Behavior skills training for family caregivers of people with intellectual or developmental disabilities: a systematic review of literature

Xiaohui Sun

Tizard Centre, University of Kent, Canterbury, Kent, UK

People with intellectual or developmental disabilities (IDDs) face lifelong challenges in areas such as self-care, learning and socializing abilities. Having such individuals at home brings family caregivers extra stress, especially when support is insufficient. Because of the inadequacy of professional support available globally, the need to increase the ability of the caregivers to provide effective support is evident. Behavior skills training (BST), an approach consisting of instruction, modelling, rehearsal, and feedback, has been used in numerous studies to train caregivers in necessary skills to help their family members with IDDs. This study conducted a systematic review of past literatures, wherein BST was used to train family caregivers of individuals with IDDs. Seventeen relevant studies were summarized according to their countries, participant characteristics, training focus, intervention details, and outcomes. Outcomes revealed that BST had been used to train caregivers of varying demographics, various target skills (e.g. discrete trial training, incidental training, etc.), order of delivery, and sometimes the introduction of novel components. Most studies showed an improvement for both caregivers and the individuals with IDDs. However, some studies did not present desired outcomes in relation to the individuals. Possible reasons and future measures were also discussed in this review.

KEYWORDS: Behavior skills training, family caregivers, intellectual or developmental disabilities

Introduction

Intellectual or developmental disabilities

Intellectual or developmental disabilities (IDDs) is a broad term covering more than one diagnosis, such as autism spectrum disorder (ASD), intellectual disabilities (IDs), and Down syndrome. According to the DSM-V (American Psychiatric Association 2013), common characteristics of those living with IDDs include possible impairments in learning abilities, restricted interests, and challenging behaviors (CBs). Furthermore, IDDs generally affects individuals throughout their lives and can present obstacles regarding their learning, language development, social interaction, emotion regulation, and self-care. Individuals with IDDs can present additional stress to their families, as there can be higher parental and financial stress (Hsiao 2018), a limiting of attention available to attend to typical developing (TD) siblings from parents (Noonan *et al.* 2018), requirements of extra support to promote positive sibling interactions (Jones *et al.* 2019), and lower overall quality of life for families (Jones *et al.* 2017). A number of studies reported that such parenting stress was largely due to the lack of support and interventions, which restricted caregivers' ability to control over situations and their confidence to help (or teach) these individuals (e.g. Zechella and Raval 2016, Porter and Loveland 2019).

It is difficult to provide a specific number about whether there is sufficient support for this population, as there are a variety of ways of providing support and interventions for those with IDDs (e.g. applied behavior analysis, ABA, Irwin and Axe 2019; neuroscience-based approaches, Cockerham and Malaia 2016; cognitive-behavioral instructional strategies, music-mediated intervention, visual support, Steinbrenner *et al.* 2020). In addition, each of these interventions can provide useful information and have applications that can benefit individuals with IDDs. ABA is a scientific approach to learning and behavior. It seeks to understand environmental variables that reliably impact behaviour and uses this understanding to implement behavior modification techniques (Bloh and Axelrod 2008). There are a

---

Correspondence to Xiaohui Sun Tizard Centre, University of Kent, Canterbury, Kent, UK. Email: sdsunxiaohui2018@163.com

DOI 10.1080/20473869.2020.1793650

number of interventions derived directly from ABA principles and considered as evidence-based practice (EBP) for individuals with IDDs (e.g. discrete trial training, Dart *et al.* 2017).

However, the lack of effective support to these individuals and their caregivers can still be observed in a great number of studies (e.g. Zechella and Raval 2016), regardless of what are the related disciplines. The support provided by behavioral approach is not always sufficient as well. Board Certificated Behavior Analysts (BCBAs), Board Certificated Assistant Behavior Analysts (BCaBAs), and Registered Behavior Technicians (RBTs) are persons with a certification in ABA (received from Behavior Analyst Certification Board, BACB), who use ABA approaches to support individuals with different needs. Given a 1.04% estimated proportion of people diagnosed with IDDs worldwide (Maulik *et al.* 2011) and the total world population (World Data Lab 2019), there is an estimated number of 7,000,000 people diagnosed with IDDs. Compared with the total number of 98,787 BCBAs, BCaBAs, and RBTs (August 2019, BACB), the support provided by ABA approach can be considered as inadequate.

In addition to the general inadequacy, the nature of support related with behavioral approaches varies depending on different regions and subsets of IDDs. Countries of low and middle incomes (annual income: low $\leq$ \$975, \$976 $\leq$ low-middle $\leq$ \$3855) presented the highest rates of IDDs (Maulik *et al.* 2011), yet the number of BCBAs in these countries was significantly lower than that of developed countries (BACB 2019). In practical application, an overwhelming majority of BCBAs work with clients who have an ASD diagnosis as compared with other specific diagnoses (Milyko 2016). Thus, effective training available for caregivers on efficacious interventions could be critical in supporting individuals with IDDs and reducing parental stress.

This article focuses on behavioral approaches to support individuals with IDDs and their family caregivers. However, there are other types of interventions available for the individuals and their caregivers. For example, Gillam *et al.* (2015) successfully trained five children diagnosed with ASD (from 8 to 12 years of age) to tell stories through *narrative intervention*. Narrative intervention is based on cognitive development and aims at increasing participants' ability to identify causal relationships, understand emotional and mental states of other people, and to rationalize information via different phases of storytelling.

Another type of intervention available for people with IDD's and their family is art therapy, which is gaining more attention as a treatment to improve individuals' imagination, abstract thinking, and the ability to organize different elements into a same topic (Martin 2009). Studies also indicated that it increased social ability and cognitive knowledge at the same time (e.g. Got and Cheng 2008). There were no such fixed training procedures reported in the use of art therapy, and there were a variety of activities used throughout. As highlighted by Martin (2009), the approach involved object relations, developmental approaches, behavioral approaches, and psychotherapy.

Emery (2004) applied art therapy to train a six-year-old boy with ASD and attained positive outcomes. The training involved teaching the participant a variety of arts-based activities by instructing the participant to complete different stages of playing with play dough (e.g. make a ball and flatten in to a circle, make into particular shapes) and drawing activities (e.g. draw disconnected figures, draw complete figures). In addition to art skills, researchers noted the improvement in the participant's other abilities during the practice such as independently learning that the figures he drew were 'people'. His social skills also improved, including tone of voice, eye contact, initiation of conversations, exhibiting the caring for others' feeling, and understanding of humour.

In addition to ASD, art therapy has also attained positive outcomes in intervening individuals with other diagnosis. For example, Got and Cheng (2008) have successfully applied art therapy to increase social skills and language comprehension for people with developmental disabilities. Beh-Pajooh *et al.* (2018) reported that it improved the externalizing behaviors for individuals with IDs.

Occupational therapy with sensory integrative approach was developed based on the theories of neuroscience, developmental psychology, occupational therapy, and education (Schaaf and Miller, 2005). It aims to improve individuals' response to sensory input by using sensory activities and modification of environment to further increase the ability to engage in play activities and social interaction during play (Williamson and Anzalone 1997). Researchers noted that appropriate level of arousal, orientation, and attention were necessary for individuals to appropriately interact with environment (Case-Smith and Bryan 1999). Case-Smith and Bryan (1999) used this intervention with five boys with ASD (mean age of 5 years old). The training included vestibular stimulation activities (e.g. swings, slides) and tactile input activities (e.g. brushing, tent). Results demonstrated an increase in goal-directed play (play in the way fitting with a game), pretended play, interaction with adults and peers, and a decrease in stereotypic behaviors and basic sensory exploration of toys. The number of related studies is currently limited, however, it has generated positive outcomes in the intervention for individuals with pervasive developmental disorders (Case-Smith and Miller 1999), ASD (Watling and Dietz 2007), and learning disabilities (Polatajko *et al.* 1991).

There are other interventions for individuals with IDDs that are commonly applied as well, including language-speech therapy (Terband *et al.* 2018), play-based intervention (Grant 2017), and medical treatment (McPheeters et al. 2011). In order to generate an exhaustive understanding of one particular type of training, this study only focused on behavioral intervention. Therefore, the training for caregivers, and the skills they were taught to train individuals with IDDs were both under the theory of behavior analysis.

Behavior skills training

Behavior skills training (BST) is an approach that typically includes instruction, modelling, rehearsal, and feedback. This approach has been widely adopted in the field of ABA. Miltenberger (2004) described each component and the application of this approach. Typically, instruction includes providing detailed descriptions of the target behavior, its antecedents, and its consequences. For behavior chains, trainers illustrate each behavior in an appropriate sequence. Miltenberger (2004) further emphasized the need for language that is easily understood and well-trained implementors to increase the effectiveness of instruction. Modelling refers to the demonstration of appropriate behaviors and is conducted either in person or via video. A number of factors can increase the efficacy of modelling, such as outcomes corresponding to target behaviors, models similar to (e.g. same age) or holding a higher status than the target participant (e.g. teachers tending to be more effective models than peers), and naturalistic settings or exemplars (e.g. in real situations, role-played real situations). In the rehearsal phase, trainees are allowed to practice the target behaviors after instruction or modelling. The feedback phase is the opportunity to reinforce the desired performance, correct errors, and provide further instructions. Typically, after delivering instructions, trainers repeat the model-rehearsal-feedback sequence until the trainees meet a mastery criterion (Miltenberger 2004).

Past studies have suggested the efficacy of using BST to train both individuals with IDDs and their support providers (e.g. tutors, caregivers, staff, parents). For example, Morgan and Wine (2018) successfully used BST to teach job skills to an 18-year-old student diagnosed with ASD, enabling him to work in a restaurant. They conducted BST in the conventional order, and repeated the last two steps (i.e. rehearsal and feedback) until the student achieved perfect implementation in all the steps for three consecutive sessions. The researchers then conducted post-training sessions in the same way as baseline sessions, where they prompted the student to conduct the target skills independently. Compared with the 0% correct rate (average of all target skills) during baseline, the student made significant improvement and was scored 99.6% (average; ranging from 95% to 100%) on completing the steps correctly.

Another researcher examined the application of BST to non-family caregivers of individuals with IDDs. In one study, Gormley *et al.* (2019) randomly assigned 54 frontline staff from six different institutions into two groups: One treatment group and one control group. Participants in the treatment group were taught by way of lectures and manuals about various topics pertaining to reinforcement, task analysis, prompting, and functional communication training (FCT). Furthermore, participants had the opportunity to observe effective modelling techniques via video and rehearse/practice scenarios with individuals role-playing the part of someone with IDD. The interactive role-play scenarios were accompanied with feedback from experienced trainers. No additional training was provided in the control group. The results revealed that after the training participants in the BST treatment group showed significantly higher scores on quizzes of ABA knowledge as well as better practical implementation of the trained skills.

In addition to studies focusing on frontline staff, there are also a number of studies indicating the potential for family caregivers to teach useful skills or change undesired behaviors at home. Results of these studies were promising. They indicated that parents became more helpful in terms of teaching communication skills (Wilkes and Lincoln 2018), reducing sleeping problems (Kirkpatrick *et al.* 2019b), and reducing disruptive behaviors (Postorino *et al.* 2017). In addition, their children with IDDs also showed improvement in skills or reduction in problem behaviors. Despite this, to date there is no systematic review examining the effectiveness of BST for family caregivers. Furthermore, compared with caregivers of TD children, BST would be more helpful for caregivers of individuals with IDDs, as individuals with IDDs often require additional interventions to gain skills or modify behaviors. For example, Jacobson *et al.* (1998) indicated that early intensive behavioral intervention (EIBI) was helpful for individuals with Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS) to obtain skills within a normal range, including intellectual, learning, social, and self-care abilities. This illustrates the requirement for implementing intervention more often. Own to the inadequacy of behavioral professionals (as mentioned previously), it would be helpful for caregivers to increase their skills to provide effective intervention, such as high procedure fidelity and generalization methods.

Instead of covering a broad scope by combining studies that involve non-family caregivers (e.g. service staff, personal care workers, teachers, volunteers, etc.), this review focused on family caregivers with the rationale that family caregivers have distinct differences

Table 1. Search terms of using BST to train family caregivers.

Group 1	behavio* skills training OR BST
Group 2	family member* OR famil* OR parent* OR sibling* OR grandparent* OR carer* OR caregiver* OR caretaker*
Group 3	autis* OR ASD OR autis* spectrum disorder OR Asperger* OR pervasive developmental disorder OR PDD-NOS OR high functioning autism OR Rett OR intellectual disab* OR developmental disab*

BST behavior skills training; ASD autism spectrum disorder; PDD-NOS pervasive developmental disorder – not otherwise specified.

from non-family caregivers. First, family caregivers have different knowledge regarding the individuals (e.g. habits and developmental history) and may also possess more knowledge regarding family dynamics in general (e.g. daily routines, Raulston *et al.* 2019). Furthermore, the nature of motivation for family caregivers and service staff likely differs in whether they are paid to provide the support (Kossyvaki *et al.* 2016), whether the development of IDDs has significant influence on them (Raulston *et al.* 2019), and whether they are under the same kind of supervision (e.g. laws, supervisors, Kossyvaki *et al.* 2016).

Current study

The current study conducts a systematic review of BST interventions to train family caregivers of individuals with IDDs. The researcher is seeking to gain a comprehensive understanding of: (a) the trainees involved in the studies (e.g. demographics) and their relationships with IDD individuals, (b) how BST was conducted and whether there existed any special components in the studies, (c) the results of BST, and (d) suggestions for future studies. With this investigation, the researcher hopes to provide potential resources for practical training for family caregivers so they are able to support individuals with IDDs.

Method

Inclusion criteria

This review only included studies meeting specific criteria. First, studies must use BST to train caregivers of individuals with IDDs (rather than the individuals themselves). These caregivers will henceforth known as trainees. Second, studies could only be included if they were published in peer-reviewed journals (i.e. no theses, dissertations, or unpublished manuscripts). Third, studies must be empirical in nature (i.e. no meta-analyses or systematic reviews). Fourth, studies must provide adequate information pertaining to procedure and intervention data to determine the treatment effectiveness and sometimes maintenance (e.g. order and specific means of different BST components, performance of trainees during different phases of training). For example, Lanovaz *et al.* (2016) was included, where the following information was provided: (a) researchers conducted BST in an instruction-modelling-rehearsal-feedback order, (b) instruction was given by oral description and written manual, and (c) no confederates were introduced and the trainees modelled directly with individuals with IDDs.

Further, it should be noted that the studies of interest included a variety of family caregivers. Although parents were generally the most common (as can be seen in the ‘Result’ section), studies were included where the trainees were non-parent caregivers. Other family members of people with IDDs and those living or working within the home who had at least some amount of interactions with the affected individuals were included as well, thus, the review also included grandparents, siblings, nannies, and adoptive parents. Service providers who already catered to individuals with IDDs (such as therapists and volunteers) were not included.

Studies were qualified as long as they met the criteria discussed above. There was not a criterion set regarding when and where the study was published. There were also no restrictions in relation to the length of studies or different phases (e.g. baseline, training, post-training sessions), as long as all these three components were included to ensure it was possible to compare participants’ performance before and after the training. Studies were not excluded according to the specific learning outcomes (i.e. performance of trainees), as the review is aiming to learn from and provide reference for both the successful and failed experiences of past studies. Studies were also not excluded based on demographics (i.e. any country), specific diagnosis (i.e. any diagnosis of IDD qualified) or specific familial relationship (i.e. any familial relationship qualified) of the individuals with IDDs.

