



A Systematic Review and Quality Appraisal of Applications of Direct Instruction with Children with Autism Spectrum Disorder

Sarah E. Frampton¹ · Greg T. Munk¹ · Laura A. Shillingsburg¹ · M. Alice Shillingsburg¹

Accepted: 29 April 2021 / Published online: 1 June 2021
© Association for Behavior Analysis International 2021

Abstract

Developed by Siegfried (“Zig”) Engelmann and colleagues, direct instruction (DI) has been recognized as an effective and replicable teaching model for decades. Although rooted in many principles of learning that behavior analysts utilize in daily practice, DI is not a common component of behavior analytic services for learners with autism spectrum disorder (ASD). This may be attributed to behavior analysts’ unfamiliarity with research evaluating the efficacy of DI with learners with ASD. This article synthesizes findings across studies evaluating DI with learners with ASD. The review addresses the contributions of the studies to date and identifies additional areas of research that may lead to more learners with ASD benefitting from DI.

Keywords Autism · Direct instruction · Education · Practice · Review

Autism spectrum disorder (ASD) is characterized by restricted and repetitive patterns of behavior as well as deficits in social interaction and social communication, with related impairments in language and problematic behaviors (American Psychiatric Association, 2013). Children with ASD often require intensive and highly specialized interventions to meet their social, language, and behavioral needs (Hyman et al., 2020). Over the past few decades, research investigating instructional approaches for children with ASD has increased substantially (Wong et al., 2015). Though research has accelerated and options for empirically supported intervention approaches have expanded, the translation to practice has been delayed by numerous barriers (Cook & Odom, 2013; Critchfield et al., 2017; Love et al., 2009; Odom et al., 2010). Some barriers reported to impede translation from research to practice include vaguely

✉ Sarah E. Frampton
sframpton@mayinstitute.org

¹ May Institute, Inc., Simmons University, 41 Pacella Park Drive, Randolph, MA 02368, USA

described procedures accompanied with technical jargon and insufficient details on treatment dosage in research articles (e.g., duration of sessions, frequency of sessions). Furthermore, time and resources needed to train implementors on new procedures has been reported as sparse. Given these barriers, one solution may lie in the adoption of procedures successfully implemented on a wide scale in similar contexts, provided ample evidence exists to support application with children with ASD. Direct instruction (DI) meets both criteria (Engelmann et al., 1988).

DI is one of the most effective instructional approaches in the education literature and has been implemented with thousands of students (Hattie, 2009; Stockard et al., 2018). Developed by Bereiter and Engelmann (1966), DI was designed to promote mastery of educational content through high rates of active group responding led by a trained teacher. DI consists of both clearly defined procedures to teach skills and carefully designed content to be taught (Barbash, 2012; Engelmann et al., 1988; Heward et al., 2006). DI includes scripts that indicate exactly how the instructor should present the lesson. These scripts were developed with an aim of promoting consistently high-quality teaching across implementors and minimize unnecessary talking, which may detract from the students' learning. Across scripts the core teaching procedures include the similar components of (1) orientation, (2) presentation, (3) practice, (4) feedback, and (5) independent practice. The teacher orients the students to the exercise by describing the content about to be taught and then presents the content by modeling desired responses. The teacher allows the students practice opportunities, either individually or in unison. The teacher immediately provides feedback in the form of praise for correct responses and error correction following incorrect responses. The scripted lessons are arranged to grow in scope and sequence as the students progress through daily exercises.

DI curricula have been developed to address educational areas such as reading, math, writing, and language (National Institute for Direct Instruction [NIFDI], 2021). Within each curriculum, a placement test guides the teacher in selecting which lessons to begin with for each student or group of students. Students should be grouped according to performance, allowing instruction to progress at a reasonable pace through daily intensive practice. The posttest for each group of lessons informs the teacher as to whether additional practice is required or if changes in group assignment are needed for any students. Use of posttest data and ongoing analysis of error patterns empowers teachers to make informed, responsive decisions to serve their students as individuals (Barbash, 2012; Watkins & Slocum, 2004). In addition, the relatively infrequent data collection allows the teacher to focus primarily on delivering the intervention with fidelity.

DI is grounded in the principles of learning and the influence of these principles can be detected in the procedures and materials (Engelmann et al., 1988). First, DI establishes responding with systematic prompts, prompt fading, and checks for discrimination with similar targets. As new content is introduced, the teacher's script calls for modeling of the desired response. Following the model, the learners are instructed to respond along with the teacher. After a practice opportunity, the learners are asked to respond in unison. An independent opportunity is often repeated later in the exercise, in discrimination with other responses. This progression of systematic prompt fading that incorporates discrimination is built into the instructions, though "on the fly" adjustments to meet students' needs remains an essential component of DI implementation

