A systematic review and evaluation of procedures for the induction of speech among persons with developmental disabilities

Teresa Mulhern, Sinéad Lydon, Olive Healy, Gerard Mollaghan, Devon Ramey & Mauro Leoni

To cite this article: Teresa Mulhern, Sinéad Lydon, Olive Healy, Gerard Mollaghan, Devon Ramey & Mauro Leoni (2016): A systematic review and evaluation of procedures for the induction of speech among persons with developmental disabilities, Developmental Neurorehabilitation

To link to this article: http://dx.doi.org/10.3109/17518423.2016.1150360

Published online: 08 Apr 2016.
A systematic review and evaluation of procedures for the induction of speech among persons with developmental disabilities

Teresa Mulherna, Sinéad Lydonab, Olive Healyb, Gerard Mollaghanb, Devon Rameyb, and Mauro Leoni
c

aSchool of Psychology, National University of Ireland, Galway, Ireland; bSchool of Psychology, Trinity College Dublin, Dublin 2, Ireland; cDisability Department, Fondazione Sospiro Onlus, CR, Italy

ABSTRACT
Objective: Deficits in vocal speech are common among those with developmental disabilities. This review examines interventions for teaching speech to individuals who presented as nonspeaking, or with low levels of vocalizations at baseline, and assesses evidence-based practice in this area. Methods: Systematic searches identified 78 studies suitable for inclusion. These studies were evaluated in terms of (a) participants, (b) intervention, (c) intervention setting, (d) intervention agent, (e) treatment efficacy, (f) generalization and maintenance of treatment effects, and (g) research rigor. Results: A variety of interventions, primarily behavioral, intended to induce vocal speech were delivered to participants with developmental disabilities aged between six months and 57 years. Treatment efficacy was variable (PND M = 52.9%; range 0%–100%); however, results indicated that behavioral interventions constituted evidence-based practice. Non-behavioral strategies were shown to have received insufficient research evaluation to date. Conclusion: Results indicate that a number of procedures can induce speech among individuals with developmental disabilities.

Introduction
Language forms the cornerstone of successful social interaction, allowing individuals to communicate their needs and influence their environment. The formation of peer relationships hinges upon the ability to communicate effectively with others, while language contributes to engagement with the environment by ameliorating educational difficulties and decreasing the likelihood of social isolation, and is also associated with cognitive abilities. Typically developing children generally demonstrate a rapid and sometimes variable trajectory in language acquisition, with children evidencing language comprehension prior to the emergence of speech. By 21 months, children typically exhibit 200 words within their vocabulary. Between 24 and 27 months, typically developing children begin to combine single words into two-word sentences, eventually producing three- and four-word sentences. By the age of three years, children’s vocabulary has typically expanded to 1,000 words, with increasing complexity and length of utterances also observed.

The seemingly effortless acquisition of vocal speech among typically developing children is in stark contrast to the delays and deficits in language development and acquisition that characterize a number of developmental disabilities, including intellectual disabilities and autism spectrum disorders (ASDs). Research suggests that between 25% and 30% of individuals with autism fail to acquire speech without direct intervention. Language is similarly impaired among those with intellectual disabilities; Arvio and Sillanpää found that 62% of a Finnish population of severe to profoundly intellectually disabled persons demonstrated significant speech defects. A significant decline in the probability of language acquisition after the age of five has been observed within developmentally disabled populations, most notably among those diagnosed with ASDs.

Research suggests that individuals with developmental disabilities demonstrate a strength and structure of peripheral oral musculature that is considered within the normal range, and that language delays or absence cannot be explained by physical abnormalities. It has been instead suggested that the failure to acquire important listener and speaker language prerequisites may be one explanation for the impairments in language observed among those with developmental disabilities. The link between language development and early social-communicative skills, such as imitation, joint attention, and pretend play, has been examined extensively in the literature and interventions to ameliorate language deficits have often targeted these specific skills. The relationship between these skills and language has been established within typically developing populations and developmentally disabled populations.

The failure to acquire spoken language has been noted to have a significant negative impact on the quality of life of those with developmental disabilities. The development of speech within this population, however minimal, constitutes one of the most significant predictors of positive outcomes, including decreases in behavioral symptom severity and improved socialization. Communication skills are related to more positive outcomes among individuals with ASD, as greater communication skills have been linked to employment, college enrolment, and independent living. Howlin and colleagues assessed the adult outcomes of 68 children with autism and found that the majority of participants (83%)...
had linguistic ability below that of a 15-year-old. These individuals demonstrated poor communication skills, persistent stereotyped behaviors, and circumscribed interests, and showed greater dependency upon family members and support services. Furthermore, few of these participants had friends, as measured by the Autism Diagnostic Interview, and few were in permanent employment. Both Howlin and colleagues and Mody et al. emphasize that the ability to speak, even a few words, in this population is the “single most important predictor of positive outcomes such as symptom reduction and increased socialization in later years”, p. 130. Furthermore, individuals with ASD who fail to speak typically display low IQ scores, weak socialization abilities, decreased motivation to communicate, deficits in joint attention and motor issues. Receptively, individuals who display delays in language also fail to orient to speech or to their names when called. Such findings indicate a broader deficit in social reciprocity and motivation, thereby indicating a close relation between social and communication skills. Venter and colleagues also found that verbal skills were a significant predictor of an individuals’ social adaptive functioning. Children who fail to acquire speech typically have limited independence, may struggle to integrate with their peers and are characterized as more withdrawn than their typically developing peers. When an individual is unable to speak, their capacity to articulate their needs, interact with their environment, and benefit from education is negatively impacted. Language deficits are also associated with impaired behavioral adjustment, with estimates that 31%–73%, or 40%–90% of young children with language deficits may demonstrate co-morbid behavioral difficulties. Such behavioral problems, associated with language delay, have been shown to persist from a young age, during childhood, adolescence, and adulthood. Given this potential for language deficits to negatively impact the lives of individuals with developmental disabilities, the current paper aimed to provide a systematic overview of the literature that has focused upon either the facilitation of vocal behavior (i.e., speech) where it does not exist or the augmentation of existing low-levels of vocalizations among persons with developmental disabilities. For the purpose of this review, speech was defined as the ability to express thoughts or needs by articulate sounds, while vocalizations were defined as the act or process of producing sounds with the voice. It is important to note here also the distinctions between the terms developmental delay, developmental disability and intellectual disability. A diagnosis of a developmental delay indicates that an individual has failed to meet significant developmental milestones within the typical age range. In comparison, the term developmental disability refers to a severe and chronic disability originating at birth, which is anticipated to persist indefinitely, resulting in substantial restrictions within an individual’s functioning in major life activities. Developmental disabilities include ASD, attention deficit hyperactivity disorder (ADHD), and intellectual disabilities. Intellectual disabilities are a specific form of developmental disability originating prior to the age of 18 that are characterized by substantial limitations in both intellectual functioning and adaptive behaviors. A further aim of the current paper was to identify evidence-based practice within this area and identify specific procedures that have been demonstrated effective in the facilitation of speech.

Method

Literature search

We identified articles suitable for inclusion by performing comprehensive searches of six electronic databases in May 2015: Scopus, PsycINFO, PsycARTICLES, Web of Science, Education Resources Information Center (ERIC), and Psychology and Behavioral Sciences Collection. Searches were carried out by inputting: intraverbal training, speech, language, mand training, tact training, contingent vocal imitation, echoic training, stimulus pairing, or echoic to mand training in combination with intervention or procedure, and in combination with autism*, pervasive developmental disorder, developmental dis*, mental retardation or intellectual disability, with no limit on year of publication. These search terms were inserted into the title, abstract and keyword search field of each database. All abstracts returned during the electronic search process were examined to ascertain their suitability for inclusion within the review. The first author, who holds an MSc in Applied Behavior Analysis (ABA), and is a PhD candidate in ABA, examined the abstract section of each study to determine suitability for inclusion in the current systematic review based upon the inclusion and exclusion criteria outlined below. If the eligibility of a study was unclear based upon the information provided in the abstract, the first author then assessed the methods section to determine if the study met the inclusion criteria. Furthermore, a review of the reference lists of all included studies was conducted in order to identify other studies meeting the inclusion criteria.

Inclusion and exclusion criteria

Searches were limited to English language, peer-reviewed journals. Studies were included within the review if they: (a) reported the application of an intervention intended to either induce vocal behavior or to augment existing low levels of vocal behavior, and (b) included the treatment of at least one participant diagnosed with a developmental disability (e.g., an ASD, intellectual disability, or ADHD) or developmental delay. We excluded studies if they: (a) only assessed vocal behavior at baseline and did not intervene, or (b) were described as a case study.