Search strategy

The search procedure involved three phases: Database search, journal search, and reference-list search. Database searched included PsycINFO, Scopus, and Web of Science. Three groups of search terms were inputted (see Table 1) into the databases, the abstracts were screened for each of the search outcomes, and finally the researcher read articles in entirety to determine whether to include each study based on the aforementioned inclusion criteria.

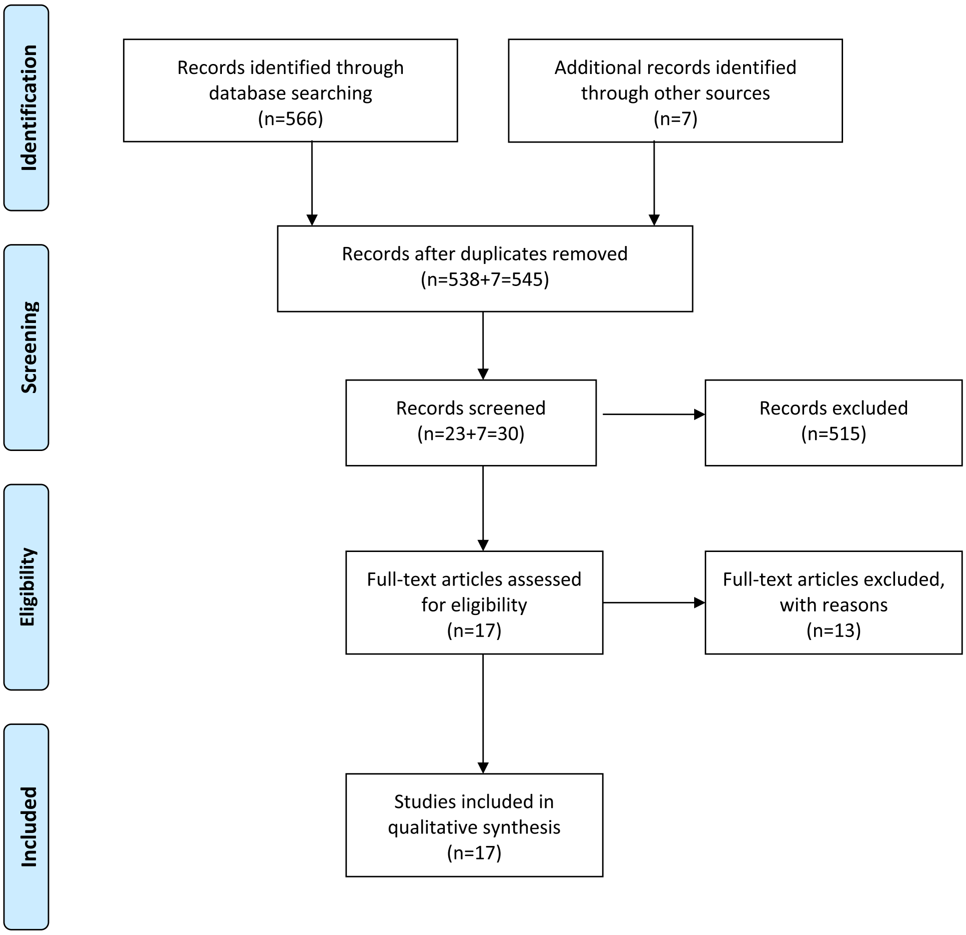

Figure 1. Search procedure and number of articles gained in different phases.

The three groups of search terms (Table 1) referred to the text appearing in the abstracts of the studies. The first group generated a list of all articles that included the phrase 'behavior skills training', 'behavior (American spelling) skills training', or 'BST'. The second group further refined the list by returning only studies related to some kind of family caregivers. Finally, the third group restricted studies to those that involved all known IDDs. As shown in Figure 1, the database search returned 566 articles, 538 of which remained after the removal of duplications. Titles and abstracts of these articles were screened, and 23 articles were retained.

The second phase involved a journal search. Based on the database search, the researcher noticed that the related studies (in database search) tended to be published in roughly the same set of journals. Among the 23 articles returned in database search, the three most common journals were the *Journal of Applied Behavior Analysis (JABA)*, *Journal of Autism and Developmental Disorders (JADD)*, and *Research in Autism Spectrum Disorder (RASD)*. The researcher examined the titles and abstracts for all articles published in these three journals (dated back to 1968 for JABA, 1971 for JADD, and 2007 for RASD, when each journal was first established) in chronological order and assessed for their relevance to the present review. On the basis of this second search strategy, the researcher realized six more articles might be related with this review, which were therefore retained for final decision. In the last phase, the researcher went through the reference section of each article previously collected (in the database and journal search) in an attempt to find more relevant articles. The results of this search strategy yielded one additional study.

Thus, the three-phase search resulted in 30 articles. Each of these articles were reviewed in their entirety and then compared against the aforementioned inclusion criteria. Following this search, the searching process was terminated with 17 articles for this systematic review.

Quality appraisal

To determine whether the BST for caregivers was an EBP, the researcher applied the Evaluative Method to evaluate the methodological quality of each included study (Reichow *et al.* 2008, Reichow 2011). Wendt and Miller (2012) found that the Evaluative Method was the most rigorous out of seven commonly used quality appraisal tools for single case design. It was rated highly in terms of its congruence with previously agreed criteria (e.g. Horner Standards, Horner *et al.* 2005),

ability to clearly identify different quality levels (i.e. strong, adequate, weak), and it's differential weighting of quality indicators (i.e. primary indicators—such as whether the independent variable was defined—are considered more important than secondary indicators—such as whether social validity was included). Moreover, the Evaluative Method is the only quality appraisal tool applicable for both single case design and group design. Since there were both single case design and group design studies, this tool was selected to determine methodological quality in this review.

Each article was categorized as strong, adequate, or weak based on a review of their primary indicators (e.g. participant characteristics, independent variables, dependent variables, baselines, etc.) and secondary indicators (e.g. blind raters, fidelity, generalization/maintenance, social validity, etc.). A detailed description of these indicators and rating criteria are available in Appendix B. The Evaluative Method tool was applied once for each included article, targeting at the information related with family caregivers (rather than individuals with IDDs).

Coding

Each article that was included was read from the beginning to the end, during which notes were taken that included a variety of information. A data extraction sheet was developed (Appendix A) to classify information of the articles under different topics to investigate whether there was something common, special, and/or important. These topics included: (a) the number of caregivers and individuals with IDDs; (b) demographic information of caregivers involving age, gender, relationship with individuals with IDDs, education, occupation, and income; (c) demographic information of individuals with IDDs involving age, gender, and diagnosis; (d) specific skills focused in the studies; (e) study design; (f) intervention details including the order of BST components, procedure for delivering the introduction, whether non-examples were included, whether role-play of children was included, how feedback was provided, whether there were any special intervention components (e.g. booster sessions), and mastery levels to move to next intervention phases and/or finish the intervention; (g) intervention outcomes including the performance of caregivers and individuals with IDDs immediately after the training; (h) generalization details including when and how the generalization probes were conducted, and the performance of caregivers and individuals with IDDs persons; (i) social validity measures; and (j) limitations and possible future studies.

Inter-rater agreement

To assess the reliability of article inclusion and quality appraisal, a graduate student was brought on as an independent rater. The rater was familiarized with the aforementioned inclusion criteria and was given 250 articles (45.87% of the total papers yield by database search) that were randomly selected from the database search results. The rater read through the titles and abstracts and decided whether each article met the established criteria. The inter-rater agreement was presented by the percentage of agreement, which was the ratio of the number of agreed articles to the total number of reviewed articles (i.e. 50). Out of the 250 given articles, the researcher found six articles that met the criteria, and the rater reported a result of seven. There were six articles that were selected by both the researcher and the rater. Therefore, the inter-rater agreement of article inclusion was 99.6%.

The rater was trained to use the quality appraisal tool (i.e. Evaluative Method) to check the methodological quality. They independently read 52% of the included articles (9 of 17) and applied the tool to the articles targeting at information related with caregivers (see Table A5, Appendix C). The inter-rater agreement was presented by the ratio of the number of agreed upon articles to the total number of the reviewed articles. The calculation of the inter-rater agreement of all primary indicators, secondary indicators, and final strength ratings were 81.48%, 94.44%, and 66.67%, respectively. To remediate any discrepancies, the articles were reviewed again by both the researcher and rater until an agreement was reached. The final decisions can be seen in Table A6 (Appendix C).

Results

Quality appraisal

Twelve studies were rated as weak (70.59%), four as adequate (23.53%), and one as strong (5.88%). An overview of the ratings for each indicator can be found in Table A6 (see Appendix C). The reviewed studies included 16 single case designs and one group design.

Within the 12 studies rated as weak, four included one unacceptable primary indicator, while the other primary indicators were all rated as high or adequate. The single case designs often did not provide detailed descriptions of interventions. This resulted in four studies (around 23.53%) being rated as unacceptable with regards to participant characteristics. All studies included generalization or maintenance probes and most studies had inter-observer agreements (16 of 17 cases) appropriately measured.

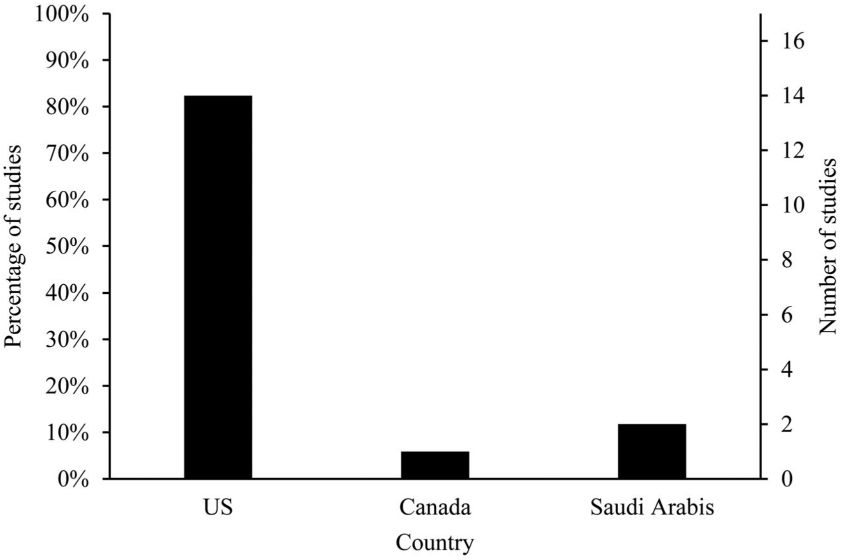

Figure 2. **Country of included studies.**

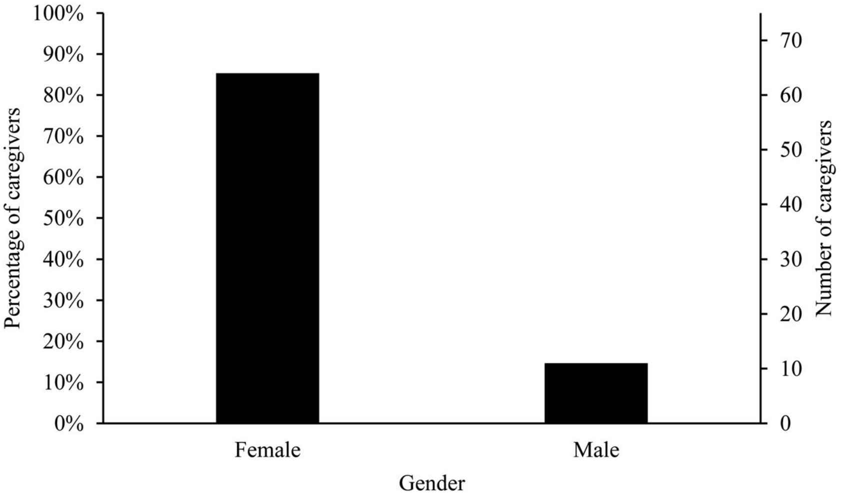

Figure 3. **Gender of caregiver participants.**

Context

An overview of the 17 included studies can be found in Appendix A, where important information for each study was presented. Results will be discussed according to the areas relevant to the research purpose including: (a) country, (b) participant characteristics, (c) training focus, (d) intervention details, (e) outcomes of post-training sessions, (f) generalization, and (g) social validity. These areas provide an overview of the components typically included in these studies, which are also useful to provide reference for researchers or practitioners to implement BST for caregivers.

Country

As suggested in Figure 2, most of the studies (i.e. 14 of 17) were conducted in the United States. Of the remaining three studies, one was conducted in Canada (Hassan *et al.* 2018), and the other two were conducted in Saudi Arabia (Eid *et al.* 2017a, Eid *et al.* 2017b).

Participant characteristics

A total of 81 caregivers and 71 individuals with IDDs were involved in the 17 studies. The average number of caregiver participants per study was 4.76 (sd = 3.71), with a mode of 3. In addition, the average number of individuals with IDDs per study was 4.18 (sd = 3.85), with a mode of 3 as well.

Caregiver participants

This section presents the information of caregivers with respect to their (a) gender, (b) relationship with individuals with IDDs, (c) age, and (d) socioeconomic status (i.e. education and income level). A total of 16 studies reported the gender of caregiver participants (expect Eid *et al.* 2017b, where 6 caregivers were included). There were 75 caregivers whose gender were reported, including 64 females and 11 males (see Figure 3).

The types of relationships the studies involved included: (a) parents (50 mothers, eight fathers, and six

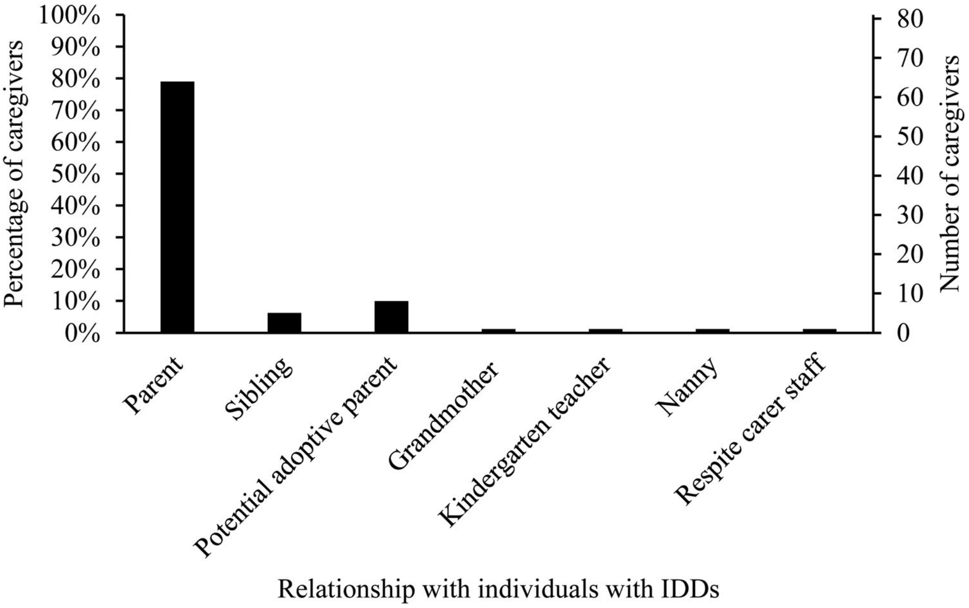

Figure 4. Relationship between the included caregivers and individuals with IDDs.