(Heward et al., 2006). Second, DI utilizes a general case model of instruction as new targets are introduced (O’Neill, 1990). For example, as a student learns to tact a cat, multiple pictures of cats are incorporated throughout the material. Examples span both 2D and 3D exemplars in some cases, such as tacting a table in a picture and then in the classroom setting. The presentation of multiple exemplars (e.g., new cat colors, new body positions, new table material) within different arrays may prevent the development of faulty stimulus control over responding. The embedded multiple exemplar approach may also promote stimulus generalization as new exemplars are presented in subsequent lessons (LaFrance & Tarbox, 2020). Third, the curriculum embeds checks for maintenance of learned concepts as lessons progress. Thus, once the tact for cat is learned, it will be incorporated in future lessons. Fourth, logical content analysis informed the design and sequence of DI lessons with an aim of “teaching more in less time” (Engelmann et al., 1988). In one set of exercises a learner may learn to use the autoclitic frame “This is” before tacting a picture. In subsequent lessons the frame will be required with tacts that were previously established in the absence of the full frame (e.g., “This is a cat”). By harnessing recombinative generalization (Goldstein, 1983), learners exponentially acquire new skills as the instructor follows the sequence of the curriculum. Finally, many DI exercises require repetition until fluency, ensuring learners continue to practice until they achieve the highest levels of competence (Binder, 1996).

DI recently met criteria for consideration as an evidence-based practice for learners with ASD (Steinbrenner et al., 2020). This milestone is important to encourage wider use of DI in the many contexts in which children with ASD receive behavior analytic services. However, heeding Cook and Odom’s (2013) findings, we know evidence of efficacy alone is not sufficient to produce adoption among the complexities of everyday life in clinical settings. It is unclear how familiar practicing behavior analysts are with DI and to what extent DI is utilized in clinical practice. DI is not listed on the Behavior Analysis Certification Board’s 5th edition Task List, which drives much of the curriculum in master’s level programs. In addition, cursory review suggests that articles on DI do not frequently appear on the pages of the mainstream behavior analytic journals. These facts suggest that more in-depth continuing education may be needed for behavior analysts to come into contact with this literature, develop competency with DI procedures, and ultimately adopt DI into their practice. We hope this review of the DI literature with learners with ASD will begin bridging this research to practice gap. Our aims with this review are to (1) characterize the scope and quality of research addressing DI with learners with ASD to highlight the contributions to date, (2) identify areas in need of additional research that may move DI closer to adoption in practice, and (3) provide preliminary recommendations for incorporating DI into clinical practice with children with ASD.

Method

Search Procedures and Reliability

To identify relevant articles for inclusion we conducted searches within three databases: PsychInfo, ERIC, and the NIFDI in May 2020. The PsychInfo and ERIC searches were

conducted using the keywords “direct instruction” and “autism.” The NIFDI search was conducted using only the keyword “autism” as all the studies in the database related to direct instruction. The first author conducted the database searches and returned 128 total studies. The second author repeated the search and returned 125 total studies; 3 fewer articles were returned on the ERIC data base at the time of the search. From the 125 total studies, 38 were duplicates across databases and were removed. The first and second author then screened the remaining articles for inclusion in the review. Articles were included if they were peer reviewed, experimental, in English, included at least one participant with a diagnosis of ASD, and evaluated an established DI curriculum. We considered established DI curricula as those listed on the NIFDI ([, 2021](#)) website. From the screening process, 77 articles failed to meet inclusion criteria. A total of 16 articles met criteria for inclusion in the review. The first and second author achieved 100% interrater reliability on inclusion and exclusion.

Coding Procedures

The first author developed coding items related to the primary domains of article information, participant characteristics, DI curricula and implementation, and experimental components. Scores for each domain were input into a survey created by the first author using Microsoft Forms ©. Codes for all domains are listed below.

Article Information

The journal in which the article was published, the first author, and the year of publication was collected.

Participant Characteristics

Diagnosis

The number of participants with a reported diagnosis of ASD was collected. The inclusion of diagnostic testing to confirm the diagnosis of ASD was also coded, whether it was reported as part of the participant characterization or specifically conducted as part of the study.

Demographics

Of the participants with a reported diagnosis of ASD, the reported age, reported gender, and reported race/ethnicity was also coded.

Characterization Methods

Of the participants with a reported diagnosis of ASD, the inclusion of additional characterization methods including IQ tests and standardized language measures was also coded. The specific measures used was also noted.

DI Curricula and Implementation

Curricula Evaluated

The specific DI curricula evaluated in the study were collected. The evaluation of the entire curricula or only partial strands was also coded.

Implementation

The context of implementation was coded as in a group (i.e., greater than 2), dyad, or one-on-one. The reported dosage was also coded, with notes related to specific dosage amounts per study.

Implementor

The status of the implementor was coded as researcher, teacher, paraprofessional, behavior technician, and other. The code researcher was applied if the implementor was an author of the article, regardless of their other credentials. In other words, if the implementor was a teacher but also an author of the paper, they were coded as a researcher. These codes were designed to differentiate between *expert* implementors, as would be expected of a study author, in contrast with *lay person* implementors.

Modifications

Inclusion of any modifications to standard DI procedures was coded and the category of modification was noted. The categories of possible modifications were noted as relates to materials, response criteria, criteria for mastery, probe/testing procedures, intervention procedures, and other.