Study classification

Studies were categorized according to the type of intervention evaluated. They were classified as either: (a) behavioral intervention, which comprised interventional strategies derived from the science of ABA; (b) non-behavioral intervention, which comprised intervention techniques drawn from disciplines other than ABA, or (c) mixed intervention, which comprised a combination of interventional strategies derived from ABA and those emerging from alternative disciplines.
Data extraction

Data were extracted on the following variables for each included study: participant characteristics (speech prior to intervention, sample size, age, diagnoses, intellectual functioning), treatment characteristics (target behavior, treatment type, treatment components, treatment setting, intervention agent, treatment efficacy, generalization, maintenance), and study characteristics (research design). Three of the authors extracted these data from each study which they entered into separate Microsoft Excel files. Subsequently, the first author compared all data to assess interrater agreement and ensure data extraction accuracy. In the instance of any inconsistencies in data extraction, the authors discussed any disagreements until a consensus had been reached.

Treatment efficacy

For studies utilizing a single-subject experimental design (SSED), a non-parametric statistic, percentage of non-overlapping data (PND), was used in order to ascertain treatment efficacy for the speech induction interventions implemented in studies. PND is calculated by determining the percentage of treatment data points that do not overlap with the baseline data points. A treatment that receives a PND score that is greater than 90% is considered highly effective, 70%–90% is considered moderately effective, 50%–70% is considered mildly effective, while a treatment that receives a PND score below 50% is deemed ineffective. In order to calculate treatment efficacy for interventions, PND is first calculated for each individual research participant and the median of these scores is then calculated and reported per study. In the current review, the level of vocal behavior in each study phase was compared (i.e., baseline was compared to intervention phase) to ascertain the impact of interventions on the speech or vocalizations of participants with developmentally disabilities or delays.

Generalization

The studies included were examined for evidence of the generalization of treatment effects as defined by Lydon et al. In this way, generalization was operationalized as a concurrent behavior increase that was observed in response to novel stimuli, people, or settings in which treatment had not been implemented.

Maintenance

Studies were also evaluated for evidence of the inclusion of maintenance measures as per Lydon et al. Maintenance was defined as the continued emission of treatment-level target behaviors following the cessation of treatment.

Research rigor and evidence-based practice

Empirical support for the various interventional techniques employed was assessed using Reichow’s criteria for evaluating research rigor and evidence-based practice. Previous reviews have used these criteria to determine the evidence-based practice (EBP) status of response redirection as a treatment for challenging behavior, early intervention for toddlers with autism, and gluten- and casein-free diets for the symptoms of autism. The application of these criteria requires the assessment of the methodological quality of each study. Studies are assessed for the presence of both primary and secondary quality indicators; Primary quality indicators include participant characteristics, independent variables, dependent variables (for both SSEDs and group designs), baseline condition, visual analysis, and experimental control (SSEDs only), comparison condition, link between research question and data analysis and statistical analysis (group designs only). Secondary quality indicators include: interobserver agreement, blind raters, fidelity, measurement of generalization or maintenance, social validity (for both SSEDs and group designs), kappa (SSEDs only), random assignment, attrition and effect size (group designs only). Studies may subsequently be categorized as “strong”, “adequate”, or “weak” depending on the presence or absence of these primary and secondary quality indicators.

EBP was assessed by analyzing the number of participants with “effective” treatment outcomes in SSEDs that were classified as “strong” (SSEDs) or adequate (SSED or group designs) and in group designs classified as “strong” (Group A) or “adequate” (Group B). Researchers characterized treatments as “effective” if they resulted in a PND score exceeding 50% while those resulting in PND scores of less than 50% were rated as “ineffective”. In the case of group studies, treatments were noted to be “effective” when they resulted in statistically significant increases in vocal behavior while those resulting in statistically non-significant increases in vocal behavior were categorized as “ineffective”. When it was not possible to calculate PND scores for a study (42.3% of the included studies), the effectiveness, or ineffectiveness, was determined by examining the author(s’) reports of effects and assessing whether the authors reported an intervention to have been effective or ineffective at increasing vocal behavior. Subsequently, EBP was computed across the studies characterized as delivering “effective” intervention using the following formula:

$$\text{EBP} = \frac{\text{Number of overall agreements}}{\text{Number of overall agreements and non-agreements}} \times 100.$$
The mean interrater agreement was found to be 94.9% (range 70%–100%) for data extraction, while the mean interrater agreement was found to be 95.1% (range 74.4%–100%) for the research rigor assessment. All disagreements were discussed and reviewed until complete agreement was reached between raters.

The reliability of PND calculations was also assessed for each study. As PND is a continuous variable, it is recommended to calculate interrater reliability for the PND metric by assessing the correlation between the scores obtained by raters.45,46 A highly significant correlation between the PND calculations of raters one and two ($r = 0.97, p < .001$), raters one and three ($r = 0.99, p < .001$), and raters two and three ($r = 0.98, p < .001$) was observed.

Results

A total of 72 articles, reporting 78 studies in total, published between the years 1967 and 2015, met the inclusion criteria. Among these, 55 (70.5%) studies employed SSEDs, while the remaining 23 studies (29.5%) utilized alternate experimental designs, including pre- and post-test designs (5.1%; $n = 4$) and randomized group designs comparing interventions without a control group (20.5%; $n = 16$). In total, 1213 participants aged between six months and 57 years received an intervention intended to induce speech or vocalizations within the included studies. Results are presented separately for each of the three categories of studies: Behavioral Intervention, Non-Behavioral Intervention, and Mixed Intervention.

Behavioral interventions

A total of 63 published articles47–109 reported the use of behavioral interventions intended to target speech or vocalizations with 968 participants diagnosed with developmental disabilities or delays. Six of these articles54,59,66,100,104,109 each reported two separate studies employing behavioral interventions, resulting in a total of 69 studies for analysis within this category. Of these, 52 utilized SSEDs, of which 84.6% ($n = 44$) provided data with which to ascertain treatment efficacy by calculating PND scores. The remaining 17 studies in this category either utilized alternative experimental designs, which were not conducive to the calculation of PND scores, or were SSEDs that failed to provide the necessary graphical information to conduct PND calculations.

Participant characteristics

Age

Participants aged between 0.6 and 57 years ($M = 5.3$ years) participated in the 69 studies categorized as "behavioral intervention". Across these studies, children (under 12 years) most frequently served as research participants ($n = 62; 89.9%$). It was uncommon for studies to include adult participants over the age of eighteen (2.9%; $n = 2$) or adolescent participants (12-18 years; 1.4%; $n = 1$) alone. Four studies (5.8%) employing behavioral interventions utilized speech induction procedures with participants from multiple age categories.

Primary diagnosis

Participants were most commonly diagnosed with autism (72.5%; $n = 50$ studies). However, eight studies (11.6%) included participants with a primary diagnosis of intellectual disability, while a further eleven studies (15.9%) provided intervention to participants with Down syndrome. The inclusion of participants diagnosed with other primary and secondary developmental disabilities or delays within such training procedures was relatively infrequent. Five of the studies (7.2%) also delivered intervention to individuals with a diagnosis of language impairment, while a further five of these studies (7.2%) also provided intervention to individuals with a diagnosis of developmental delay.

Intellectual functioning

The treatment of participants of normal intellectual functioning (IQ > 80) was most frequently reported by the included studies (71%; $n = 49$). A further four studies provided intervention to individuals of intellectual functioning within the normal range in combination with individuals of varying intellectual disability, ranging from borderline intellectual disability (IQ of 70–80) to severe intellectual disability. Two studies treated participants with mild intellectual disabilities (IQ of 50–70) only.60,73 Two studies treated participants with moderate intellectual disabilities (IQ of 35–49) only.47,107 Lancioni and colleagues85 included participants with a severe intellectual disability (IQ ≤ 35) only, and 11 studies included participants with a range of intellectual disability severities ranging from mild to severe.