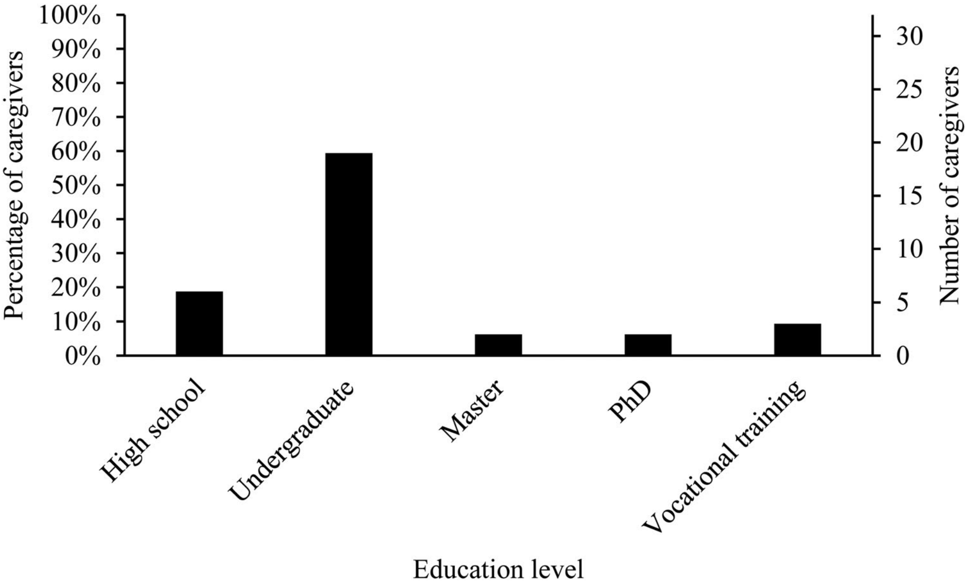

Figure 5. Education level of caregiver participants.

parents without gender reported), (b) siblings (one brother, four sisters), (c) potential adoptive parents (two potential adoptive fathers and six potential adoptive mothers), (d) a grandmother, (e) a kindergarten teacher, (f) a nanny, and (g) a respite caregivers staff (see Figure 4). The last three caregivers (i.e. kindergarten teacher, nanny, respite care staff) were included because they were not paid directly for providing special service to individuals with IDDs and had similar knowledge of their developmental history as typical family caregivers. There was also a case in which the precise relationship between the caregivers and the individual(s) with IDDs was ambiguous (Eid *et al.* 2017b did not report the genders of the parents).

There were 12 studies reported specific ages for only 30 of the caregivers. Amongst these caregivers, ages ranged from 24 to 68 years ($m = 39.33$, $sd = 8.56$). Some studies did not report specific ages for their caregivers, but reported age ranges. This applied to 25 caregivers, of which 18 ranged from 26 to 47 years old, two from 30 to, four from 37 to 47, and one from 60 to 65.

The highest level of formal education achieved was reported in 12 studies, for 32 caregivers. Figure 5 showed that these levels consisted of: (a) high school diplomas ($n = 6$), (b) undergraduate degrees from college/university ($n = 19$), (c) postgraduate degrees including Masterversity (d$n = 2$) or PhD ($n = 2$), and (d) trade/technical/vocational training ($n = 3$).

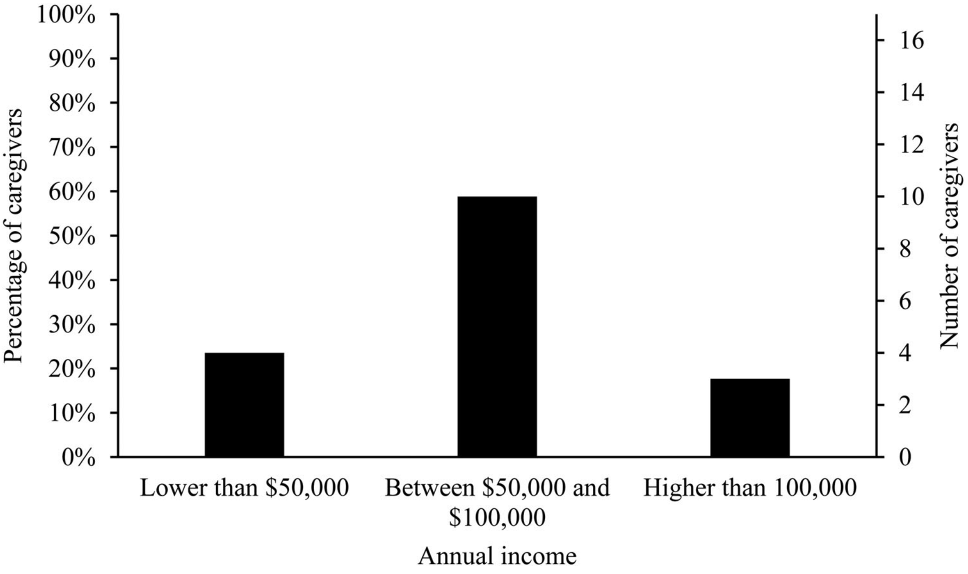

Figure 6. Income level of caregiver participants.

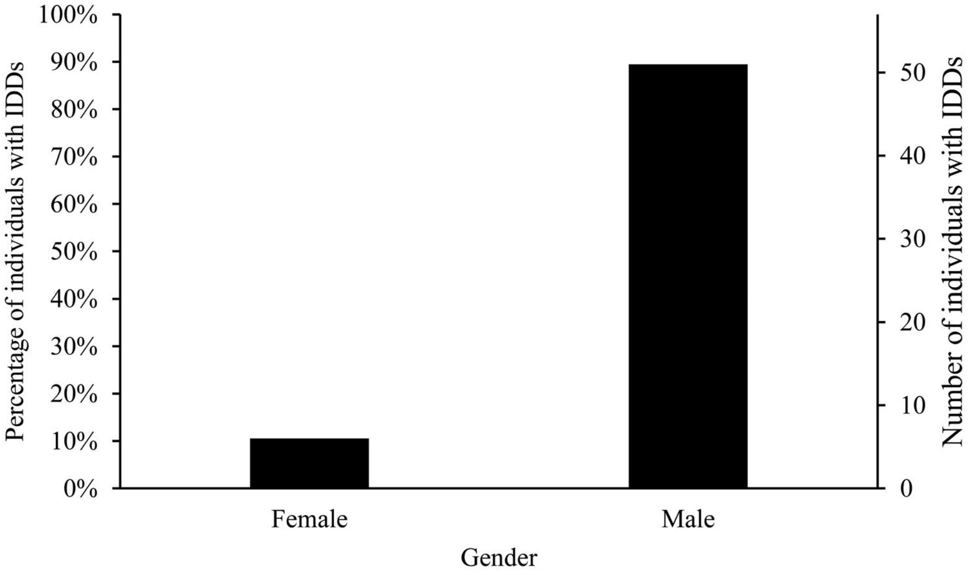

Figure 7. Gender of individuals with IDDs.

As demonstrated in Figure 6, only five studies reported 20 caregiversining (degree (iplomas (studies, for 32 caregivers. es. rs, ages ranged from r, ten earned between $50,000 and $1,00,000 per year, while three earned over $1,00,000 per year. In addition, authors of one study conducted in the US (Lafasakis and Sturmey 2007) reported that three participants had low income, without range specified.

Participants with IDDs

With regards to participants with IDDs, an analysis of their gender, age, and diagnosis follows. As suggested in Figure 7, the 71 individuals with IDDs, gender was indicated in 57 cases (in 14 studies) such that individuals with IDDs were predominantly male ($n = 51$) rather than female ($n = 6$).

Age was reported in 45 cases (also in 14 studies) such that participants had a mean age of 6.71 years (sd = 3.96). In some cases, an age range instead of specific age was provided. Four participants were reported as being from six to eight years (Hassan *et al.* 2018) while 18 were reported to be 18 months (Matthews *et al.* 2018) old on average.

Specific diagnoses were reported for 43 individuals in 14 studies, with details demonstrated in Figure 8. The most prevalent diagnosis was ASD, as 37 individuals were reported to have this diagnosis. Of these 37 individuals with ASD, 33 were reported to have an ASD diagnosis alone, while the remaining four were reported to have another condition present—either, brain injury, severe ID, Attention Deficit Hyperactivity Disorder (ADHD), or Phelan-McDemid Syndrome. In addition to ASD, there were also individuals who were

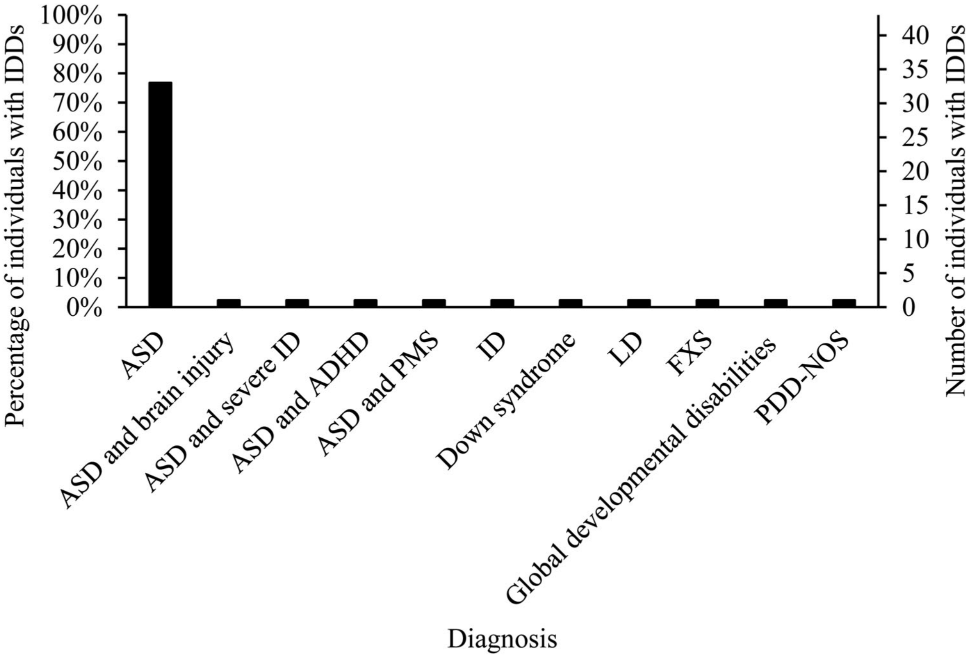

Figure 8. Diagnosis of individuals with IDDs.

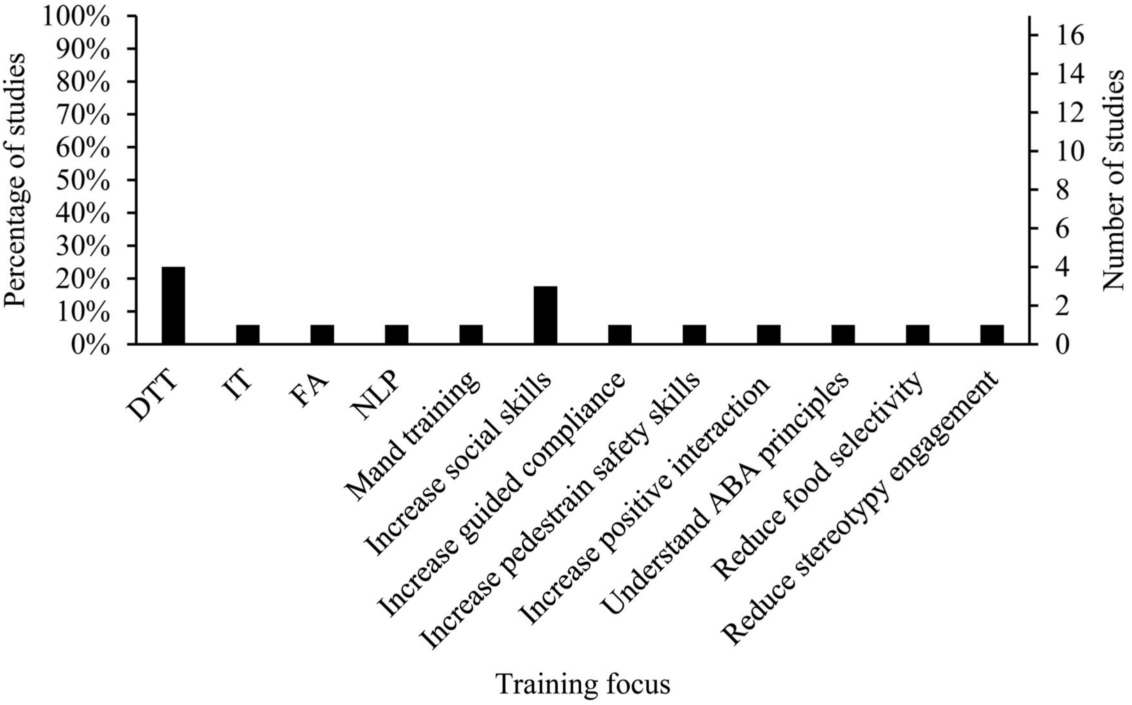

Figure 9. Training focus of included studies.

diagnosed with ID, Down syndrome, learning disabilities, Fragile X syndrome, global developmental delay, and PDD-NOS, with one person diagnosed with each of these disabilities.

Training focus

The training focuses were specific techniques or training targets. As presented in Figure 9, specific techniques included: (a) discrete trial training (DTT, Lafasakis and Sturmey 2007, Ward-Horner and Sturmey 2008, Eid *et al.* 2017a, Subramaniam *et al.* 2017), (b) incidental teaching (IT, Hsieh *et al.* 2011), (c) functional assessment (FA, Shayne and Miltenberger 2013), and (d) natural language paradigm (NLP, Eid *et al.* 2017b). The training targets aimed at improving certain skills or reducing undesired behaviors. They were typically achieved by a series of steps designed by researchers, consisting of: (a) mand training (Loughrey *et al.* 2014), (b) increasing guided compliance (Miles and Wilder 2009), (c) increasing social skills of individuals with IDDs (Stewart *et al.* 2007, Dogan *et al.* 2017, Hassan *et al.* 2018), (d) increasing pedestrian safety skill of individuals with IDDs (Harriage *et al.* 2016), (e)

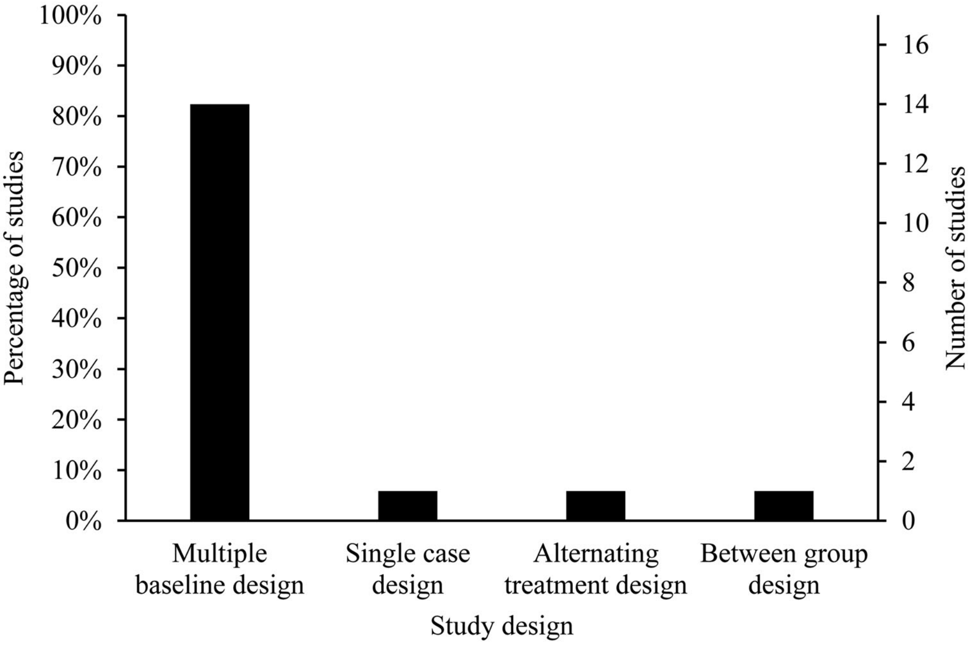

Figure 10. Study design of included studies.

increasing positive interaction between individuals with IDDs and TD siblings (Kryzak and Jones 2017), (f) reducing stereotypy engagement (Lanovaz *et al.* 2016), (g) reducing food selectivity (Seiverling *et al.* 2012), and (h) understanding ABA principles and procedures (Matthews et al. 2018). As will be seen, the most frequent training focus in relation to specific technique was DTT (four studies) and in terms of training targets was to increase social skills of individuals with IDDs (three studies).

