.

Rogers, S. J., & Dawson, G. (2010). *Early start Denver model for young children with autism*. New York: Guilford Press.

Rogers, S. J., & Vismara, L. A. (2008). Evidence-based comprehensive treatments for early autism. *Journal of Clinical Child and Adolescent Psychology*, *37*, 8–38. doi:10.1080/15374410701817808.

Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioral treatment for children with autism: four-year outcome and predictors. *American Journal on Mental Retardation*, *110*, 417–438.

Schlinger, H. D. (1995). *The behavior-analytic view of child development*. New York: Plenum Press.

Smith, R., Michael, J., & Sundberg, M. L. (1996). Automatic reinforcement and automatic punishment in infant vocal behavior. *Analysis Verbal Behav*, *13*, 39–48.

Spradlin, J. E., & Brady, N. C. (1999). Early childhood autism and stimulus control. In P. M. Ghezzi, W. L. Williams, & J. E. Carr (Eds.), *Autism: behavior analytic perspectives* (pp. 49–65). Reno: Context Press.

Stock, R. A., Schulze, K. A., & Mirenda, P. (2008). A comparison of stimulus-stimulus pairing, standard echoic training, and control procedures on the vocal behavior of children with autism. *Analysis Verbal Behav*, *24*, 123–133.

Stone, W. L., & Yoder, P. J. (2001). Predicting spoken language level in children with autism spectrum disorders. *Autism*, *5*, 341–361. doi:10.1177/1362361301005004002.

Sundberg, M. L., Michael, J., Partington, J. W., & Sundberg, C. A. (1996). The role of automatic reinforcement in early language acquisition. *Analysis Verbal Behav*, *13*, 21–37.

Sundberg, M. L., Partington, J. W. (1998). *Teaching language to children with autism or other developmental disabilities.* Pleasant Hill, CA: Behavior Analysts, Inc.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of langauge milestones. *Child Development*, *72*, 748-767.

Toth, K., Munson, J., Meltzoff, A. N., & Dawson, G. (2006). Early predictors of communication development in young children with autism spectrum disorder: joint attention, imitation, and toy play. *Journal of Autism and Developmental Disorders*, *36*, 993–1005. doi:10.1007/s10803-006-0137-7.

Turati, C., Simion, F., Milani, I., & Umiltà, C. (2002). Newborns' preference for faces: what is crucial? *Developmental Psychology*, *38*, 875–882. doi:10.1037/0012-1649.38.6.875.

Vannest, K. J., Davis, J. L., & Parker, R. I. (2013). *Single case research in schools: practical guidelines for school-based professionals (school-based practice in action)*. New York: Routledge.

Ward, S. J., Osnes, P. J., & Partington, J. W. (2007). The effects of a delay of noncontingent reinforcement during a pairing procedure in the development of stimulus control of automatically reinforced vocalizations. *Analysis Verbal Behav*, *23*, 103–111.

Warren, Z., McPheeters, M. L., Sathe, N., Foss-Feig, J. H., Glasser, A., & Veenstra-VanderWeele, J. (2011). A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics*, *127*, 1303–1311. doi:10.1542/peds.2011-0426.

Watkins, C. L., Pack-Teixeira, L., & Howard, J. S. (1989). Teaching intraverbal behavior to severely retarded children. *Analysis Verbal Behav*, *7*, 69–81.

Yoon, S., & Bennett, G. M. (2000). Effects of a stimulus-stimulus pairing procedure on conditioning vocal sounds as reinforcers. *Analysis Verbal Behav*, *17*, 75–88.

Yoon, S., & Feliciano, G. M. (2007). Stimulus-stimulus pairing and subsequent mand acquisition of children with various levels of verbal repertoires. *Analysis Verbal Behav*, *23*, 3–16.