Treatment characteristics

Setting

Behavioral interventions were most commonly delivered in participants’ natural settings, such as schools, day care centers and participants’ own homes (55.1%; $n = 38$). One study was implemented within a leisure-based day program, which was designed to provide leisure activities and medical services for participants in addition to providing respite services.84 Interventions were also commonly conducted across multiple settings (17.4%; $n = 12$), such as within the participants’ own home, school and university clinics.56 A further 16 studies reported behavioral interventions were implemented within clinical settings, such as laboratories, therapy rooms, clinics, and psycho-education units attached to universities (23.2%). Two studies failed to provide data regarding the setting in which treatment was delivered [2.9%].69,107

Intervention agents

Behavioral interventions were most commonly delivered by the experimental team (52.2%; $n = 36$). The experimental team was typically comprised of the researchers, in addition to appropriately trained personnel that were external to the participant’s typical service team. However, a significant proportion of the behavioral research also employed multiple intervention agents across the study phases (36.2%; $n = 25$), including school staff and family members and members of the experimental team. A number of studies also reported on interventions conducted by family members exclusively (5.8%; $n = 4$) or solely by school or service staff members (5.8%; $n = 4$).
Treatment approaches

In total, 71% (n = 49) of the behavioral intervention studies implemented treatment packages including reinforcement components. Within the current literature, reinforcement is typically delivered contingent upon the emission of vocalizations, both whole words and part words. Typically, within the included studies when reinforcement is utilized, the participant speaks aloud a target word, or a successive approximation of the word, and receives the requested item in the case of requests/mands, or receives an alternate reinforcer such as social praise in the case of labels/tacts. 10

A further 62.3% (n = 43) of the studies included prompting procedures in the treatment package delivered. Prompting procedures involve the alteration of antecedent conditions in order to occasion correct responding. 11 The prompting procedures within the cited literature include verbal, echoic, gesture, model, physical, and least-to-most prompts to occasion target responses from the participants. A common prompting procedure within the included literature is that of echoic prompts, in which the experimenter provides a verbal model of the target word to the participant, which they then must echo. 93 Least-to-most prompting procedures are also used within the included studies, whereby participants are provided with the least intrusive level of prompts required to produce the targeted vocalizations, such as gesture prompts. However, if the participant fails to provide the target vocal response, a more intrusive prompt is introduced, such as a partial physical prompt. 93 An example of physical prompts employed to induce speech, involved the interventionist applying pressure to the participant’s mouth to induce production of the target sound. 48 Gesture prompts are used within sign language training programs that targeted vocalizations concurrently. If a participant fails to emit a manual sign, the interventionist initiates a prompt sequence that begins within a gesture prompt. 53 Gesture prompts are also employed within Picture Exchange Communication System (PECS) training programs, wherein the interventionist provides a gesture to the appropriate PECS symbol for the participant to select. 56 Gesture prompts are further used to signal to the participant that it is their turn to communicate, for example, the interventionist points towards an item relevant to the communication act, or towards the participant’s mouth. 62 A further 17.4% (n = 12) of the included studies employed shaping procedures, which involve reinforcing successive steps towards a target behavior until a terminal goal is achieved. 111 For example, initially the first syllable of a word is the target vocalization resulting in the delivery of reinforcement. Mastery of this vocalization is then followed by targeting a more complex vocalization that more closely approximate the terminal word. This process of shaping successive approximations of the target word is continued until the whole word is vocalized. 88 Other behavioral interventions employed include stimulus–stimulus pairing (n = 10; 14.5%), which involves simultaneously presenting a neutral stimulus (such as vocalizations) along with highly preferred stimuli such as toys and food. 67 Extinction procedures are utilized by 2.9% of studies (n = 2), to eliminate inappropriate behaviors that had previously functioned as forms of communication, with a further 2.9% (n = 2) employing punishment-based interventions within multi-component treatment packages. Of the studies that employ punishment-based interventions, one employs positive punishment practices in the form of verbal reprimands in order to decrease inappropriate or incorrect vocal mands. 70 The remaining study 87 indicates that punishment-based procedures are taught to parents within a parent-training program designed to equip parents with knowledge and skills required to teach pre-speech and speech skills to their children. Each week parents are provided with one-hour lectures and are taught various ABA-based procedures. Within the fourth week of parent training, parents are provided with information on the implementation of punishment. However, the authors fail to explicitly outline if, or how, these procedures are applied to facilitate speech within their research (see Wolchik and Weitz, 66 p. 378). A further 2.9% (n = 2) of studies utilize echoic training, a procedure designed to induce verbal operants (tacts or mands) in the learner. Within this procedure the learner is provided with a target word delivered via echoic prompts, the learner then emits the echoic response, which is reinforced. 17

A number of behavioral intervention studies incorporated milieu teaching within their procedural framework (n = 5). 60,62,71–73 Milieu teaching procedures involve increasing learning opportunities for language skills by embedding the child’s environment with such opportunities and by capitalizing on participants’ interests and enhancing their motivation to access reinforcers. 112 Incidental teaching is used within milieu teaching and is combined with time delay prompts in addition to modelling procedures to evoke mands, or requests. Milieu teaching includes the development of play-based routines within which the child’s communication needs naturally occur and teaching prompts can be delivered. 60

An additional procedure used within the reviewed studies was PECS. 66 PECS is a functional communication system, which focuses upon the facilitation of communication within non-verbal populations. Communication between individuals is based upon the learner approaching a communication partner and exchanging pictures in order to communicate. PECS is a form of child-initiated communication, which centers upon the systematic fading of prompts in order to minimize prompt dependence and is implemented over six phases. 63

Percentage of non-overlapping data (PND)

Of the 53 behavioral intervention studies employing single-subject research designs, it was possible to calculate PND scores for 83% (n = 44) of these studies. Of these studies, over a third delivered interventions that were “highly effective”, achieving PND scores of over 90% (43.2%; n = 19). While each of these 19 studies utilized multi-component intervention packages, all utilized reinforcement-based procedures. A further 36.4% (n = 16) of studies evaluating behavioral interventions yielded PND scores ranging between 70% and 90%, indicative of “moderately effective” intervention. A further 22.7% (n = 10) of behavioral intervention studies yielded PND scores between 50% and 70%, and therefore could be classified as “mildly effective”. However, a considerable number (n = 22; 50%) of studies yielded PND scores below 50%, and were classified as “ineffective”. Mean treatment efficacy scores for each intervention strategy are presented in Table 1.

Table 2 presents the PND scores of all studies that allowed for the calculation of this statistic. Table 2 also provides
<table>
<thead>
<tr>
<th>Treatment efficacy</th>
<th>Average PND score (range)</th>
<th>*Total no. of studies</th>
<th>*Total no. of applications</th>
<th>Procedure</th>
<th>Evidence-based practice Z score</th>
<th>Evidence-bases ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Highly effective (PND &gt; 90%)</strong></td>
<td>100%</td>
<td>1</td>
<td>4</td>
<td>Function-based video modelling</td>
<td>8</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>96.3% (96.3%)</td>
<td>3</td>
<td>32</td>
<td>High probability behavior sequencing</td>
<td>40</td>
<td>Promising</td>
</tr>
<tr>
<td></td>
<td>96.7% (94.5%–100%)</td>
<td>4</td>
<td>25</td>
<td>Multiple exemplar training</td>
<td>34</td>
<td>Promising</td>
</tr>
<tr>
<td><strong>Moderately effective (PND 70%–90%)</strong></td>
<td>85.7% (71.4%–100%)</td>
<td>7</td>
<td>271</td>
<td>Milieu teaching</td>
<td>129</td>
<td>Evidence-based practice</td>
</tr>
<tr>
<td></td>
<td>85.7% (85.7%)</td>
<td>2</td>
<td>23</td>
<td>Contingent adult imitation</td>
<td>27</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>77.3% (64.3%–100%)</td>
<td>9</td>
<td>96</td>
<td>Manipulating motivating operations</td>
<td>110</td>
<td>Evidence-based practice</td>
</tr>
<tr>
<td></td>
<td>70.8% (19.2%–100%)</td>
<td>12</td>
<td>131</td>
<td>Behavior shaping</td>
<td>132</td>
<td>Evidence-based practice</td>
</tr>
<tr>
<td><strong>Mildly effective (PND 50%–70%)</strong></td>
<td>68.2% (0%–100%)</td>
<td>51</td>
<td>503</td>
<td>Prompting</td>
<td>368</td>
<td>Evidence-based practice</td>
</tr>
<tr>
<td></td>
<td>62.6% (47.8%–90%)</td>
<td>3</td>
<td>4</td>
<td>Echoic training</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>61.8% (0%–100%)</td>
<td>52</td>
<td>453</td>
<td>Reinforcement</td>
<td>292</td>
<td>Evidence-based practice</td>
</tr>
<tr>
<td><strong>Ineffective (PND &lt; 50%)</strong></td>
<td>46.3% (16.7%–70%)</td>
<td>3</td>
<td>4</td>
<td>Extinction</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>37.3% (0%–100%)</td>
<td>10</td>
<td>26</td>
<td>Stimulus–stimulus pairing</td>
<td>6</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>34.6% (0%–100%)</td>
<td>12</td>
<td>181</td>
<td>Picture Exchange Communication System (PECS)</td>
<td>85</td>
<td>Evidence-based practice</td>
</tr>
</tbody>
</table>

*Note. A Z-score below 30 is considered to lack evidence-based support, a Z-score between 30 and 60 is considered to be "promising", while a Z-score of 60 or more is deemed as evidence-based practice. *Total no. of studies and applications refers to all single-subject experimental designs and group designs that employed the procedure; PND and evidence-based practice was calculated for a subset of these.*
Table 2. A summary of participants’ vocal speech capabilities prior to intervention, targeted behavior(s), intervention(s) applied, and treatment efficacy (Percentage of Non-Overlapping Data scores) across the studies reviewed.