Intervention details

The operation of BST across studies included different study designs, orders of BST components, instructions, various use of confederates, and special intervention components.

Study design

Figure 10 displayed that a multiple baseline design was the most common study design and was applied in 14 of 17 studies (e.g. Lafasakis and Sturmey 2007). In addition, each of the following designs was observed exactly once: Single case design (Stewart et al. 2007), alternating treatment design (Lanovaz *et al.* 2016), and comparison with controlled groups design (between group design, Matthews *et al.* 2018).

Order of BST components

The order of the four conventional components of BST (i.e. instruction, modelling, rehearsal, feedback) varied between studies (see Figure 11). Researchers typically delivered instruction once and repeated the other components until caregivers attained certain criteria. The instruction-modelling-rehearsal-feedback (I-M-R-F) sequence was observed as the most common order, which was applied in 11 (of the 17) studies (e.g. Seiverling *et al.* 2012). Moreover, there were three slightly different variations, including: (a) instruction-rehearsal-feedback-modelling (I-R-F-M, Ward-Horner and Sturmey 2008, Miles and Wilder 2009), (b) I-R-M-F (Hsieh *et al.* 2011, Loughrey *et al.* 2014), and (c) instruction-rehearsal-feedback (I-R-F, Shayne and Miltenberger 2013). One special case was Matthews *et al.* (2018), where caregivers were only provided one trial of instruction-modelling, and then were required to rehearse at home without any observation or feedback (I-M-R).

Instruction

The delivery of instructions varied between studies with respect to: (a) the ratio of trainers to trainees, (b) the use of technology, and (c) whether baseline performance was discussed prior to the instruction phase of BST. In some cases, a single trainer delivered instructions to a single trainee (e.g. Eid *et al.* 2017b) while in other cases, a single trainer delivered instructions to an entire group (e.g. Lafasakis and Sturmey 2007). Technologically based visual aids (e.g. PowerPoint as seen in Loughrey *et al.* 2014) were sometimes used, while sometimes instructions did not rely on as much technology and instead incorporated verbal instructions and printed handouts (e.g. Hsieh *et al.* 2011). Lastly, in some cases, participants were required to perform the target skills independently during baseline sessions. When giving instructions, trainers discussed their baseline performance and provided some feedback (e.g. Ward-Horner and Sturmey 2008).

Confederates

Caregivers in five studies rehearsed the target skills with confederates before meeting a mastery level. In the studies

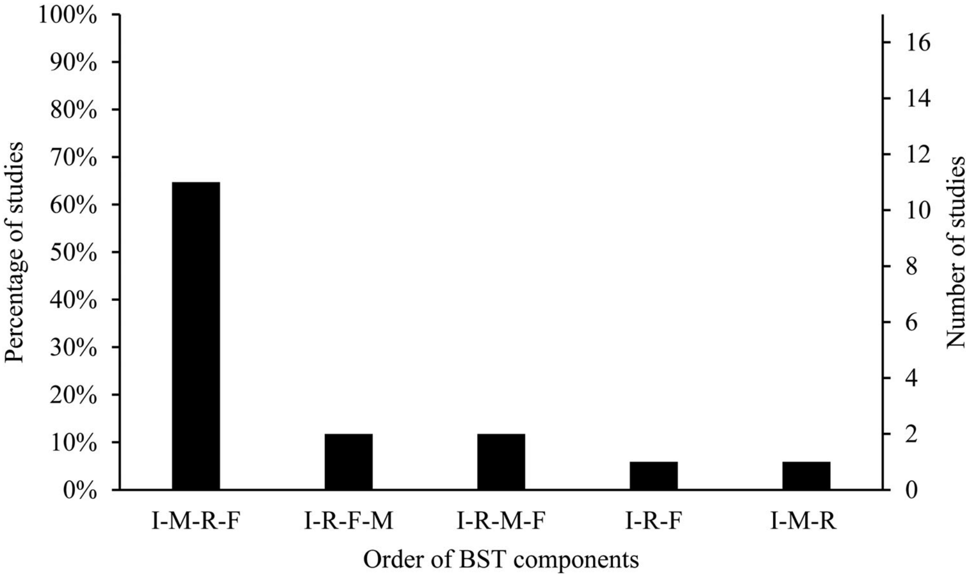

Figure 11. Order of BST components applied in studies.

reviewed, confederates included undergraduate students (Stewart *et al.* 2007), graduate students (Miles and Wilder 2009), the author of the investigation (Kryzak and Jones 2017), and other trainers without clear descriptions (Harriage *et al.* 2016, Subramaniam *et al.* 2017). Some studies involved confederates in a different way. For instance, the primary trainer modelled with a secondary trainer (i.e. the confederate) and caregivers directly rehearsed with individuals with IDDs, rather than with the confederates (Shayne and Miltenberger 2013).

Special intervention components

In some cases, additional components were added to conventional ones. In these studies, improvement was also observed in the performance of caregivers. Although this does not necessarily mean that such components were effective, they did illustrate alternative ways to provide training.

Loughrey *et al.* (2014) made use of peer-training. A mother was trained by researchers with BST directly, and after meeting the mastery level, she trained her husband using the same method with which she was trained. During assessment, both individuals showed improvement. Hassan *et al.* (2018) separated participants into model parent and observer parent groups. Model parents were trained by conventional BST, while observer parents learned by observing the training of the other group (the same handouts were provided to each group illustrating correct implementation). Results suggested that parents in both groups exhibited improvement of the target skill. Dogan *et al.* (2017) added special training sessions for caregivers who failed to meet the mastery criteria. After post-training sessions, the authors introduced these caregivers into training booster (T-B) sessions. In the T-B sessions, caregivers rehearsed with the instructor and received feedback, who played the role of individuals with IDDs. If caregivers still failed after two rehearsal-feedback rounds, self-monitoring procedure was conducted. In this procedure, all steps of the taught skill were listed, and caregivers checked whether they implemented each step correctly.

Termination criteria

The criteria upon which investigators decided to terminate training varied between studies. Most commonly, termination was deemed appropriate when caregivers implemented 80% of the target skill correctly (e.g. Shayne and Miltenberger 2013, Harriage *et al.* 2016, Lanovaz *et al.* 2016, Dogan *et al.* 2017). Similarly, some studies terminated training once caregivers achieved a 90% (Eid *et al.* 2017b), 95% (Eid *et al.* 2017a), 100% (Kryzak and Jones 2017) correct implementation. Another common criterion was based on the percentage of correct implementations for a number of consecutive sessions. These percentages ranged from 80% to 100% for typically two to three consecutive trials. The investigation by Matthews *et al.* (2018) involved a unique case in which training was simply terminated when all content was taught to the trainees without any requirements for the caregivers' implementation.

Training outcomes

During the post-training sessions (i.e. those that took place immediately after training), all caregivers' performance was recorded, and the results suggested improvement (Figure 12).

As indicated in Figure 13, the performance of individuals with IDDs was reported in 15 studies, 11 of which indicated improvement for all individuals involved. The other four studies showed improvements for some but not all individuals, including one of three

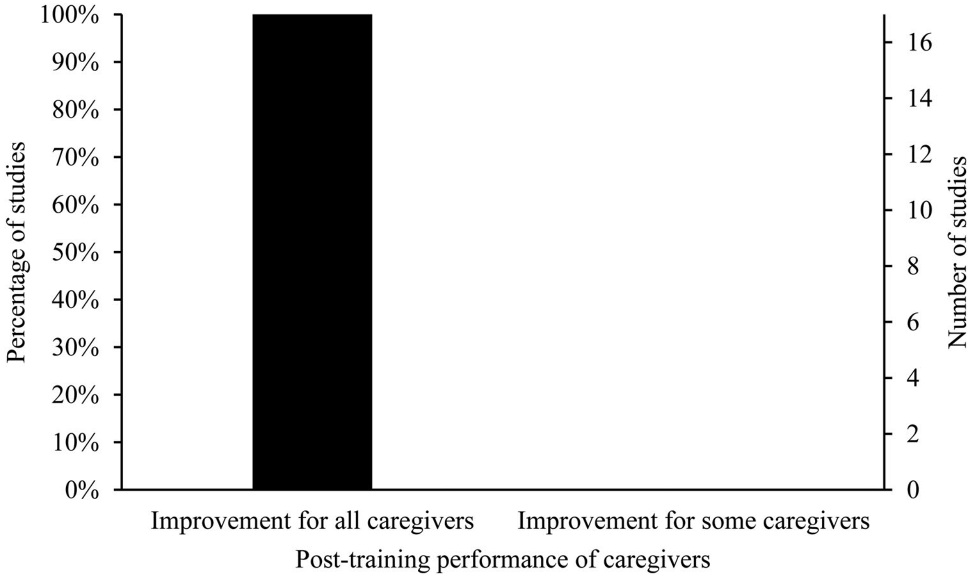

Figure 12. **Training outcomes of caregivers in studies.**

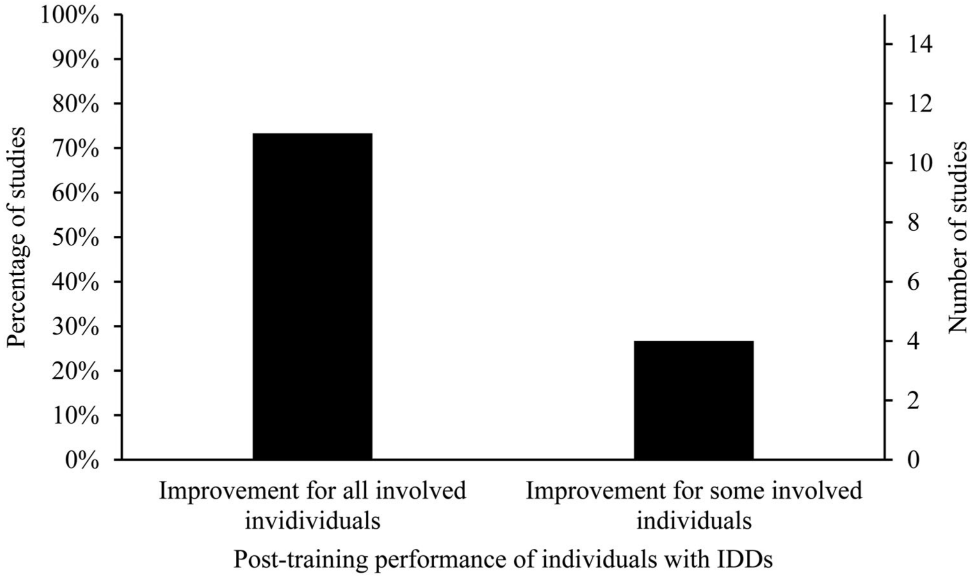

Figure 13. **Training outcomes of individuals with IDDs in studies.**

(Ward-Horner and Sturmey 2008, Hsieh *et al.* 2011), two of three (Miles and Wilder 2009), and three of four (Dogan *et al.* 2017). Two studies did not have the performance of individuals with IDDs reported (Shayne and Miltenberger 2013, Eid *et al.* 2017b).

Generalization

Excluding Lafasakis and Sturmey (2007) and Hassan *et al.* (2018), all other studies included generalization measures. The probes occurred at different time points and implemented different methods. The time points of generalization probes ranged from 1 to 26 weeks. Most studies ($n = 8$) included probes within 5 weeks after conclusion of the training (Seiverling *et al.* 2012, Shayne and Miltenberger 2013, Loughrey *et al.* 2014, Harriage *et al.* 2016, Lanovaz *et al.* 2016, Eid *et al.* 2017a, Eid *et al.* 2017b, Dogan *et al.* 2017). Some studies included the probes at different times for different participants. For example, Miles and Wilder (2009) conducted probes after 3–6 weeks, and Kryzak and Jones (2017) conducted after 2, 6, and 14 weeks. Two studies assessed generalization after 3 months (Stewart *et al.* 2007, Matthews *et al.* 2018) and one study did so after 26 weeks (Subramaniam *et al.* 2017). In addition, Ward-Horner and Sturmey (2008) measured generalization during their training phases when they provided novel exemplars to the caregivers and asked them to teach using the skill they had been taught.

As illustrated in Figure 14, the most common way to assess generalization involved observing the caregivers implementing the trained skills in conditions similar to those observed during baseline/post-training phases (Shayne and Miltenberger 2013, Loughrey *et al.* 2014, Lanovaz *et al.* 2016, Eid *et al.* 2017a, Eid *et al.* 2017b,

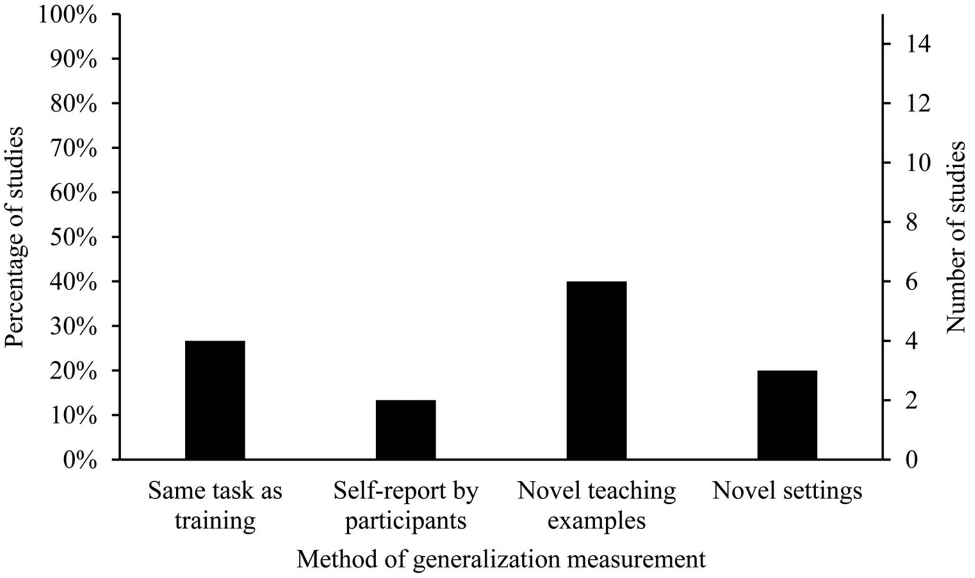

Figure 14. Method of generalization measurement in studies.

Dogan *et al.* 2017). In addition, self-report was used to measure generalization, where researcher consult with caregivers directly (via phone call, Stewart *et al.* 2007; via video conference, Subramaniam *et al.* 2017). Novel teaching examples were also commonly used, wherein caregivers were provided new teaching targets and were required to teach with the trained skills. For example, Ward-Horner and Sturmey (2008) required caregivers to teach new skills. Seiverling *et al.* (2012) used novel food items to access caregivers' ability to reduce food selectivity. Dogan *et al.* (2017) provided new scripts for caregivers to teach social skills. In some cases, researchers assessed generalization by having caregivers perform the taught skills in novel settings. Examples include increasing individuals' guided compliance during different activities (Miles and Wilder 2009), teaching pedestrian safety skills on different street types (Harriage *et al.* 2016), and teaching social interaction skills in a different room (Kryzak and Jones 2017).

All studies but one presented improvement for all caregivers who underwent training (based on a comparison with baseline performance). The single exception was the study by Shayne and Miltenberger (2013), in which improvement was still observed for six of the eight caregivers.

Social validity

Ten (of 17) studies included some form of social validity measures. As reflected in Figure 15, most frequently, investigators used questionnaires consisting of Likert-type scales ($n = 9$, e.g. Seiverling *et al.* 2012). One special case was Kryzak and Jones (2017), who measured social validity by comparing the interaction between individuals with IDDs and their siblings, with the interaction between TD siblings.

With the exception of Stewart *et al.* (2007), where caregivers reported some discomfort to individuals with IDDs, all other studies presented a positive attitude toward the training.

Limitations mentioned in the included studies

There were some limitations mentioned by the original authors in the included articles, which may function as barriers to the improvement of caregivers or the individuals with IDDs. The authors also mentioned practical operations that could serve to avoid the following limitations with careful design.

First, some studies were lacking in skill or preference assessments. In some cases, the authors reported that this was responsible for the lack of improvement in some individuals with IDDs. For example, in Ward-Horner and Sturmey (2008), only one of the three individuals showed better performance after the training. The authors highlighted that such failure was largely due to the excessive difficulty of target skills and this could have been avoided had skill assessment taken place. In Hsieh *et al.* (2011), two of the three individuals with IDDs did not demonstrate improvement likely because the reinforcers were not effective. Again, had preference assessment conducted, this problem might have been avoided.

Second, choice of target behavior was not always optimal and potentially reduced the social validity of the training. In Lanovaz *et al.* (2016), individuals with IDDs displayed a 30% reduction of mild stereotypy engagement, which seems impressive. However, since the topography of the stereotypy engagement was not severe from the beginning, such improvement might not bring much meaningful change for the affected individuals and their families. Therefore, future studies should attempt to include behaviors/skills of high significance.

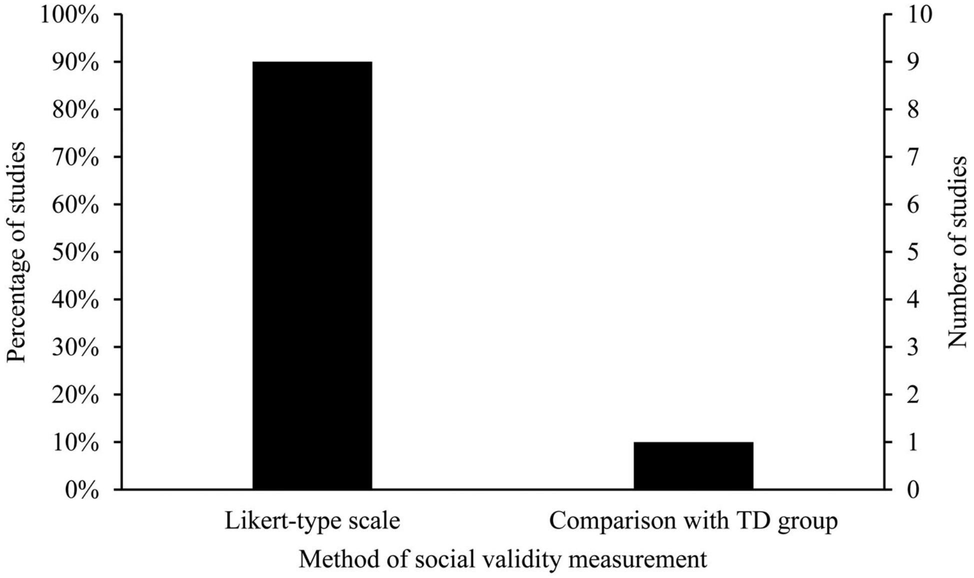

Figure 15. Method of social validity measurement in studies.

Third, it was possible that the effectiveness of the training on trainee competence might have been overestimated due to biasing factors. In one study (Shayne and Miltenberger 2013), participants were shown videos and taught to conduct functional analysis (FA), which included recording the instances of CBs. With each subsequent video, the number of CBs present increased. If this pattern was recognized by participants, they might adjust their level of effort accordingly and thus their scores might not reflect their typical performance. Practical use could avoid such biasing factors by careful design.

Discussion

In this review, the researcher sought to investigate the use of BST to train caregivers of individuals with IDDs. After three phases of search, the researcher ended up with 17 published studies that were included and summarized based on various criteria, which were typically involved in these studies and helpful for practical replication. The criteria contained country of origin, participant demographics, training focus, intervention procedures, outcomes, generalization, and social validity.

There were 14 studies conducted in the US, one in Canada, and two in Saudi Arabia, which suggested an uneven development of BST studies involving training family caregivers. Such distributions can also be noticed in the application of BST studies in training individuals with IDDs themselves, training non-family caregivers, and other ABA approaches. A literature search was conducted in the PsycINFO database, and the results were promising. There were 45 studies related to BST for individuals with IDDs, 42 of those authors were from academic institutions in the US. Similarly, the ratio of BST toward non-family caregivers (e.g. frontline staff, social workers, volunteers) was 27 of 33.

Parents, particularly mothers, were the main participants in the included studies. This suggests that mothers were more likely to be the primary caregivers, and therefore, had the potential to suffer higher parenting stress. Although some studies indicated a higher parenting stress of mothers (e.g. Dabrowska and Pisula 2010), it was not consistent across literatures. Non-examples included Hudock (2012), where fathers and mothers of autistic children presented similar stress levels. Except familial relationship, other demographic information of caregivers was also analyzed in this review and the involved caregivers presented various gender, age, education level, and income level. Moreover, the included studies trained caregivers with different skills or targets. However, most studies presented similar and promising outcomes, which suggested that BST can be applicable to caregivers with different demographics.

An ASD diagnoses represented a majority of the diagnoses reported (37 of 43). However, most behavioral interventions were not exclusively intended specifically or exclusively for those with ASD. For example, Harriage *et al.* (2016) trained three parents on how to increase the pedestrian safety skill for their children, who had different diagnoses (ASD, language impairment, and ADHD). The parents were trained with same procedures and contents, regardless of the diagnoses of their children. Results were promising that all parents and children showed better performance after intervention. Similar examples can also be noticed in Loughrey *et al.* (2014), Lanovaz *et al.* (2016), and Dogan *et al.* (2017). Therefore, rather than only include studies focusing on ASD, the researcher decided to include all diagnoses in this review. Similar with caregiver participants, individuals with IDDs also presented different gender, age, and specific diagnosis. However, most studies showed an improvement in their performance, which suggested that BST to caregivers was helpful to support different individuals with IDDs.

In addition to family caregivers, studies indicated that BST was effective in training individuals whose diagnosis varied (e.g. ID, Travis and Sturmey (2013); mental retardation, Miltenberger *et al.* 1999). Similarly, promising outcomes can also be noticed in BST for non-family caregivers, who supported individuals with different diagnosis (e.g. frontline staff for ID, Gormley *et al.* 2019; volunteers for developmental disabilities, Davis et al. 2019). While a lot of attention has been focused on ASD interventions, including BTS and the ABA field as a whole, these approaches have shown to be effective for other diagnoses (Milyko 2016). For example, Lang *et al.* (2009) conducted a systematic review of using ABA approaches to treat individuals with developmental disabilities, examining 10 empirical studies with 53 participants involved. The studies investigated differential reinforcement, non-contingent reinforcement, shaping, FCT, blocking, and scheduled exercise, of which 80% generated positive outcomes.

The most common study design ($n = 14$) was a multiple baseline design and was typically compared across three or more participants. Consistent with other studies, this design could increase the replicability and predictability of the treatment effect (Harvey *et al.* 2004). The delivery of BST in different studies was generally consistent, however, there was some variation between studies and some that customized novel BST procedures (e.g. booster training sessions, peer training, and observer training). These variations provided practitioners with different ways to implement BST for caregivers.

Outcomes and generalization of this review overall support the notion that BST is generally effective. It demonstrates the potential of family caregivers to administer therapeutic techniques (typically applied by professionals) for individuals with IDDs. Social validity measurements suggest that BST is helpful to increase caregivers' confidence in having control over situations and ability to provide help. This is particularly useful because, as noted in the introduction, there seems to be an inadequate number of specialized professionals targeted towards individuals with IDDs. If family caregivers can effectively deliver similar treatment, it will help to address this deficiency and further reduce parenting stress. Therefore, this review suggests that BST is potentially a critical role in training caregivers for individuals with IDDs.

Implication

This review provides detailed reference information on all available studies involving BST training applied to caregivers. This includes demographics regarding the individuals with IDDs as well as the caregivers themselves. Although the included caregivers differed in relation to age, gender, familial relationship, education, and income level, most studies shared a positive outcome of training efficacy. Similarly, individuals with IDDs varied while most studies presented the improvement of their performance. In addition, trainers can refer to these studies when attempting to make use of BST for caregivers. For example, they can follow training procedures that were empirically reported to have achieved desirable outcomes. They may also find benefit from reviewing the studies wherein the BST was ineffective and modify their own procedures accordingly.

There are also some recommendations this review could provide regarding to practical training family caregivers of individuals with IDDs. Demographics of caregivers do not seem relevant to the training efficacy. Before giving instructions, trainers can ask caregivers to implement the target skills independently and video their performance. During instruction, trainers can discuss the caregivers' performance with them, which would help caregivers avoid making similar mistakes. Instruction only need to be given once, and it is better to provide both oral description and written manuscripts. After instruction, trainers can repeat a modelling-rehearsal-feedback or rehearsal-modelling-feedback sequence, until the caregivers attain some criterion (e.g. an 80% correct implementation for three consecutive sessions). Further, when rehearsing the newly taught skills, trainers could introduce confederates to rehearse with caregivers until they attain certain mastery level, then transition to involve with individuals with IDDs. If the caregivers have difficulty attaining the termination criteria, trainers can increase the number of sessions and implement other means of feedback (e.g. self-examination) to help the caregivers improve their implementation. After the training is terminated, trainers can re-visit their trainees to ensure maintenance and examine if the skills have generalized and provide necessary help.

Some special components could also be referenced by practical training. In addition to researchers providing demonstrations, modelling can also be provided to trainees by video (Shayne and Miltenberger 2013). When teaching skills to siblings of individuals with IDDs, trainers can consider using rewards to increase siblings' motivation of accurate performance (Kryzak and Jones 2017). According to Matthews *et al.* (2018), it would be useful to combine the BST training with the teaching of ABA principles, which can increase caregivers' understanding of the rationale behind each training components. When the resource of professionals is limited, caregivers might be able to increase their skills simply by being trained by peers who have already received BST training and have mastered the target skills (peer training, Loughrey *et al.* 2014), or by observing the training procedure of others (observer group, Hassan *et al.* 2018).

As mentioned previously, there were some limitations mentioned by the original authors of the included studies, which might function as barriers to the improvement of caregivers and individuals with IDDs. Practitioners can avoid these limitations and improve their training by careful design. For example, when selecting training targets, trainers should conduct skill assessment determine the difficulty of the target skills (Ward-Horner and Sturmey 2008), or consider the meaningfulness of the target behaviors (Lanovaz *et al.* 2016). In addition, conducting a preference assessment is necessary for individuals with IDDs to find effective reinforcers for training (Hsieh *et al.* 2011).

In the broader literature, BST also showed the potential to effectively train individuals with IDDs and non-family caregivers. Examples of this was shown with Ryan *et al.* (2019), who elevated conversation interactions for adults with ASD. Moreover, Stannis *et al.* (2019) successfully taught individuals with IDs how to correctly respond to bullying. As discussed in the introduction section, the results of training other caregivers were also consistent with family caregivers. Further examples contain frontline staff (Gormley *et al.* 2019) and volunteers (Davis *et al.* 2019).

Limitations and future studies

While this review was useful for collecting and summarizing information about what is known regarding the use of BST on caregivers of individuals with IDDs, there are certain limitations that should be claimed. These limitations involved the quality appraisal method, generality of results, effect size, EBP, and other useful resources.

First, the inclusion of blind rater and Kappa value tends to bring adverse effect on the quality appraisal. Although the inclusion or removal would not change the final ratings of the included studies, there are still considerations for these two indicators. The blind rater is questioned because it is typically included in between-subject designs (e.g. randomized controlled trial, Husain *et al.* 2014, Andreou *et al.* 2017). In Reichow's Evaluative Method, blind rater was designed for both between- and within-subject research designs, however, the post-training performance (e.g. after attained the termination criteria) of caregivers would look systematic and clearly different from those untrained. Therefore, it would be easy for raters to notice the independent variables and hard to be considered blind. Kappa values are typically relevant to studies where results reported in the form of a category (McHugh 2012). Attaining different number of items results in different final ratings, which are qualitatively described (e.g. mild, moderate, severe). Kappa value could be used in these studies to calculate the probability that raters assign the same rating by chance, however, the dependent measure of the included studies were presented by a numerical value (i.e. percentage of correct performance) and no categories were involved. Although there are limitations of the inclusion of these two indicators, the removal would also be questioned for the consistency with other literatures using Evaluative Method as the quality appraisal tool, and future studies with blind rater or Kappa value included. Thus, this limitation could only be solved by the establishment of another quality appraisal tool.

Second, certain demographics may have been over-represented. For example, the number of females reported ($n = 56$) was greatly higher than males ($n = 10$) and the most common relationship between caregivers and individuals with IDDs were parents ($n = 42$). This suggests that there may have more information on female parent caregivers than male, non-parent caregivers. In terms of individuals with IDDs, individuals were mostly male ($n = 51$ compared to $n = 6$ females) and young (defined as age 12 or under; $n = 39$ compared to $n = 6$ for adults or teenagers aged 13+). The most common diagnosis in the studies was ASD ($n = 37$ compared to $n = 6$ other diagnoses). Thus, the present understanding of BST training with respect to male caregivers, non-parent caregivers, older individuals with IDDs, and individuals with IDDs except ASD is more limited than the understanding of BST as it applies to parents, female caregivers, and young individuals with ASD. Future studies that examine these demographics would therefore be of particularly high value.