<table>
<thead>
<tr>
<th>Article</th>
<th>Vocal speech prior to intervention</th>
<th>Targeted behavior</th>
<th>Intervention</th>
<th>Median PND score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barrett and Sisson, 47</td>
<td>Expressive language between 12 and 30 months.</td>
<td>Sentence triads – echolics.</td>
<td>(1) Oral speech condition. (2) Total communication condition. (3) Modified total communication condition.</td>
<td>(1) 52.9%</td>
</tr>
<tr>
<td>Bauer and Jones, 48</td>
<td>Expressive language between 4 and 10 months.</td>
<td>Echosics and mands.</td>
<td>Echosics: (1) Exaggerated model and physical prompt condition. (2) Exaggerated model only condition. (3) Time delay prompt condition.</td>
<td></td>
</tr>
<tr>
<td>Mands: (1) Full prompt condition. (2) Partial prompt condition. (3) Time delay prompt condition.</td>
<td>(1) 92.3% (2) 96.88% (3) 100%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bourret et al., 51</td>
<td>Two participants had no vocal behavior, while the third participant engaged in vocal verbal behavior in limited circumstances.</td>
<td>Mands.</td>
<td>(1) Prompting and reinforcement condition. (2) Behaviour shaping, stimulus fading, prompting and reinforcement condition. (3) Prompt fading and reinforcement condition.</td>
<td>(1) 88.1%</td>
</tr>
<tr>
<td>Carbone et al., 53</td>
<td>Two participants could independently mand using sign language (10-15 items), while the third participant required partial and full physical prompts to produce signs.</td>
<td>Mands.</td>
<td>Reinforcement, manipulating motivating operations, and prompting procedures.</td>
<td>64.3%</td>
</tr>
<tr>
<td>Carroll and Klatt, 54</td>
<td>Study 1: Participant 1 demonstrated a few vocal sounds, but did not have an echoic, mand, or tact repertoire. Participant 2 demonstrated an echoic, mand, and tact repertoire, but had a limited intraverbal repertoire.</td>
<td>Study 1: One-syllable vocalizations.</td>
<td>Study 1: (1) Stimulus-stimulus pairing and reinforcement condition. (2) Direct reinforcement condition for one participant.</td>
<td>(1) 5.6% (2) 0%</td>
</tr>
<tr>
<td>Charlop-Christy et al., 56</td>
<td>Participant 1 had an expressive language equivalent of 1 year 2 months (Minnesota Child Development Inventory) and very little spontaneous speech. Participant 2 had a communication equivalent of 10 months (Vineland Adaptive Behaviour Scales) and no spontaneous speech. Participant 3 had no spontaneous speech.</td>
<td>Mands and echolics.</td>
<td>PECS training with reinforcement and prompting procedures.</td>
<td>50%</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Article</th>
<th>Vocal speech prior to intervention</th>
<th>Targeted behavior</th>
<th>Intervention</th>
<th>Median PND score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dada and Alant, 113</td>
<td>Participants had fewer than 15 intelligible words.</td>
<td>Tacts.</td>
<td>Aided language simulation – pointing to spoken word on a communication board.</td>
<td>58.4%</td>
</tr>
<tr>
<td>Esch et al., 59</td>
<td>Study 1: None of the three participants had any speech or echoic behavior at baseline.</td>
<td>One-syllable vocalizations.</td>
<td>Study 1: (1) Echoic condition. (2) Echoic with antecedent pairings condition.</td>
<td>Study 1: (1) 0% (2) 0%</td>
</tr>
<tr>
<td></td>
<td>Study 3: Participants 2 and 3, from the first study, participated in Study 3.</td>
<td>One- and two-syllable vocalizations.</td>
<td>Study 3: Differential reinforcement procedures.</td>
<td>Study 3: 50%</td>
</tr>
<tr>
<td>Esch et al., 58</td>
<td>No speech or echoic behavior at baseline.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Franco et al., 62</td>
<td>Expressive language between 5 and 9 months.</td>
<td>Child-initiated intentional communication.</td>
<td>Prelinguistic milieu teaching (PMT) techniques.</td>
<td>100%</td>
</tr>
<tr>
<td>Ganz et al., 63</td>
<td>One participant could imitate some word approximations, while the remaining two participants had no intelligible speech.</td>
<td>Intelligible mands and word approximations of mands.</td>
<td>PECS training with backward chaining and prompting procedures.</td>
<td>0%</td>
</tr>
<tr>
<td>Gazdag and Warren, 64</td>
<td>Limited vocalizations and imitation skills.</td>
<td>Vocal imitations.</td>
<td>Adult contingent vocal imitation with incidental teaching procedures.</td>
<td>85.7%</td>
</tr>
<tr>
<td>Greenberg et al., 66</td>
<td>Study 1: Three participants had no speech, while the fourth participant had limited one-syllable word approximations. Study 2: Both participants were able to imitate consonant-vowel combinations.</td>
<td>Word approximations of mands.</td>
<td>Study 1: PECS training and differential reinforcement procedures.</td>
<td>Study 1: 0%</td>
</tr>
<tr>
<td>Hung, 70</td>
<td>One participant was non-verbal, while the remaining participant had a limited verbal repertoire.</td>
<td>‘Yes’ and ‘no’ mands.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iacono et al., 71</td>
<td>Average word production was 33 words per participant.</td>
<td>Spoken words and vocalizations.</td>
<td>Interactive and milieu teaching, incidental teaching, responsivity education, and prompting procedures.</td>
<td>71.4%</td>
</tr>
<tr>
<td>Kelley et al., 74</td>
<td>Two participants emitted one-word utterances, while the third participant emitted one- to three-word utterances.</td>
<td>Mands and tacts.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kodak and Clements, 75</td>
<td>High levels of vocal stereotypy, however the participant demonstrated little functional language.</td>
<td>Mands and tacts.</td>
<td>(1) Mand training – manipulating motivating operations, prompting, and reinforcement. (2) Tact training – prompting and reinforcement. (3) Echoic training – prompting and reinforcement.</td>
<td>(1) 73.8% (2) 84.3% (3) 90%</td>
</tr>
<tr>
<td>Koegel et al., 78</td>
<td>No functional speech or language skills.</td>
<td>Echocs, tacts, and spontaneous utterances.</td>
<td>Reinforcement, prompting, manipulating motivating operations, behavior shaping, and multiple exemplar training.</td>
<td>(1) 95.5%</td>
</tr>
<tr>
<td>Article</td>
<td>Vocal speech prior to intervention</td>
<td>Targeted behavior</td>
<td>Intervention</td>
<td>Median PND score</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Koegel et al., 76</td>
<td>Participants demonstrated limited intelligible speech production.</td>
<td>One-syllable vocalizations.</td>
<td>(1) Naturalistic condition – multiple exemplar training, manipulating motivating operations, reinforcement, prompting, and behavior shaping. (2) Analogue condition – Prompting, reinforcement, and behavior shaping.</td>
<td>(1) 100% (2) 28%</td>
</tr>
<tr>
<td>Koegel et al., 79</td>
<td>No functional speech.</td>
<td>Echoics (phonemes and words) and mands.</td>
<td>Individualized orienting cues and prompting procedures.</td>
<td>100%</td>
</tr>
<tr>
<td>Kroeger and Nelson, 81</td>
<td>Participant engaged in echolalia but demonstrated limited language in the form of one or two word phrases.</td>
<td>Echoics, mands and tacts.</td>
<td>Incidental teaching, differential reinforcement, discrete trial training, and prompting procedures.</td>
<td>77%</td>
</tr>
<tr>
<td>Laski et al., 83</td>
<td>Participants 1-4 were functionally nonverbal, and could only imitate a few vocalizations and words upon request. Participants 5-8 were echolalic, and had larger vocabularies, but rarely spoke spontaneously.</td>
<td>Echoics, tacts, and spontaneous speech.</td>
<td>Natural Language Paradigm – task variation and multiple exemplars modelling, following the child’s lead, and reinforcement of verbal attempts.</td>
<td>94.5%</td>
</tr>
<tr>
<td>LeBlanc et al., 84</td>
<td>One participant demonstrated negligible verbal communication. Participant 2 displayed echolalia and had a limited mand and tact repertoire, while the remaining participant also displayed echolalia and inappropriate speech.</td>
<td>Mands.</td>
<td>Natural Language Paradigm – prompting and reinforcement procedures.</td>
<td>100%</td>
</tr>
<tr>
<td>Miguel et al., 86</td>
<td>Participants demonstrated limited verbal behavior in the form of mands, tacts, and intraverbals. Two participants had no vocal imitation skills, while the remaining participant demonstrated vocal imitation abilities.</td>
<td>One-syllable vocalizations.</td>
<td>Stimulus-stimulus pairing.</td>
<td>37.5%</td>
</tr>
<tr>
<td>Milliotis et al., 87</td>
<td>One participant had low-frequency vocal play and no echoic behavior, while the remaining participant had low-frequency vocal play and limited vocal behaviour under echoic control.</td>
<td>One-syllable vocalizations.</td>
<td>(1) Stimulus-stimulus pairing – one target sound per trial. (2) Stimulus-stimulus pairing – one non-target sound per trial. (3) Stimulus-stimulus pairing – three presentations of a target sound per trial. (4) Stimulus-stimulus pairing – three presentations of a non-target sound per trial.</td>
<td>(1) 87.1% (2) 4.6% (3) 53% (4) 4.2%</td>
</tr>
<tr>
<td>Oakes et al., 118</td>
<td>Participant had low levels of vocal behaviour.</td>
<td>Verbal communication acts.</td>
<td>Naturalistic parent-implemented language intervention – follow-in commenting, environmental arrangement, function-based extinction, reinforcement, and prompting procedures.</td>
<td>16.7%</td>
</tr>
<tr>
<td>Plavnick and Ferrer, 89</td>
<td>None of the participants had any form of functional communication. Two participants displayed delayed echolalia, but no functional speech. Two participants had no history of emitting vocal behaviour.</td>
<td>Mands.</td>
<td>(1) Function based video modelling, reinforcement, and prompting procedures. (2) Non-function based video modelling, reinforcement, and prompting procedures.</td>
<td>(1) 100% (2) 25%</td>
</tr>
</tbody>
</table>
Table 2. (Continued).