Third, although this review made an exhaustive exploration of the topic and assessed the quality of the included studies, it cannot be concluded whether BST was an EBP based on the limited number of studies (see Appendix D of Reichow's 2011). This does not mean BST for caregivers is or is not an EBP, as such conclusion can only be made when there are sufficient number of related empirical studies. For future empirical studies using BST to train caregivers, researchers can improve their study quality by referencing to studies whose quality was rated as adequate or high (Loughrey *et al.* 2014, Harriage *et al.* 2016, Kryzak and Jones 2017, Subramaniam *et al.* 2017, Matthews *et al.* 2018). As mentioned previously, four studies were rated as weak, with all primary indicators rated as high but *participant characteristics* as unacceptable. In addition, 11 of the 17 included studies were rated as unacceptable. There are two common weakness in these studies: Lacking the age of caregivers and information about interventionists. Most studies simply described interventionists as 'experimenter' or 'trainer', which failed to provide useful information and reduced their replicability. Therefore, additional research would be necessary for BST to be considered as an EBP for training caregivers of individuals with IDDs. To increase the methodological quality of future empirical studies, researchers should pay extra attention that the trainees in the included studies were caregivers (rather than individuals with IDDs), and the information should also include age and gender for caregivers.

Finally, some relevant studies might have been left out of this review because of the methodology. Only studies published in peer-reviewed journals were included to ensure the quality of the results. Grey literatures (e.g. book chapters, experiment report, thesis, unpublished manuscripts) were not include, though they might also provide useful information. This limitation is general to systematic reviews, as the nature of this type of review requires researchers to adhere to inclusion criteria strictly (Tomlinson *et al.* 2018).

Conclusion

To conclude, this systematic review included a comprehensive search of using BST to train family caregivers of individuals with IDDs. The search resulted in 17 related articles. The researcher summarized these studies in relation to various characteristics (e.g. training focus, intervention details, outcomes) to come up with a comprehensive understanding of the study focus. Results of this study suggested that BST could be used to train family caregivers across various demographics and generally brought improvement in their ability to support individuals with IDDs. In addition to the improvements made by family caregivers, the individuals with IDDs also showed improvements. Therefore, this study suggested the efficacy of BST for training caregivers and the potential for caregivers to provide effective support for individuals with IDDs. Under the current situation where there are an insufficient number of related professionals as compared with the need for support (as mentioned in the 'Introduction' section), this study provides a referential approach to reduce this gap and to support individuals with IDDs and their families.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders*. 5th ed. Washington, DC: American Psychiatric Association.

Andreou, C., Wittekind, C., Fieker, M., Heitz, U., Veckenstedt, R., Bohn, F. and Moritz, S. 2017. Individualized metacognitive therapy for delusions: A randomized controlled rater-blind study. *Journal of Behavior Therapy and Experimental Psychiatry*, 56, 144–151.

Behaviour Analyst Certification Board. 2019. www.bacb.com/

Beh-Pajooh, A., Abdollahi, A. and Hosseinian, S. 2018. The effectiveness of painting therapy program for the treatment of externalizing behaviors in children with intellectual disability. *Vulnerable Children and Youth Studies*, 13, 221–227.

Bloh, C. and Axelrod, S. 2008. IDEIA and the means to change behavior should be enough: Growing support for using applied behavior analysis in the classroom. *Journal of Early and Intensive Behavior Intervention*, 5, 52–56.

Case-Smith, J. and Bryan, T. 1999. The effects of occupational therapy with sensory integration emphasis on preschool-age children with autism. *The American Journal of Occupational Therapy*, 53, 489–497.

Case-Smith, J. and Miller, H. 1999. Occupational therapy with children with pervasive developmental disorders. *American Journal of Occupational Therapy*, 53, 506–513.

Chang, C. 2014. Cohen's kappa for capturing discrimination. *International Health*, 6, 125–129.

Cockerham, D. and Malaia, E. 2016. Neuroscience-based approaches to teaching students on the autism spectrum. *Zeitschrift für Psychologie*, 224, 290–293.

Dabrowska, A. and Pisula, E. 2010. Parenting stress and coping styles in mothers and fathers of pre-school children with autism and down syndrome. *Journal of Intellectual Disability Research: JIDR*, 54, 266–280.

Dart, E., Radley, K., Furlow, C. and Murphy, A. 2017. Using behavioral skills training to teach high school students to implement discrete trial training. *Behavior Analysis: Research and Practice*, 17, 237–249.

Davis, S., Thomson, K. and Connolly, M. 2019. A component analysis of behavioral skills training with volunteers teaching motor skills to individuals with developmental disabilities. *Behavioral Interventions*, 34, 431–450.

Dogan, R., King, M., Fischetti, A., Lake, C., Mathews, T. and Warzak, W. 2017. Parent-implemented behavioral skills training of social skills. *Journal of Applied Behavior Analysis*, 50, 805–818.

Eid, A. M., Alhaqbani, O. A., Asfahani, S. M., Alaql, M., AlSaud, A. N., Mohtasib, R. S., Aldhalaan, H. M. and Fryling, M. J. 2017a. Learning by doing and learning by observing: Training parents in Saudi Arabia to implement the Natural Language Paradigm. *Journal of Developmental and Physical Disabilities*, 29, 557–565.

Eid, A. M., Aljaser, S. M., AlSaud, A. N., Asfahani, S. M., Alhaqbani, O. A., Mohtasib, R. S., Aldhalaan, H. M. and Fryling, M. 2017b. Training parents in Saudi Arabia to implement discrete trial teaching with their children with autism spectrum disorder. *Behavior Analysis in Practice*, 10, 402–406.

Emery, M. J. 2004. Art therapy as an intervention for autism. *Art Therapy*, 21, 143–147.

Gillam, S. L., Hartzheim, D., Studenka, B., Simonsmeier, V. and Gillam, R. 2015. Narrative intervention for children with autism spectrum disorder (ASD). *Journal of Speech, Language, and Hearing Research: JSLHR*, 58, 920–933.

Gormley, L., Healy, O., O'Sullivan, B., O'Regan, D., Grey, I. and Bracken, M. 2019. The impact of behavioural skills training on the knowledge, skills and well-being of front line staff in the intellectual disability sector: A clustered randomised control trial. *Journal of Intellectual Disability Research: JIDR*, 63, 1291–1304.

Got, I. L. S. and Cheng, S.-T. 2008. The effects of art facilitation on the social functioning of people with developmental disability. *Art Therapy*, 25, 32–37.

Grant, R. J. 2017. *Play-based intervention for autism spectrum disorder and other developmental disabilities*. New York, NY: Routledge/Taylor & Francis Group.

Harriage, B., Blair, K. and Miltenberger, R. 2016. An evaluation of a parent implemented in situ pedestrian safety skills intervention for individuals with autism. *Journal of Autism and Developmental Disorders*, 46, 2017–2027.

Harvey, M. T., May, M. E. and Kennedy, C. H. 2004. Nonconcurrent multiple baseline designs and the evaluation of educational systems. *Journal of Behavioral Education*, 13, 267–276.

Hassan, M., Simpson, A., Danaher, K., Haesen, J., Makela, T. and Thomson, K. 2018. An evaluation of behavioral skills training for teaching caregivers how to support social skill development in their child with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 48, 1957–1970.

Horner, R., Carr, E., Halle, J., Mcgee, G., Odom, S. and Wolery, M. 2005. The Use of Single-Subject Research to Identify Evidence-Based Practice in Special Education. *Exceptional Children*, 71, 165–179.

Hsiao, Y. 2018. Autism spectrum disorders: Family demographics, parental stress, and family quality of life. *Journal of Policy and Practice in Intellectual Disabilities*, 15, 70–79.

Hsieh, H., Wilder, D. and Abellon, O. 2011. The effects of training on caregiver implementation of incidental teaching. *Journal of Applied Behavior Analysis*, 44, 199–203.

Hudock, R. 2012. *Examining parenting stress in parents of children with autism spectrum disorders* (Order No. 3527577). Available from ProQuest Dissertations & Theses Global (1074791266).

Husain, N., Chaudhry, N., Fatima, B., Husain, M., Amin, R., Chaudhry, I. B., Ur Rahman, R., Tomenson, B., Jafri, F., Naeem, F. and Creed, F. 2014. Antidepressant and group psychosocial treatment for depression: A rater blind exploratory RCT from a low income country. *Behavioural and Cognitive Psychotherapy*, 42, 693–705.

Irwin, C. and Axe, J. 2019. Overview of applied behavior analysis and early intervention for autism spectrum disorder. In *Behavioral interventions in schools: Evidence-based positive strategies*. 2nd ed. Washington, DC: APA, pp. 205–226.

Jacobson, J. W., Mulick, J. A. and Green, G. 1998. Cost–benefit estimates for early intensive behavioral intervention for young children with autism—general model and single state case. *Behavioral Interventions*, 13, 201–226.

Jones, S., Bremer, E. and Lloyd, M. 2017. Autism spectrum disorder: Family quality of life while waiting for intervention services. *Quality of Life Research*, 26, 331–342.

Jones, E., Fiani, T., Stewart, J., Sheikh, R., Neil, N. and Fienup, D. 2019. When one sibling has autism: Adjustment and sibling relationship. *Journal of Child and Family Studies*, 28, 1272–1282.

Kirkpatrick, B., Louw, J. and Leader, G. 2019. Efficacy of parent training incorporated in behavioral sleep interventions for children with autism spectrum disorder and/or intellectual disabilities: A systematic review. *Sleep Medicine*, 53, 141–152.

Kossyvaki, L., Jones, G. and Guldberg, K. 2016. Training teaching staff to facilitate spontaneous communication in children with autism: Adult Interactive Style Intervention (AISI). *Journal of Research in Special Educational Needs*, 16, 156–168.

Kryzak, L. and Jones, E. 2017. Sibling self-management: Programming for generalization to improve interactions between typically developing siblings and children with autism spectrum disorders. *Developmental Neurorehabilitation*, 20, 525–537.

Lafasakis, M. and Sturmey, P. 2007. Training parent implementation of discrete-trial teaching: Effects on generalization of parent teaching and child correct responding. *Journal of Applied Behavior Analysis*, 40, 685–689.

Lang, R., Rispoli, M., Machalicek, W., White, P. J., Kang, S., Pierce, N., Mulloy, A., Fragale, T., O'Reilly, M., Sigafoos, J. and Lancioni, G. 2009. Treatment of elopement in individuals with developmental disabilities: A systematic review. *Research in Developmental Disabilities*, 30, 670–681.

Lanovaz, M., Rapp, J. T., Maciw, I., Dorion, C. and Prégent-Pelletier, É. 2016. Preliminary effects of parent-implemented behavioural interventions for stereotypy. *Developmental Neurorehabilitation*, 19, 1–196.

Loughrey, T. O., Contreras, B. P., Majdalany, L. M., Rudy, N., Sinn, S., Teague, P., Marshall, G., McGreevy, P. and Harvey, A. C. 2014. Caregivers as interventionists and trainers: Teaching mands to children with developmental disabilities. *The Analysis of Verbal Behavior*, 30, 128–140.

Martin, N. 2009. Art therapy and autism: Overview and recommendations. *Art Therapy*, 26, 187–190.

Matthews, N. L., Orr, B. C., Harris, B., McIntosh, R., Openden, D. and Smith, C. J. 2018. Parent and child outcomes of JumpStart (TM), an education and training program for parents of children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 56, 21–35.

Maulik, P., Mascarenhas, M., Mathers, C., Dua, T. and Saxena, S. 2011. Prevalence of intellectual disability: A meta-analysis of population-based studies. *Research in Developmental Disabilities*, 32, 419–436.

McHugh, M. 2012. Interrater reliability: The Kappa statistic. *Biochemia Medica*, 22, 276–282.

McPheeters, M. L., Warren, Z., Sathe, N., Bruzek, J. L., Krishnaswami, S., Jerome, R. N. and Veenstra-VanderWeele, J. 2011. A systematic review of medical treatments for children with autism spectrum disorders. *Pediatrics*, 127, e1312.

Miles, N. and Wilder, D. 2009. The effects of behavioral skills training on caregiver implementation of guided compliance. *Journal of Applied Behavior Analysis*, 42, 405–410.

Miltenberger, R. 2004. Behavioral skills training procedures. In R. Miltenberger (Ed.), Behavior modification: Principles and procedures. Belmont, CA: Wadsworth.

Miltenberger, R., Roberts, J., Ellingson, S., Galensky, T., Rapp, J., Long, E. and Lumley, V. 1999. Training and generalization of sexual abuse prevention skills for women with mental retardation. *Journal of Applied Behavior Analysis*, 32, 385–388.

Milyko, K. 2016. The typical and not-so-typical jobs in applied behavior analysis. Purdue University Global.

Morgan, C. and Wine, B. 2018. Evaluation of behavior skills training for teaching work skills to a student with autism spectrum disorder. *Education and Treatment of Children*, 41, 223–232.

Noonan, H., O'Donoghue, I. and Wilson, C. 2018. Engaging with and navigating limbo: Lived experiences of siblings of adults with autism spectrum disorders. *Journal of Applied Research in Intellectual Disabilities: JARID*, 31, 1144–1153.

Polatajko, H. J., Law, M., Miller, J., Schaffer, R. and Macnab, J. 1991. The effect of a sensory integration program on academic achievement, motor performance, and self-esteem in children identified as learning disabled: Results of a clinical trial. *The Occupational Therapy Journal of Research*, 11, 155–176.

Porter, N. and Loveland, K. A. 2019. An integrative review of parenting stress in mothers of children with autism in Japan. *International Journal of Disability, Development & Education*, 66, 249–272.

Postorino, V., Sharp, W., McCracken, C., Bearss, K., Burrell, T., Evans, A. and Evans, A. 2017. A systematic review and meta-analysis of parent training for disruptive behavior in children with autism spectrum disorder. *Clinical Child and Family Psychology Review*, 20, 391–402.

Raulston, T., Hieneman, M., Caraway, N., Pennefather, J. and Bhana, N. 2019. Enablers of behavioral parent training for families of children with autism spectrum disorder. *Journal of Child and Family Studies*, 28, 693–703.

Reichow, B. 2011. Development, procedures, and application of the evaluative method for determining evidence-based practices in autism. In B. Reichow, P. Doehring, D. V. Cicchetti, and F. R. Volkmar, Eds. *Evidence-based practices and treatments for children with autism*. New York, NY: Springer.

Reichow, B., Volkmar, F. and Cicchetti, D. 2008. Development of the evaluative method for evaluating and determining evidence-based practices in autism. *Journal of Autism and Developmental Disorders*, 38, 1311–1319.

Ryan, G., Brady, S., Holloway, J. and Lydon, H. 2019. Increasing appropriate conversation skills using a behavioral skills training package for adults with intellectual disability and autism spectrum disorder. *Journal of Intellectual Disabilities: JOID*, 23, 567–580.

Schaaf, R. and Miller, L. 2005. Occupational therapy using a sensory integrative approach for children with developmental disabilities. *Mental Retardation and Developmental Disabilities Research Reviews*, 11, 143–148.

Seiverling, L., Williams, K., Sturmey, P. and Hart, S. 2012. Effects of behavioral skills training on parental treatment of children's food selectivity. *Journal of Applied Behavior Analysis*, 45, 197–203.

Shayne, R. and Miltenberger, R. 2013. Evaluation of behavioral skills training for teaching functional assessment and treatment selection skills to parents. *Behavioral Interventions*, 28, 4–21.

Stannis, R. L., Crosland, K. A., Miltenberger, R. and Valbuena, D. 2019. Response to bullying (RTB): Behavioral skills and in situ training for individuals diagnosed with intellectual disabilities. *Journal of Applied Behavior Analysis*, 52, 73–83.

Steinbrenner, J. Hume, K. Odom, S. Morin, K. Nowell, S. Tomaszewski, B. Szendrey, S. McIntyre, N. YüCesoy-ÖZkan, S. and Savage, M. 2020. *Evidence-based practices for children, youth, and young adults with Autism.* The University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Institute, National Clearinghouse on Autism Evidence and Practice Review Team.

Stewart, K., Carr, J. and LeBlanc, L. 2007. Evaluation of family-implemented behavioral skills training for teaching social skills to a child with Asperger's disorder. *Clinical Case Studies*, 6, 252–262.

Subramaniam, S., Brunson, L., Cook, J., Larson, N., Poe, S. and St. Peter, C. 2017. Maintenance of parent-implemented discrete-trial instruction during videoconferencing. *Journal of Behavioral Education*, 26, 1–26.

Terband, H., Coppens-Hofman, M. C., Reffeltrath, M. and Maassen, B. A. M. 2018. Effectiveness of speech therapy in adults with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities: JARID*, 31, 236–248.

Tomlinson, S., Gore, N. and McGill, P. 2018. Training individuals to implement applied behavior analytic procedures via eelehealth: A systematic review of the literature. *Journal of Behavioral Education*, 27, 172–222.

Travis, R. and Sturmey, P. 2013. Using behavioural skills training to treat aggression in adults with mild intellectual disability in a forensic setting. *Journal of Applied Research in Intellectual Disabilities*, 26, 481–488.

Ward-Horner, J. and Sturmey, P. 2008. The effects of general-case training and behavioral skills training on the generalization of parents' use of discrete-trial teaching, child correct responses, and child maladaptive behavior. *Behavioral Interventions*, 23, 271–284.

Watling, R. and Dietz, J. 2007. Immediate effect of Ayres's sensory integration–based occupational therapy intervention on children with autism spectrum disorders. *The American Journal of Occupational Therapy*, 61, 574–583.

Wendt, O. and Miller, B. 2012. Quality Appraisal of Single-Subject Experimental Designs: An Overview and Comparison of Different Appraisal Tools. *Education and Treatment of Children*, 35, 235–268.

Wilkes, G. and Lincoln, M. 2018. Parent-mediated intervention training delivered remotely for children with autism spectrum disorder (ASD) has preliminary evidence for parent intervention fidelity and improving parent knowledge and children's social behaviour and communication skills. *Australian Occupational Therapy Journal*, 65, 245–246.

Williamson, G. and Anzalone, M. 1997. Sensory integration: A key component of the evaluation and treatment of young children with severe difficulties in relating and communicating. *Zero to Three*, 17, 29–36.

World Data Lab 2019. Current world population clock.

Zechella, A. and Raval, V. 2016. Parenting children with intellectual and developmental disabilities in Asian Indian families in the United States. *Journal of Child and Family Studies*, 25, 1295–1309.

Appendix A

Overview of included studies

Study	Country	Caregivers	Individuals with IDDs	Training focus	Study design	Setting	BST order	Instruction	Confederates	Special components	Termination criteria	Training outcome	Generalization	Social validity
Stewart *et al.* (2007)	US	1 mother, 1 sister	1 10-year-old male, with ASD and ADHD	Increase social skills	Case study design	Not at home, settings not clearly reported	Instruction-modelling-rehearsal-feedback, and repeat the last three procedures	Instruction on computer	Undergraduate students, during modelling and rehearsal	No special components	80% correct implement for two consecutive sessions	Improvement for both caregivers and individuals with IDDs in both post-training sessions	Measured via phone consultation after three months; Improvement for both caregivers and individuals with IDDs	Measured via phone consultation after three months; Increased confidence of teaching, but children show mild discomfort
Lafasakis and Sturmey (2007)	US	3 mothers	3 4-year-old males, 1 with ASD, 1 with mental retardation, 1 with Down Syndrome	DTT	Multiple baseline design	Not at home, settings not clearly reported	Instruction-modelling-rehearsal-feedback, and repeat the last three procedures	Instruction started with discussing baseline performance	No confederates	No special components	90% correct implementation for two consecutive sessions	Improvement for both caregivers and individuals with IDDs in post-training sessions	Not measured	Not measured
Ward-Horner and Sturmey (2008)	US	2 mothers, 1 father	3 males; age: 4, 4, 5; diagnosis: 2 ASD, 1 ASD and brain injury	DTT	Multiple baseline design	Clinic for special service outside school	Instruction-rehearsal-feedback-modelling, and repeat rehearsal-modelling	Instruction started with discussing baseline performance	No confederates	No special components	90% implementation for three consecutive sessions	Improvement for all caregivers and one of the three individuals with IDDs in post-training sessions; Some IDDs' performance was not improved because the target skill extended the current repertoire	Measured by novel teaching examples, after each training session; Improvement for both caregivers and individuals with IDDs	Not measured
Miles and Wilder (2009)	US	1 kindergarten teacher, 1 nanny, 1 mother	1 female and 2 males; age: 6, 6, 4; 1 ASD, 1 learning disability, 1 TD	Increase guided compliance, by prompt hierarchy	Multiple baseline design	Not at home, settings not clearly reported	Instruction-rehearsal-feedback-modelling, and repeat rehearsal-modelling	Instruction started with discussing baseline performance	No confederates	No special components	100% correct implementation for three consecutive sessions	Improvement for both caregivers and individuals with IDDs in post-training sessions	Measured in novel settings, for five uninterrupted trials, after 3–6 weeks; Generalization shown in all caregivers and two of the three individuals with IDDs	Not measured
Hsieh *et al.* (2011)	US	1 respite care staff, 1 father, 1 mother	2 females and 1 male; age: 8, 8, 10; 2 ASD and severe ID, 1 ASD and Phelan-McDemid Syndrome	Incidental teaching (IT)	Multiple baseline design	Not at home, settings not clearly reported	Instruction-rehearsal-modelling-feedback, and repeat the last three procedures	Instruction given by verbal description and written handouts	No confederates	No special components	Terminate at the first 100% correct implementation	Improvement for all cares and one of the three IDDs in post-training sessions; some IDDs' performance was not improved because of the selection of ineffective reinforcers	Measured by novel teaching examples, after 2–3 weeks; Generalization shown in all caregivers and individuals with IDDs in generalization	Not measured
Seiverling *et al.* (2012)	US	3 mothers	3 males; age: 4, 8, 5; all ASD	Reduce food selectivity	Multiple baseline design	Dinner table at home	Instruction-modelling-rehearsal-feedback, and the last three procedures	Instruction given by written handouts	No confederates	No special components	90% correct implementation, and individuals with IDDs ate the food	Improvement for both caregivers and individuals with IDDs in post-training sessions	Measured by novel feeding items, conducted every week and lasting for 3–4 weeks; Generalization shown in all caregivers and individuals with IDDs in generalization	Social validity measured by 5-item rating scales, which were related with the acceptability and outcome of the training; Caregivers rated agree to strongly agree for social validity items
Shayne and Miltenberger (2013)	US	8 potential adoptive parents, 2 males and 6 females	No individual with IDDs participated in the study	FA by watching videos and conducting ABC record	Multiple baseline design	Researchers' office	Instruction-rehearsal-feedback, and repeat the last two procedures	Instruction given by verbal description and written handouts	No confederates	No real individuals with IDDs presented, caregivers only observed such individuals and their behaviors in videos	Terminate at the first 80% correct implementation	Improvement for all caregivers in post-training sessions	Generalization measured by watching using different videos, after 1–2 weeks; Generalization shown in six of eight caregivers in generalization sessions	Social validity measured by 5-items scales; Caregivers rated agree to strongly agree for social validity items
	US	2 mothers, 1 father		Mand training					No confederates					

(*Continued*)

Study	Country	Caregivers	Individuals with IDDs	Training focus	Study design	Setting	BST order	Instruction	Confederates	Special components	Termination criteria	Training outcome	Generalization	Social validity
Loughrey *et al.* (2014)			1 female and 1 male; age: 3, 12; 1 ASD, 1 FXS		Multiple baseline design	University ASD treatment centre	Instruction-rehearsal-modelling-feedback, and repeat modelling-rehearsal-feedback	Instruction given by PowerPoint presentation and written handouts		No special components	80% correct implementation for three consecutive sessions	Improvement for both caregivers and individuals with individuals with IDDs, in both post-training sessions	Generalization measured by novel teaching examples, after 5 weeks; Generalization shown in all caregivers and individuals with IDDs in generalization	Social validity measured by 5-items scales; Caregivers rated agree to strongly agree for social validity items
Harriage *et al.* (2016)	US	2 mother, 1 father	3 males; age: 14, 15, 23; 1 ASD and language impairment, 1 ASD and ADHD, 1 ASD	Pedestrian safety skill	Multiple baseline design	Home, different street types in local community	Instruction-modelling-rehearsal-feedback, and repeat the last three procedures	Instruction given at home before moving to streets	Confederates: 1) at home: modelled and rehearsed with researcher; 2) at real street: no confederates	No special components	Terminate at the first 80% correct implementation	Improvement for all caregivers and individuals with IDDs in post-training sessions	Generalization measured at novel street types, after 2 weeks; Generalization shown in all caregivers, high in individuals with IDDs at the same street types as trained, but low generalization at new street types (maybe because the street in training was in community, which was quiet and with less cars)	Social validity measured by 5-items rating scales; Caregivers rated agree to strongly agree for social validity items
Lanovaz *et al.* 2016)	US	2 mother, 1 father	3 males, age: 6, 6, 4; 2 ASD, 1 global developmental delay	Reduce stereotypy engagement	Alternating treatment design	Home	Instruction-modelling-rehearsal-feedback, and repeat the last three procedures	Instruction: written and oral	No confederates	No special components	Terminate at the first 80% correct implementation	Improvement for both caregivers and individuals with IDDs, in both post-training sessions	Generalization measured with the same task during training, every one to three weeks; Generalization shown in all caregivers and individuals with IDDs in generalization	Social validity not measured, however the authors doubt the significance of a 30% reduction of stereotypy
Eid *et al.* (2017a)	Saudi Arabia	3 mothers	1 female and 2 males; age: 6, 6, 4; all ASD	Discrete trial training (DTT)	Multiple baseline design	Not at home, settings not clearly reported	Instruction-modelling-rehearsal-feedback, and repeat the last three components	Instruction: written and oral	No confederates	No special components	Terminate at the first 95% correct implementation	Improvement for all caregivers and individuals with IDDs in post-training sessions	Generalization with the same task during training, after 2 weeks; Generalization shown in all caregivers and two of three individuals with IDDs;	Social validity measured by 5-items rating scales; Caregivers rated strongly agree for social validity items
Eid *et al.* (2017b)	Saudi Arabia	6 parents, without gender clearly reported;	6 without gender reported; age: 3, 3, 4, 5, 6, 7; diagnosis not reported	NLP	Multiple baseline design	Not at home, settings not clearly reported	Instruction-modelling-rehearsal-feedback, and repeat the last three procedures	One-to-one instruction	No confederates	Special components: 1) Caregivers were separated into model parents and observer parents, model parents were directly trained by BST, observer parents observe the whole process models being trained and had access to same handouts; 2) feedback was given during rehearsal	Terminate at the first 90% correct implementation	Improvement for both model and observer parents, performance of individuals with IDDs not reported in post-training sessions	Generalization measured with the same task during training, after 3 weeks; Improvement for both model and observer parents, performance of individuals with IDDs not reported in generalization sessions	Social validity measured by 5-item rating scales; Caregivers rated agree to strongly agree for social validity items

(*Continued*)

Study	Country	Caregivers	Individuals with IDDs	Training focus	Study design	Setting	BST order	Instruction	Confederates	Special components	Termination criteria	Training outcome	Generalization	Social validity
Dogan *et al.* (2017)	US	4 mothers	4 males, age: 9, 9, 10, 12; 2 ASD, 1 ASD and ADHD, 1 PDD-NOS	Increase social skills	Multiple baseline design	Not at home, settings not clearly reported	Instruction-modelling-rehearsal-feedback, and repeat the last three procedures	Instruction: written and oral	Confederates: graduate students	Special components: (a) booster sessions, (b) self-monitoring procedure	Terminate at the first 80% correct implementation	Improvement for all caregivers and three of four individuals with IDDs in post-training sessions;	Generalization measured by teaching novel social skills, after 1 month; Generalization shown in all caregivers and individuals with IDDs in generalization sessions	Social validity measured by 5-item rating scales; Caregivers rated agree to strongly agree for social validity items
Kryzak and Jones (2017)	US	3 female siblings, 1 male sibling	1 female and 3 males, age: 6, 8, 10, 12; all ASD	Increase social skills (interaction with siblings)	Multiple baseline design	Individuals' (with IDDs) room at home	Instruction-modelling-rehearsal-feedback, and repeat the last three procedures	Instruction: not clearly reported	Confederates: author, modelling and rehearsal for one round with confederates, then move to real individuals with IDDs	Special components: siblings were given reward for accurate performance	Terminate at first 100% correct implementation	Improvement for all sibling caress and IDDs in post-training sessions;	Generalization measured in novel settings, including siblings' room, and backyard, after 2, 6, and 14 weeks; Improvement for all sibling caregivers and two of four individuals with IDDs in generalization sessions;	Social validity measured by compare the interaction of individuals with IDDs and siblings and of TDs and siblings; Interaction between three of the four sibling-IDD dyads exceeded common TD sibling dyads; one of the four was comparable
Subramaniam *et al.* (2017)	US	4 mothers	4 males; age: 2, 7, 8, 11; all ASD	DTT	Multiple baseline design	Clinic	Instruction-modelling-rehearsal-feedback, and repeat the last three procedures	Instruction: written and oral	Confederates: 1) trainer; 2) model and rehearsal for three rounds, then move to real individuals with IDDs	No special components	Terminate at 90% correct implementation for three consecutive sessions	Improvement for all caregivers and individuals with IDDs, in post-training sessions	Generalization measured via videoconferencing, where participants give descriptive feedback, after 26 weeks; Generalization shown in all caregivers and individuals with IDDs in generalization sessions	Social validity not measured
Hassan *et al.* (2018)	Canada	7 mothers, 1 grandmother	8 without age reported; age: all 6 – 8; all ASD	Increase social skills	Multiple baseline design;	Quite rooms, not at home, but also not clearly reported	Instruction-modelling-rehearsal-feedback, and repeat the last three components	Instruction: written and oral	Confederates: 1) trainer; 2) modelled with confederates, rehearsed with real individuals with IDDs	Special components: Observation group, where parents outside the training program observed and provide feedback	Terminate at 100% correct implementation for three consecutive sessions	Improvement for all caregivers and individuals with IDDs in post-training sessions	Generalization not measured	Social validity measured by 5-item rating scales; Caregivers rated agree to strongly agree for social validity items
Matthews *et al.* (2018)	US	15 mothers, 3 fathers	18 males; age: mean of 40 months; diagnosis not reported	Understand and implement ABA procedures. Use service system to establish comprehensive intervention program	Experimental design (without random assignment of participants)	Community-based settings	Instruction-modelling-rehearsal-feedback	Instruction: written and oral, about basic ABA knowledge	No confederates	Special components: 1) BST embedded within an ABA training program; 2) models in video examples; 3) rehearsed at home; 4) no feedback for rehearsal	Terminate when all pre-designed content was taught	Improvement for all caregivers and individuals with IDDs in post-training session	Generalization measured by video recording the interaction between parents and children, after 3 months; Generalization shown in all caregivers and individuals with IDDs in generalization sessions	Social validity not measured

ASD autism spectrum disorder; ADHD attention deficit hyperactivity disorder; DTT discrete trial training; IDDs intellectual and developmental disabilities; TD typical developing; IT incidental teaching; FA functional assessment; ABC antecedent-behavior-consequence; FXS Fragile X syndrome; NLP natural language paradigm; ABA applied behavior analysis.