<table>
<thead>
<tr>
<th>Article</th>
<th>Vocal speech prior to intervention</th>
<th>Targeted behavior</th>
<th>Intervention</th>
<th>Median PND score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roche et al., 90</td>
<td>One participant displayed echolalia and infrequent one-word requests, but mainly resorted to babbling-like sounds, gestures, and/or leading to communicate.</td>
<td>Mands.</td>
<td>(1) Speech generating devices, contriving motivating operations, prompting, and reinforcement procedures.</td>
<td>(1) 0%</td>
</tr>
<tr>
<td></td>
<td>The remaining participant had no speech, but was able to make babbling-like sounds, and sometimes used sounds, gestures, and/or leading to communicate.</td>
<td></td>
<td>(2) Speech generating devices, contriving motivating operations, time delay, prompting, and reinforcement procedures.</td>
<td>(2) 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3) Contriving motivating operations and reinforcement procedures.</td>
<td>(3) 75%</td>
</tr>
<tr>
<td>Rogers et al., 91</td>
<td>Participants were documented to use less than five functional words per day. The MSEL expressive language age of participants was between 6 and 18 months.</td>
<td>Word approximations, echoics, mands, and tacts.</td>
<td>(1) Denver model – naturalistic behavioral teaching alternated with didactic (massed trials) teaching; prompting, behaviour shaping, and reinforcement procedures.</td>
<td>(1) 75%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2) PROMPT model – use of auditory and tactile cues; prompting, behaviour shaping, and reinforcement procedures.</td>
<td>(2) 83.3%</td>
</tr>
<tr>
<td>Ross and Greer, 93</td>
<td>No spontaneous speech and no vocal imitation.</td>
<td>Echoics and mands.</td>
<td>High probability behaviour sequencing, modelling, and reinforcement procedures.</td>
<td>96.3%</td>
</tr>
<tr>
<td>Schlosser et al., 94</td>
<td>Little or no functional speech; participants were required to have less than ten spoken words to qualify for inclusion.</td>
<td>Mands.</td>
<td>(1) Speech condition – speech generating device with speech output, prompting, reinforcement, and corrective feedback.</td>
<td>(1) 19%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2) No speech condition – speech generating device without speech output, prompting, reinforcement, and corrective feedback.</td>
<td>(2) 4.8%</td>
</tr>
<tr>
<td>Sigafos et al., 96</td>
<td>One participant had not made any vocalizations for one month prior to the study. The remaining two participants could vocalize and make speech-like sounds, but could not make intelligible words.</td>
<td>Mands.</td>
<td>Acquisition training – prompting and reinforcement procedures.</td>
<td>(1) 0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1) Speech generating device with output and reinforcement procedures.</td>
<td>(2) 0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2) Speech generating device without output and reinforcement procedures.</td>
<td></td>
</tr>
<tr>
<td>Stock et al., 97</td>
<td>All three participants lacked an echoic and tact repertoire. Two participants were able to mand for at least 10 items via picture exchange, but the third participant was only able to lead, point, or use gestures to mand.</td>
<td>One-syllable vocalizations.</td>
<td>(1) Stimulus-stimulus pairing condition – vocal model paired with preferred edible.</td>
<td>(1) 94.1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2) Echoic training condition – contingent reinforcement.</td>
<td>(2) 47.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3) Control condition – preferred edibles provided non-contingently.</td>
<td>(3) 17.4%</td>
</tr>
<tr>
<td>Thomas et al., 96</td>
<td>No functional words, however, participants had a restricted range of monosyllables. None communicated using vocalizations or gestures.</td>
<td>Vocal approximations, echoics and mands.</td>
<td>Modelling, prompting, and reinforcement.</td>
<td>66.7%</td>
</tr>
<tr>
<td>Article</td>
<td>Vocal speech prior to intervention</td>
<td>Targeted behavior</td>
<td>Intervention</td>
<td>Median PND score</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Tincani, 99</td>
<td>Both participants could vocally imitate, but did not use functional speech without prompts.</td>
<td>Mands.</td>
<td>(1) PECS training condition – prompting and reinforcement. (2) Sign language training condition – prompting and reinforcement.</td>
<td>(1) 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2) 100%</td>
</tr>
<tr>
<td>Travis and Geiger, 101</td>
<td>Study 1: Neither participant used vocal speech to communicate nor did they use an augmentative communication system. Study 2: Participant was unable to functionally communicate using pictures.</td>
<td>Word and vocal approximations.</td>
<td>Study 1: PECS training, prompting, and reinforcement. Study 2: Contingent reinforcement for word or vocal approximations.</td>
<td>Study 1: 0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Study 2: 100%</td>
</tr>
<tr>
<td>Valentino et al., 103</td>
<td>Both presented with some verbal language, however, participants demonstrated little intentional communication.</td>
<td>Mands and tacts.</td>
<td>PECS training and reinforcement.</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>The participant echoed two sounds, but did not emit tacts or intraverbals and mands were limited to gestures.</td>
<td>Intraverbals.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1) Echoloc prompting plus reinforcement condition. (2) Echoloc and model prompting plus reinforcement condition.</td>
<td>(1) 71%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2) 88.4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(3) 0%</td>
</tr>
<tr>
<td>Valentino et al., 104</td>
<td>Study 1: Both participants demonstrated limited echoic behavior and had no functional speech. Study 2: The participant demonstrated no echoic behaviors or functional speech.</td>
<td>Study 1: Vocalizations.</td>
<td>Study 1: Sign extinction and differential reinforcement. Study 2: Sign extinction and differential reinforcement.</td>
<td>Study 1: 70%</td>
</tr>
<tr>
<td></td>
<td>Study 1: All three participants failed to demonstrate any speaker behaviour or vocal imitative skills. Study 2: All three participants failed to demonstrate any speaker behaviour or vocal imitative skills.</td>
<td>Study 2: Vocalizations.</td>
<td></td>
<td>Study 2: 52.2%</td>
</tr>
<tr>
<td></td>
<td>Study 1: One-syllable vocalizations. Study 2: One-syllable vocalizations.</td>
<td>Study 1: One-syllable vocalizations. Study 2: One-syllable vocalizations.</td>
<td>Study 1: Stimulus-stimulus pairing. Study 2:</td>
<td>Study 1: 100%</td>
</tr>
<tr>
<td></td>
<td>Study 2:</td>
<td>Study 2:</td>
<td></td>
<td>Study 2:</td>
</tr>
<tr>
<td></td>
<td>Study 1:</td>
<td></td>
<td>(1) Stimulus-stimulus pairing. Study 2: (1) 50%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Study 1:</td>
<td></td>
<td></td>
<td>(2) 0%</td>
</tr>
</tbody>
</table>

Note. Only studies for which it was possible to calculate a PND score (Behavioral Intervention n = 44 studies arising from 38 papers; Non-Behavioral Intervention n = 1; Mixed intervention n = 1) are presented in this table.
information on speech prior to intervention and target behavior during intervention for studies where treatment efficacy was reported.

**Generalization**

Twelve (17.4%) studies within this category employed generalization measures. Of these, generalization was most commonly assessed across settings ($n = 5$). Other forms of generalization included generalized outcomes across behaviors ($n = 4$) across persons ($n = 3$) and across stimuli ($n = 3$). Of these studies, three assessed generalization across multiple variables. The outcomes of generalization assessments were largely positive, with 91.7% ($n = 11$) of studies indicating that the generalization of behavioral improvements was observed for all participants, with the final two studies indicating that there were positive outcomes for some study participants but not all.

**Maintenance**

A greater number of studies that implemented behavioral interventions assessed for the maintenance of behavior change (34.8%; $n = 24$). Maintenance was assessed less than two months ($n = 3$; 12.5%); between two and six months ($n = 5$; 20.8%); between six and twelve months ($n = 6$; 25%); and more than twelve months post-intervention ($n = 2$; 8.3%). A further 20.8% ($n = 5$) of studies assessed maintenance outcomes at two or more time periods. Three studies provided data on maintenance outcomes, however, no information was provided regarding the length of follow-up post-intervention.