Appendix B

Evaluative method indicator definitions and ratings

See Tables A1, A2, and A3.

Table A1. Primary indicators of research report rigor.

Study Design	Primary Quality Indicator	Rating Criteria High (H)	 Acceptable (A)	 Unacceptable (U)
Group and single case design	Participant Characteristics (PART)	1. Age and gender are provided for all caregivers (mean age, common sexual name acceptable). 2. Diagnosis of individuals with IDDs (if included) are operationalized by including specific diagnosis and diagnostic instrument (ADOS, ADI-R, CARS, DSM-IV, and ICD-10 acceptable) used to make the diagnosis, or an operational definition of behaviors and symptoms of the participants. 3. Information on the characteristics of interventionist are provided (the ability to determine who did the intervention is minimal a criterion) and information on any secondary participants (e.g. peers) is provided. 4. If the study provides standardized test scores, the measures used to obtain those scores are indicated.	Meet criteria 1, 3, and 4.	Not meet all of criteria 1, 3, and 4.
Group and single case research design	Independent variable (IV)	Define independent variable with replicable precision (i.e. one could reproduce the intervention given the description). If a manual is used, the study passes.	Define many elements of the independent variable but omits specific details.	Did not sufficiently define the independent variable.
Group and single case research design	Dependent variable (DV)	1. The variables are defined with operational precision. 2. The details necessary to replicate the measures are provided. 3. The measures are linked to the dependent variable. 4. The measurement data is collected at appropriate times during the study for the analysis being conducted.	Meet three of the four criteria.	Meet fewer than three criteria.
Single case research design	Baseline condition (BL)	1. Encompass at least three measurement points. 2. Appear through visual analysis to be stable. 3. Have no trend or a counter-therapeutic trend. 4. Have conditions that are operationally defined with replicable precision.	Meet three of the four criteria.	Meet less than three criteria.
Single case research design	Visual analysis (VA)	1. Have data that are stable. 2. Contain less than 25% overlap of data points between adjacent conditions, unless behavior is at ceiling or floor levels in the previous condition. 3. Show a large shift in level or trend between adjacent conditions that coincide with the implementation or	Two criteria were met at least 66% of the graphs.	Two or fewer criteria were met on less than 66% of the graphs.

(*Continued*)

Study Design	Primary Quality Indicator	Rating Criteria High (H)	 Acceptable (A)	 Unacceptable (U)
		removal of the IV. If there was a delay in change at the manipulation of the IV, the study is acceptable as high quality if the delay was similar across different conditions or participants (± 50% of delay).		
Single case research design	Experimental control (EC)	Contains at least three demonstrations of the experimental effect, occurring at three different points in time and changes in the DVs vary with the manipulation of the IV in all instances of replication. If there was a delay in change at the manipulation of the IV, the study is accepted as high quality if the delay was similar across different conditions or participants (± 50% of delay).	At least 50% of the demonstrations of the experimental effect meet the criteria: there are two demonstration of the experimental effect at two different points in time and changes in the DVs vary with the manipulation of the IV.	Less than 50% of the demonstrations of the experimental effect meet the above criteria: there are fewer than two demonstrations of the experimental effect occurring at two different points in which changes in the DVs vary with the manipulation of the IV.
Group research design	Comparison condition (CC)	Define the conditions for the comparison group with replicable precision, including a description of any other interventions participants receive.	Vaguely describes the conditions for the comparison group, information on other interventions may not be reported.	Not report the conditions for the comparison group or has no control or comparison group.
Group research design	Link between research question and data analysis (LRQ)	Data analysis is strongly linked to the research questions and uses correct units of measure.	Data analysis is poorly linked to the research questions but uses correct units for a majority of the outcome measures.	Data analysis is linked weakly or not at all to the research questions and uses the correct unit for only a minority of the outcome measures.
Group research design	Statistical analysis (SA)	Proper statistical analyses were conducted with an adequate power and sample size ($n > 10$) for each statistical measure.	Proper statistical analyses were conducted for at least 75% of the outcome measures, or in which proper statistical analyses were conducted on 100% of outcome measures but with inadequate power or a smaller sample size.	Statistical analysis was not done correctly, the sample size too small or the power was inadequate.

IDDs intellectual and developmental and developmental disabilities; ADOS autism diagnostic observation schedule; ADI-R autism diagnostic interview of outcome measures but with inadequate power or a smaller sample size.ry with the manipulation of torders – 4th edition; ICD-10 international classification of diseases bilith revision.

Table A2. Secondary indicators of research report rigor.

Study Design	Secondary Quality Indicator	Rating Criteria (Is evidence of the indicator)
Group and single case research design	Blind raters (BR)	Raters are blind to the treatment condition of the participants.
Group and single case research design	Fidelity (FD)	Treatment or procedural fidelity is continuously assessed across participants, conditions, and implementers, and if applicable, has measurement statistics $>.80$.
Group and single case research design	Generalization/ maintenance (GM)	Outcome measures are collected after the final data collection to assess generalization or maintenance.
Group and single case research design	Social validity (SV)	Contains at least four of the following features: 1. Socially important DVs (i.e. society would value the changes in outcome of the study). 2. Time- and cost-effective intervention (i.e. the ends justify the means). 3. Comparisons between individuals with and without disabilities. 4. A behavioral change that is large enough for practical value (i.e. it is clinically significant). 5. Consumers who are satisfied with the results. 6. IV manipulation by people who typically come into contact with the participant. 7. A natural context.
Single case research design	Interobserver agreement (IOA)	IOA is collected across all conditions, raters, and participants with reliability $>.80$.
Group research design	Random assignment (RA)	Participants are assigned to groups using a random assignment procedure.
Group research design	Interobserver agreement (IOA)	IOA is collected across all conditions, raters, and participants with reliability $>.80$.
Group research design	Attrition (ATR)	Articulation is comparable (does not differ between groups by more than 25%) across conditions and less than 30% at the final outcome measure.
Group research design	Effect size (ES)	Effect sizes are reported for at least 75% of the outcome measures and are $>.40$.

Table A3. Research report strengthen ratings.

Rating	Study design	Rating criteria
Strong	Group research design	Received high quality grades on all primary quality indicators and showed evidence of four or more secondary quality indicators.
	Single case research design	Received high quality grades on all primary quality indicators and showed evidence of three or more secondary quality indicators.
Adequate	Group research design	Received high quality grades on four or more primary quality indicators with no unacceptable quality grades on any primary quality indicators and showed evidence of at least two secondary quality indicators.
	Single case research design	Received high quality grades on four or more primary quality indicators with no unacceptable quality grades on any primary quality indicators and showed evidence of at least two secondary quality indicators.
Weak	Group research design	Received fewer than four high quality grades on primary quality indicators or showed evidence of less than two secondary quality indicators.
	Single case research design	Received fewer than four high quality grades on primary quality indicators or showed evidence of less than two secondary quality indicators.

Appendix C

EBP ratings by different raters and final decision

See Tables A4, A5, and A6.

Table A4. EBP ratings for each included study by the author.

Study	Study design	Primary indicators									Secondary indicators									Strength
		PA	IV	DV	BL	VA	EC	CC	LQR	SA	BR	FD	GM	SV	IOA	KA	RA	ATR	ES	
Stewart et al. (2007)	S	H	H	H	U	U	U	–	–	–	×	×	√	√	√	×	–	–	–	W
Lafasakis and Sturmey (2007)	S	U	H	H	H	A	H	–	–	–	×	×	√	×	√	×	–	–	–	W
Ward-Horner and Sturmey (2008)	S	U	H	U	H	H	H	–	–		×	×	√	×	√	×	–	–	–	W
Miles and Wilder (2009)	S	U	H	H	H	H	H	–	–	–	×	√	√	×	√	×	–	–	–	W
Hsieh et al. (2011)	S	U	H	H	H	H	H	–	–	–	×	√	√	×	√	×	–	–	–	W
Seiverling et al. (2012)	S	U	H	H	A	U	H	–	–	–	×	×	√	√	√	×	–	–	–	W
Shayne and Miltenberger (2013)	S	U	H	H	A	U	H	–	–	–	×	√	√	√	√	×	–	–	–	W
Loughrey et al. (2014)	S	A	H	H	H	H	H	–	–	–	×	√	√	√	√	×	–	–	–	A
Harriage et al. (2016)	S	H	H	H	H	H	H	–	–	–	×	√	√	√	√	×	–	–	–	S
Lanovaz et al. 2016)	S	U	H	H	H	H	H	–	–	–	×	√	√	×	√	×	–	–	–	W
Eid et al. (2017a)	S	U	H	H	U	H	H	–	–	–	×	×	√	√	√	×	–	–	–	W
Eid et al. (2017b)	S	U	H	H	U	H	H	–	–	–	×	×	√	√	√	×	–	–	–	W
Dogan et al. (2017)	S	H	H	H	U	U	H	–	–	–	×	√	√	√	√	×	–	–	–	W
Kryzak and Jones (2017)	S	H	H	H	H	A	H	–	–	–	×	√	√	√	√	×	–	–	–	A
Subramaniam et al. (2017)	S	A	H	H	A	H	H	–	–	–	×	√	√	×	√	×	–	–	–	A
Hassan et al. (2018)	S	U	H	H	U	U	H	–	–	–	×	√	×	√	√	×	–	–	–	W
Matthews et al. (2018)	G	H	H	H	–	–	–	A	H	H	√	×	√	×	×	–	×	√	√	A

Table A5. EBP ratings for each included study by the secondary rater.

Study	Study Design	Primary Indicators									Secondary Indicators									Strength
		PA	IV	DV	BL	VA	EC	CC	LQR	SA	BR	FD	GM	SV	IOA	KA	RA	ATR	ES	
Stewart et al. (2007)	S	U	H	H	U	U	U	–	–	–	×	×	√	√	√	×	–	–	–	W
Ward-Horner and Sturmey (2008)	S	U	H	U	H	H	H	–	–	–	×	×	√	×	√	×	–	–	–	W
Hsieh et al. (2011)	S	A	H	H	H	H	H	–	–	–	×	√	√	×	√	×	–	–	–	A
Seiverling et al. (2012)	S	U	H	H	A	U	A	–	–	–	×	×	√	×	√	×	–	–	–	W
Harriage et al. (2016)	S	U	H	H	H	H	H	–	–	–	×	√	√	√	√	×	–	–	–	W
Dogan et al. (2017)	S	U	H	H	A	H	H	–	–	–	×	√	√	√	√	×	–	–	–	W
Kryzak and Jones (2017)	S	H	H	H	H	A	H	–	–	–	×	√	√	√	√	×	–	–	–	A
Subramaniam et al. (2017)	S	U	H	H	U	H	H	–	–	–	×	√	√	√	√	×	–	–	–	W
Hassan et al. (2018)	S	H	H	H	U	U	H	–	–	–	×	√	√	√	√	×	–	–	–	W

Table A6. Final EBP Ratings for Each Included Study.

Study	Study Design	Primary Indicators									Secondary Indicators									Strength
		PA	IV	DV	BL	VA	EC	CC	LQR	SA	BR	FD	GM	SV	IOA	KA	RA	ATR	ES	
Stewart et al. (2007)	S	U	H	H	U	U	U	–	–	–	×	×	√	√	√	×	–	–	–	W
Lafasakis and Sturmey (2007)	S	U	H	H	H	A	H	–	–	–	×	×	√	×	√	×	–	–	–	W
Ward-Horner and Sturmey (2008)	S	U	H	U	H	H	H	–	–		×	×	√	×	√	×	–	–	–	W
Miles and Wilder (2009)	S	U	H	H	H	H	H	–	–	–	×	√	√	×	√	×	–	–	–	W
Hsieh et al. (2011)	S	U	H	H	H	H	H	–	–	–	×	√	√	×	√	×	–	–	–	W
Seiverling et al. (2012)	S	U	H	H	A	U	H	–	–	–	×	×	√	√	√	×	–	–	–	W
Shayne and Miltenberger (2013)	S	U	H	H	A	U	H	–	–	–	×	√	√	√	√	×	–	–	–	W
Loughrey et al. (2014)	S	A	H	H	H	H	H	–	–	–	×	√	√	√	√	×	–	–	–	A
Harriage et al. (2016)	S	H	H	H	H	H	H	–	–	–	×	√	√	√	√	×	–	–	–	S
Lanovaz et al. 2016)	S	U	H	H	H	H	H	–	–	–	×	√	√	×	√	×	–	–	–	W
Eid et al. (2017a)	S	U	H	H	U	H	H	–	–	–	×	×	√	√	√	×	–	–	–	W
Eid et al. (2017b)	S	U	H	H	U	H	H	–	–	–	×	×	√	√	√	×	–	–	–	W
Dogan et al. (2017)	S	H	H	H	U	U	H	–	–	–	×	√	√	√	√	×	–	–	–	W
Kryzak and Jones (2017)	S	H	H	H	H	A	H	–	–	–	×	√	√	√	√	×	–	–	–	A
Subramaniam et al. (2017)	S	A	H	H	A	H	H	–	–	–	×	√	√	×	√	×	–	–	–	A
Hassan et al. (2018)	S	U	H	H	U	U	H	–	–	–	×	√	√	√	√	×	–	–	–	W
Matthews et al. (2018)	G	H	H	H	–	–	–	A	H	H	√	×	√	×	×	–	×	√	√	A

Appendix D

Criteria for Evidence-Based practice (EBP), Reichow (2011)

Level of EBP	Example criteria
Established ($\geq$ 60 points from the EBP status formula)	• Five SSED studies of strong research report strength with a total sample size of at least 15 participants across studies conducted by at least three research teams in three different geographic locations • Ten SSED studies of adequate research report strength with a total sample size of at least 30 different participants across studies conducted by at least three research teams in three different geographic locations • Two group design studies of strong research report strength conducted by in different geographic locations • Four group design studies of at least adequate research report strength conducted in at least two different research teams • One group design study of strong research report strength and three SSED studies of strong research report strength with at least 8 different participants • Two group design studies of at least adequate research report strength and six SSED studies of at least adequate research report strength with at least 16 different participants
Promising ($>$ 30 points from the EBP status formula)	• Five SSED studies of at least adequate research report strength with a total sample size of at least 16 different participants across studies conducted by at least two research teams in two different geographic locations • Two group design studies of at least adequate research report strength • One group research report of at least adequate research report strength rating and at least three SSED studies of at least adequate strength rating with at least 8 participants

SSED single-subject experimental design.