The results of these maintenance assessments indicated mixed outcomes across studies, however half of the studies suggested positive maintenance outcomes for all participants ($n = 12$). With a further ten studies indicating positive maintenance outcomes for some, but not all participants. Two studies included data that demonstrated either poor, or no maintenance effects for the participants.

**Evidence-based practice**

An assessment of the research rigor of the 69 studies within the behavioral intervention category, using Reichow’s criteria, led to eleven studies (15.9%) being rated as “Strong” and 25 studies (36.2%) being rated as “Adequate”. The majority of studies ($n = 33$; 47.8%) were rated as “Weak.”

Overall, studies employing behavioral interventions achieved a Z score of 416 [(4 * 30) + (8 * 15) + (24 * 4) + (40 * 2)], indicating that this intervention category may be considered evidence-based practice for the treatment of speech deficits among persons with developmental disabilities. Given the evidence-based practice status of behavioral interventions, the various behavioral strategies were also examined, the results of which are presented in Table 1. A number of individual behavioral interventions were identified as being evidence-based, including the utilization of prompting (Z score = 368), reinforcement (Z score = 292), milieu teaching (Z score = 129), manipulating motivating operations (Z score = 110) and behavior shaping procedures (Z score = 132).

**Non-behavioral interventions**

Two studies, including 52 participants, delivered non-behavioral interventions intended to teach speech to persons with developmental disabilities. Of these studies, Dada and Alant utilized a multiple probe design across activities, while Casenhiser and colleagues employed a randomized between groups experimental design.

**Participant characteristics**

**Age**

Participants aged between 2 and 12.1 years ($M = 4.1$ years) participated in the non-behavioral studies.

**Primary diagnosis**

Dada and Alant treated the speech delay of one participant with Down syndrome. Casenhiser and colleagues treated 51 participants with a diagnosis of pervasive developmental disorder.

**Intellectual functioning**

Participants included within Casenhiser and colleagues’ study were of normal intellectual functioning (IQ > 80). While the diagnosis of Down syndrome is typically associated with impaired intellectual functioning, Dada and Alant did not provide information on the participant’s level of intellectual functioning.

**Treatment characteristics**

**Setting**

Of the studies employing non-behavioral interventions, Dada and Alant implemented intervention in a natural setting (i.e., the child’s school). Casenhiser and colleagues did not provide information regarding the treatment setting.

**Intervention agents**

The intervention agent in Dada and Alant was a speech, language and hearing therapist. Casenhiser and colleagues employed multiple intervention agents, including speech and language pathologists, occupational therapists, a clinical social worker and the participants’ own family, to deliver non-behavioral interventions.

**Treatments used**

Dada and Alant employed aided language stimulation which involves the intervention agent pointing to symbols placed on a communication board while simultaneously presenting participants with ongoing spoken language stimulation. This study describes aided language stimulation as an input strategy which is aimed at increasing receptive language abilities, such that individuals exposed to this therapy are provided with more input (i.e., comments or statements) than output (i.e., not emphasizing output or expression from the child). The communication boards focused upon three activities within the participants’ daily lives and intervention was provided within a group setting to all participants simultaneously.

Casenhiser and colleagues employed a social-interaction-based treatment program called MEHRIT (Milton and Ethel
Harris Research Imitative). This program focuses on the development of language via functional social interaction exhibited during play therapy sessions. Within MEHRIT sessions, the topic of interaction is determined by the child’s play preferences. The therapeutic goals outlined within MEHRIT sessions include: (1) enhancing a child’s ability to regulate; (2) enhancing ability to attend to social interactions; (3) increasing engagement in reciprocal interactions; (4) augmenting problem-solving abilities in social interactions, and (5) increasing the ability to use ideas and language in a functional manner. Participants were provided with two hours of therapy per week for the duration of the study, during which families were provided with training on how to facilitate social interaction and communication with their children. These communication acts were coded into nine categories including: commenting, labelling, responding, directing, sharing, obtaining information, rejecting or protesting, social conventions and routines, and spontaneous social expressions. Communication acts were selected based upon two criteria; they were deemed developmentally appropriate for the participant in question, and were also relevant to the deficits exhibited by the participants.

PND
It was possible to calculate PND scores for the study conducted by Dada and Alant. The results indicated that the aided language stimulation program evaluated by Dada and Alant was mildly effective, achieving a PND score of 58.4% (see Table 2).

Generalization
Neither of the studies assessed for the generalization of treatment effects.

Maintenance
Only Dada and Alant assessed maintenance, reporting maintenance outcomes at an unspecified amount of time post-intervention. The results indicated positive maintenance outcomes for all participants.

Evidence-based practice
The study conducted by Dada and Alant was considered to have “adequate” research rigor. The procedures utilized within this study could not be considered evidence-based practice due to insufficient research evaluation. The study conducted by Casenhiser and colleagues, however, was categorized as exhibiting “weak” research rigor.

Mixed intervention
Seven studies delivered mixed interventions, including both behavioral and non-behavioral procedures, intended to teach speech to 193 persons with developmental disabilities. Of these, two utilized SSEDs and five employed alternative experimental designs, including action research designs and randomized group designs.

Participant characteristics
Age
All mixed intervention studies included children aged between 1.8 and 7.3 years as participants (M = 4.3 years).

Primary diagnosis
The majority of participants in these studies had a primary diagnosis of autism (n = 6; 85.7%). The remaining participants included those with a diagnosis of intellectual disability (n = 3; 42.9%), Down syndrome (n = 1; 14.3%), and Fragile X Syndrome (n = 1; 14.3%)

Intellectual functioning
The majority of the studies intervened with participants of normal intellectual functioning (n = 4). Two studies involved the implementation of speech induction procedures for participants with a range of cognitive abilities ranging from severely intellectually disabled to participants of normal intellectual functioning. One study involved the implementation of intervention to an individual with Fragile X syndrome. Fragile X syndrome is typically accompanied by a diagnosis of an intellectual disability. However, Oakes and colleagues did not provide data on the level of intellectual functioning of their participant.

Treatment characteristics
Setting
Of the seven studies included, one failed to provide information on intervention setting. Among the remaining studies, interventions were conducted across clinical settings, including a laboratory that was attached to a University [n = 1], natural settings, including the child’s school [n = 2], home [n = 1], and multiple settings, including the child’s home and school [n = 2].

Intervention agents
Mixed interventions were most frequently delivered by multiple intervention agents [n = 5]. The intervention agents within these studies included parents, a number of specialists including speech and language pathologists, developmental psychologists, occupational therapists, and special educators. A board certified music therapist acted as the primary intervention agent in research conducted by Lim and Draper. Oakes and colleagues utilized family members as the primary intervention agents.

Treatment used
Sandiford et al. provided melodic-based communication therapy to participants. Within this study, 25 target words, and 25 stimulus items based upon the target words were selected. During therapy, participants listened to a CD recording of a target word that was set to a specific melody, the therapist presented the stimulus item to the child simultaneously. Therapy then progressed from listening to a recording of the word set to the melody, to hand over hand clapping of the melody. The next stage of training involved independent clapping of the melody while singing along with the recording
with the clinician. This progressed to singing with the clinician without either clapping, to singing while the clinician mouthed the word silently, to singing the word independently. The final stage of treatment involved the clinician singing a question (i.e., “What is this?”) with the melodic version of the target word, and the participant answering the question independently with the target word.

Bernard-Opitz et al.\textsuperscript{115} utilized computerized visual feedback using the IBM SpeechViewer system. This procedure provides participants with visual feedback regarding their pitch, loudness, and quality of sound productions,\textsuperscript{115} and was combined with both prompting and reinforcement procedures. Within the computer assisted condition, the researcher or parent sat next to the participant and turned off access to the microphone if the participant emitted inappropriate vocalizations or sounds. During the procedure, a picture was presented on the screen, which the parent or researcher then drew the participants’ attention to and provided a vocal model of the sound or word. Participants attempted to produce the targeted sounds within sessions by speaking into a microphone. Throughout the procedure, approximations of the target word were reinforced with praise and on-screen animations.

Chandler and colleagues\textsuperscript{116} utilized musical interaction therapy in combination with reinforcement within their study. Musical interaction therapy involves the therapist singing a commentary on the actions of the participant, which is theorized to contribute to a dialogue and manufacture shared intention and social timing. In addition, the therapist imitates the child in an attempt to encourage reciprocal play. Flexible action songs are also used in an attempt to provide the child with an opportunity to direct, either verbally or non-verbally, what happens. Chandler and colleagues\textsuperscript{116} contend that music provides emphasis, rhythm, ritual, pace, pause, change of mood and increases their willingness to engage, thereby creating a framework which generates communicative dialogue.

Lim and Draper\textsuperscript{117} employed an intervention involving “music ABA verbal behavior” with their participants, a procedure that involved singing verbal instructions to the participant while simultaneously presenting pictures of the target words. Each song lyric contained target words for the participant and presentation of the song lyric was combined with the use of a number of behavioral interventions, including shaping, reinforcement, and prompting procedures. The authors found that the use of music acted as an effective antecedent variable, providing participants with increased motivation to respond, and also functioned as a potent reinforcer enhancing verbal production. By pairing target words with music, it is theorized that vocal verbal behavior is established as a form of automatic reinforcement, thereby increasing the frequency of such vocal behavior among participants.

Van der Schuit and colleagues\textsuperscript{118} provided the KLINc Studio intervention to participants. The KLINc studio intervention was designed to establish the least restrictive environment possible, while simultaneously providing the child with a supportive and appropriate environment for communication. Within this environment, a child is provided with the opportunity to choose and engage in activities that they can engage in independently. Van der Schuit and colleagues\textsuperscript{118} intervention combined a number of methods, including anchored instruction with learning activities that were centered around a theme or event, and interactive storytelling. The authors also employed augmentative and alternative means of communication and computer-based technologies within their intervention, in combination with the Reggio Emilia approach, which is an educational philosophy based upon providing children with a self-guided curriculum within a supportive and enriching environment, and multiple behavioral interventions, including multiple exemplar training.\textsuperscript{122} Participants are then provided with numerous opportunities for the training of both receptive and productive language skills.

Oakes and colleagues\textsuperscript{118} incorporated speech-language pathology strategies within behavioral interventions to facilitate speech with their participant. These strategies included teaching follow-in commenting and interpreting and expanding. Follow-in commenting involves adults within the child’s environment making comments about the child’s activity or exploration with their environment, while interpreting involves the parent interpreting the child’s utterances as meaningful words while expanding involves the parent expanding upon the child’s syllable structures. These procedures were combined with reinforcement and prompting procedures.

Solomon and colleagues\textsuperscript{120} evaluated the efficacy of the PLAY Project Home Consultation model in addition to community services with their participant pool. PLAY is described as a parent-mediated, developmental framework that focuses upon play interactions between the primary caregiver and the child. The PLAY treatment involves coaching, modelling, and video feedback. PLAY services consist of a 3-hour monthly home visit over a period of 12 months. During these visits, PLAY specialists teach the primary caregiver how to identify their child’s social cues and respond contingently to these cues to effectively engage their child in reciprocal exchanges.\textsuperscript{120}

**PND**

It was not possible to calculate PND scores for the majority of the mixed intervention studies due to the experimental design utilized\textsuperscript{116,117,119–121} or a failure to provide data graphically.\textsuperscript{115} However, it was possible to calculate a PND score for the intervention conducted by Oakes and colleagues,\textsuperscript{118} which indicated that the intervention delivered, was ineffective (PND = 16.7%; see Table 2).

**Evidence-based practice**

Of the seven mixed intervention studies, no study was categorized as “strong” based upon the guidelines provided by Reichow’s evidence-based practice criteria. However, three studies were classified as having “adequate” research rigor,\textsuperscript{117,119,120} while the remaining four\textsuperscript{115,116,118,121} were methodologically weak.

None of the procedures utilized by studies in this category\textsuperscript{115–121} could be considered evidence-based practice due to insufficient research evaluation.

**Generalization**

None of the studies implementing mixed interventions assessed the generalization of treatment effects.


**Maintenance**

One study, conducted by Oakes and colleagues,\textsuperscript{118} assessed maintenance of treatment gains three months post treatment. However, the results of this assessment indicated that participants failed to demonstrate maintenance of treatment gains.

**Discussion**

Communication deficits comprise a core characteristic of several developmental disabilities,\textsuperscript{5,10} and negatively impact upon social skills, academic skills, and independent living.\textsuperscript{1,3,123} The current systematic review revealed that a relatively large body of research has evaluated treatments to produce vocal behavior in this population over the past five decades. Our analysis revealed that the majority of research examining interventions to induce vocal speech among individuals with developmental disabilities has evaluated behavioral procedures, most commonly reinforcement, prompting and behavior shaping procedures. Of the three intervention types assessed (i.e., behavioral intervention, non-behavioral intervention, mixed intervention), only behavioral interventions were found to constitute EBP for remediating speech deficits among persons with developmental disabilities according to Reichow’s\textsuperscript{41} evaluative criteria.

In order to provide a comprehensive review of speech induction procedures, the current review included a range of studies spanning over 40 years (1967–2015). Earlier research studies provided the data necessary to calculate treatment efficacy (PND scores), and demonstrated treatment efficacy and research rigor scores similar to more recently published studies. Earlier published studies also employed interventional procedures that are still in use today. For example, the earliest of the studies included in this review, conducted by Hingtgen and colleagues in 1967,\textsuperscript{69} employed a multi-component treatment package including a combination of reinforcement, modelling, prompting, prompt fading, and behavior chaining procedures, all procedures that are still widely used. Thus, as the earlier studies within the current review demonstrated no distinguishable differences in relation to interventional strategies, treatment efficacy, or methodological rigor, there was no sufficient justification to exclude them from the current review, as doing so would have negatively impacted upon the current review’s rigor and comprehensiveness.

A number of interesting findings have emerged from the current analysis. Our results reveal that there exists a variety of effective interventional procedures for speech induction suitable for use with individuals diagnosed with developmental disabilities. Given the positive outcomes associated with the development of speech,\textsuperscript{1,3,4,123} this finding is an encouraging one, as the applications of such procedures may result in improved quality of life for persons with developmental disabilities. While our EBP analysis was limited by the frequent utilization of multi-component behavioral interventions, and the difficulties in making inferences about the efficacy and empirical support for the various intervention components in such cases, the EBP status of behavioral interventions for targeting vocal behavior is evident. These results highlight that clinicians should consider implementing interventions derived from the science of ABA when attempting to aid language acquisition. The knowledge of the average efficacy of each treatment procedure and their individual status as EBP (see Table 1), and the ability to examine specific interventional outcomes in the context of participants’ baseline vocalizations and abilities (see Table 2), may be of use for clinicians considering, or seeking, interventions to remedy the speech deficits of persons with developmental disabilities. Our analysis allowed for the discrimination between effective and ineffective interventions; there exist a number of highly effective behavioral interventions for inducing speech among persons with developmental disabilities including intensive tact instruction, contingent adult imitation, function-based video modelling, echoic training, and high probability behavior sequencing. However, of these “highly effective” treatments, none may be considered EBP based upon Reichow’s\textsuperscript{41} guidelines, although contingent adult imitation and high probability behavior sequencing may be classified as “promising”. This finding highlights the need for further rigorous empirical evaluation of the procedures. Furthermore, the classification of video modelling as “highly effective” based upon PND calculations (see Table 1) should be interpreted with caution. While PND calculations indicated that this procedure is effective in the facilitation of speech within developmentally disabled populations, the generalizability of these results can be called into question, as this procedure was only implemented in one study.\textsuperscript{89} As such, while the results of the current research indicate that this procedure is effective, it currently lacks sufficient evidence to be considered EBP. Further research must be conducted in order to more accurately determine its efficacy and evidence base. Interestingly, our analysis of PND calculations also suggests that a number of commonly used behavioral procedures may be considered ineffective methods for promoting language acquisition including stimulus-stimulus pairing, PECS, and extinction.

Our findings on the utility of PECS for inducing speech or promoting the acquisition of language are also of interest. PECS is one of the most widely used augmentative and alternative communication strategies for persons with developmental disabilities in practice. However, analysis of treatment efficacy of studies utilizing PECS within SSLEDs in the current review revealed that this intervention may not at present be considered an effective method of promoting vocal behavior acquisition (mean efficacy = 34.6%, range 0%–100%). On the other hand, the criteria employed for evaluating research rigor and evidence-based practice within our analysis suggest that PECS may currently be considered an EBP. Previous articles and review papers\textsuperscript{124–126} have suggested that PECS constitutes both an effective and empirically supported treatment. However, these papers have focused primarily upon general communication outcomes, such as symbol requests, spontaneous picture exchange requests and non-vocal social interaction, rather than its effects on vocal behavior acquisition as in the current review. A review conducted by Hart and Banda\textsuperscript{126} noted that the effectiveness of PECS, estimated using PND, for increasing speech was mixed in the research literature. Similarly, Preston and Carter’s\textsuperscript{127} review of PECS as a language intervention suggested that its effects on speech development and language acquisition remain unclear. These results suggest that while PECS may constitute a useful communication strategy for persons with...
developmental disabilities it may not always result in an increase in speech or vocal communication. Future research may wish to examine additional intervention strategies that could be combined with PECS in order to increase the efficacy of the intervention to promote vocal verbal language.

The evaluation of non-behavioral interventions for promoting speech development among persons with developmental disabilities has also yielded some interesting findings. Currently, this body of literature is quite limited and the number of non-behavioral interventions which have been evaluated in research studies is small. Future research including a greater number of participants is required to further our knowledge of the efficacy of these procedures. Furthermore, our analysis suggests that specific non-behavioral interventions such as musical interaction therapy, and melodic based communication therapy combined with behavioral interventions, such as schedules of reinforcement to promote speech production, may be considered promising. However, disentangling the role of motivation and reinforcement within mixed intervention approaches in future research may help identify the active components in such treatments. Substantiation of the findings regarding mixed interventions, including music-based therapies, requires further research, given the limited number of studies currently available to determine treatment efficacy, and corresponding empirical support.

The frequency with which individuals diagnosed with autism participated in the reviewed research evaluations of interventions targeting speech deficits was also notable. The majority of reviewed studies included a participant sample which were primarily composed of individuals diagnosed with ASD (74.4%; n = 58). This is perhaps unsurprising as research suggests that between 25% and 30% of persons with autism fail to acquire speech.\textsuperscript{11,12} Furthermore, other reviews of behavioral research have also indicated an increase in the inclusion of participants with autism or high rates of behavioral intervention with persons with autism.\textsuperscript{42-128,131} However, speech deficits or delays are also highly prevalent among persons with other developmental disabilities such as Down syndrome, intellectual disabilities and fragile X syndrome.\textsuperscript{13,132} The inordinate focus on persons with ASD in the extant body of research is somewhat problematic as research suggests that the learning abilities and responsibility of individuals with ASD may differ substantially from those of individuals with other developmental disabilities. For example, Eldervik and colleagues\textsuperscript{133} noted that children diagnosed with intellectual disabilities may be more responsive to behavioral intervention than their peers with autism for numerous reasons including the greater ease of identifying reinforcers, the greater efficacy of social reinforcement, better ability to sustain attention, lesser engagement in stereotypy, fewer issues with stimulus overselectivity, and the greater likelihood of the generalization of behavior change observed among persons with intellectual disabilities as compared to individuals with autism. Therefore, the findings of research conducted with individuals diagnosed with autism analyzed as part of the current review may not be generalizable to other populations. Future research may wish to examine whether differential treatment outcomes are achieved with different developmental disabilities (e.g., autism, intellectual disabilities, Down syndrome) and whether particular interventions are better suited to certain populations.

Finally, the poor methodological quality of many of the research studies reviewed is also worth noting. Of the 78 published studies reviewed, 50% (n = 39) were classified as being methodologically "weak", based upon Reichow's criteria\textsuperscript{41} for evaluating research rigor, while only 14.1% (n = 11) were classified as being methodologically "strong". The most common variables which affected their methodological rigor included a failure to provide adequate information regarding participant characteristics or the treatment variables, failure to employ a control condition for group studies, and weak baseline conditions for SSEDs. Studies also regularly failed to assess IOA, treatment fidelity or social validity. The measurement of the methodological quality among SSEDs has not received as much interest or attention as it has for other types of experimental designs.\textsuperscript{134} However, issues relating to methodological rigor of SSEDs do not appear confined to this body of literature alone; similar reviews have noted methodological weaknesses within these designs.\textsuperscript{42,135,136} Such findings regarding the general quality of behavioral intervention research are concerning. Rigorous SSEDs are capable of adequately controlling for threats to internal validity, poor implementation of SSEDs may be insufficient to appropriately evaluate training procedures\textsuperscript{137,138} and lead to unreliable findings and conclusions. Future research may wish to further investigate the methodological quality among a broader sample of behavioral intervention studies in order to identify common flaws and provide suggestions for improved conduct in this area.

This current review had a number of limitations. Key among these was that it was not possible to calculate PND scores for all included studies. PND statistics are not applicable to groups designs (n = 23; 29.5%) and a considerable percentage of the studies employing SSEDs (n = 10; 12.8%) failed to provide graphical data from which treatment efficacy could be ascertained. Furthermore, the accuracy of composite PND scores calculated for each intervention type may have been influenced by the number of applications of the treatment. It is likely that the accuracy of composite PND figures increases with the number of applications from which data can be derived. In this way, some treatments which achieved high composite PND scores, such as function-based video modelling, which had a mean PND of 100% resulting from four applications, may have had less reliable composite PND scores than other interventions which performed less well but which were applied more frequently. For example, PECS had a mean PND of 34.6% resulting from 181 applications. Finally, the decision to employ Reichow’s criteria\textsuperscript{41} for evaluating methodological rigor and ascertaining EBP may also be critiqued. Previous reviews have noted the stringency of these criteria\textsuperscript{42,138} which may adversely impact upon the determination on EBP. There exist a multitude of alternative measures for assessing EBP for persons with developmental disabilities\textsuperscript{139-142} and it is possible that the results of our analysis would have differed had we employed an alternate, less exigent set of criteria. Also, Lydon et al.\textsuperscript{42} previously noted that these criteria failed to discriminate between highly effective and ineffective treatments when establishing evidence-based practice. The current analysis therefore addressed this issue by only including effective training procedures, as characterized by their PND scores or by the authors’
evaluation, within EBP analysis. Future research may wish to evaluate this body of research using an alternate evaluation methodology in order to address these issues.

Finally, our adoption of a broad definition of vocal behavior in the current study, which encompasses both functional and non-functional vocalizations, may also be critiqued. Subramanian and Wendt have previously noted the ambiguity of the term “speech” and have called for clarification on whether speech “refers to functional, imitated, or spontaneous compared to non-directed vocalizations that do not carry meaning” (p. 24). However, the exclusion of articles for which it was not possible to ascertain that the targeted vocalizations were functional would have resulted in a much-reduced sample of studies. Furthermore, it would have led to the omission of all studies evaluating the use of stimulus–stimulus pairing procedures, which focus on the acquisition of syllable-consonant vocalizations that may be non-functional during the initial stages of training. However, the facilitation of early vocalizations during stimulus–stimulus pairing is a key step in the development of functional, useful speech as these vocalizations can be successively shaped into meaningful language. Future research may wish to adopt a narrower, more stringent definition of speech for the evaluation of this body of research.

The findings of the current review suggest a number of areas relating to the treatment of speech deficits where additional research is required. While studies employing behavioral interventions may be classified as evidence-based practice, the use of multicomponent interventions within the studies reviewed made it difficult to draw firm conclusions about the EBP status of the individual behavioral procedures. For this reason, and due to the varying treatment efficacy of the behavioral procedures utilized within these interventions (see Table 1), future research including component analyses of behavioral interventions to ameliorate speech deficits would be of interest and would further our understanding of the active or essential components in these treatment approaches. Furthermore, the variable efficacy of the behavioral interventions outlined (ranging from 0% to 100%) suggests that research studies are necessary to examine the variables that predict successful or unsuccessful language acquisition in response to behavioral intervention. A better understanding of what differentiates “responders” and “non-responders” may allow us to develop more effective interventions or may aid in the selection of interventions based on client characteristics. Finally, the current analysis has suggested that there exist a number of interventions (e.g., auditory motor mapping) that appear effective in inducing speech but that have received insufficient research evaluation to date. Future research studies further examining such approaches with a greater number of participants are recommended.

Funding
This research was conducted at The National University of Ireland Galway and Trinity College Dublin, and was supported by the National University of Ireland, Galway Doctoral Research Scholarship Scheme.

Declaration of interest
The authors report no declaration of interest.

References


46. Subramanian A, Wendt O. PECS has empirical support, but limitations in the systematic review process require this conclusion to be interpreted with caution. Evidence-Based Communication Assessment and Intervention 2010;4:22–26. doi: 10.17489/3505341

47. Barrett RP, Sisson LA. Use of the alternating treatments design as a strategy for empirically determining language training approaches with mentally retarded children. Research in Developmental Disabilities 1987;8:401–412*


57. Chezan LC, Drasgow E. Pairing vocalizations with preferred edibles and toys may produce a modest increase in the frequency of vocalizations in three young children with autism. Evidence-Based Communication Assessment and Intervention 2010;4:101–104. doi: 10.1080/17498539.2010.501186


82. Lancioni GE, Singh NN, O’Reilly MF, Sigafos J, Oliva D, Montriont G. Evaluating a computer system used as a microswitch for word utterances of persons with multiple disabilities. Disability and Rehabilitation 2004;26:1286–1290. doi: 10.1080/09638280412331280307


84. LeBlanc LA, Geiger KB, Sautter RA, Sidener TM. Using the natural language paradigm (NLP) to increase vocalizations of older adults with cognitive impairments. Research in Developmental Disabilities 2007;28:437–444. doi: 10.1016/j.ridd.2006.06.004


124. Ganz JB, Davis JL, Lund EM, Goodwyn FD, Simpson RL. Meta-analysis of PECS with individuals with ASD: investigation of


* indicates studies included in the review