Experimental Components

Experimental Design

The type of design included was scored broadly as group, single subject design, case study, or mixed methods. If a single subject design was included, the specific design applied was also coded.

Measures

The inclusion of an evaluation of maintenance and generalization was coded, as reported by the authors. The inclusion of reliability data, procedural fidelity, and social validity data were also coded.

Quality Indicators

We utilized the standards developed by the Council for Exceptional Children ([CEC]; Cook et al., 2014) to evaluate indicators of quality research studies. The CEC identified

eight indicators, each with multiple features applicable to group and/or single-subject design. The first and third author jointly reviewed the CEC indicators and discussed their applicability to the focus of this review (e.g., DI and ASD). The first author created a scoring sheet in which each feature of each indicator was scored as “met” or “not met.” If indicated as “not met,” the rationale was noted. We adhered to the CEC guidelines and considered the entire indicator “not met” if a single component of a feature was judged insufficient.

Effect Size Calculation

To further assess the efficacy of DI, we evaluated the magnitude of change of dependent measures by calculating or extracting a reported effect size. Effect sizes are a quantifiable measure of the extent to which the independent variable was associated with change to the dependent variable (Parker & Vannest, 2009). In group design, calculations of effect size are a common way of describing differences between experimental and control groups. In single-subject designs, effect sizes can be calculated by comparing data points in the baseline condition to data points following intervention (Parker & Vannest, 2009). For studies using a single-subject design we conducted a nonoverlap of all pairs (NAP; Parker & Vannest, 2009), the analysis using an online NAP calculator (Vannest et al., 2016). For group design studies that utilized inferential statistics we extracted the reported effect size. No additional calculations were deemed necessary because results were published in peer reviewed journals.

NAP scores are calculated by comparing each data point in the baseline condition to each data point in subsequent conditions to assess the proportion of data points with no overlaps out of the total comparisons. The higher the score, the less overlap between conditions, suggestive of a stronger experimental effect (Parker & Vannest, 2009). The scores were calculated for each participant with ASD by comparing the baseline phase and intervention or treatment phase only; maintenance data were not included in the analysis. If the effects of DI were evaluated across multiple behaviors for one participant, a NAP score was calculated for each evaluated behavior (i.e., tier on the graph) and then averaged.

Reliability

A random number generator was used to randomly select 50% of the studies to score for reliability on coding study characteristics and quality appraisal. For coding study characteristics, the first and second author independently coded the items according to the coding sheet following calibration training. We calculated point-by-point agreement for each coded item. The number of coded items with agreement for each study were divided by the total number of items and multiplied by 100. The mean initial agreement score across studies was 90% (range: 84%–100%). Disagreements were resolved through discussion until consensus was reached.

For quality indicators, the first and third author independently coded the same randomly selected articles following calibration training. We calculated point-by-point agreement for each feature of each of the eight indicators. The number of features scored with agreement for each study were divided by the total number of features applicable to the study (24 for group designs; 22 for single-

subject designs) and multiplied by 100. The mean score across studies was 95% (range: 83%–100%). Disagreements were resolved through discussion until consensus was reached.

For effect size calculations, the first author trained the second author on use of the NAP calculator (Vannest et al., 2016). We used a random number generator to select 50% of the single subject design studies to score for reliability. We calculated proportional IOA for each participant's NAP score by dividing the smaller recorded by the larger recorded and converting the sum into a percentage. Mean IOA across studies was 94% (range: 74%–100%). Agreement below 80% on a study triggered a rescore, which occurred for one study. Following the rescore, agreement reached 98%.

Results

Article Information

The search returned 16 studies from 10 journals published between the years of 1993 and 2020. Of the published studies, only one study was published within a behavior analysis journal (Frampton et al., 2020). Fourteen studies were published in journals focused on ASD (e.g., *Journal of Autism and Developmental Disorders*) or special education (e.g., *Exceptionality*). One article was published in the *Journal of Direct Instruction*.

Participant Characteristics

Diagnosis

Across the 16 studies, 92 participants had a reported diagnosis of ASD. In 63% of the studies, diagnoses of ASD were supported by details from diagnostic testing or use of a specific screening tool like the Childhood Autism Rating Scale (CARS; Schopler et al., 1988). For the remaining studies, diagnoses of ASD were either reported by caregivers or identified by eligibility for special education services.

Demographics

Participants ranged in age from 4 to 17 years, though most participants fell into the 4–6 and 7–12 age ranges. Of note, participants in the 4–6 age range primarily participated in evaluations of *Language for Learning* (Engelmann & Osborn, 2008a). This curriculum requires the fewest prerequisite skills and is designed for pre-kindergarten to second grade. However, *Language for Learning* was also evaluated with participants ages 10–15 (e.g., Frampton et al., 2020; Shillingsburg et al., 2015). Across studies, the majority of participants were male ($N = 69$) with only a few participants identified as female ($N = 7$). (gender was not reported in three studies). Non-white/Caucasian participants made up 44% of participants (e.g., Black, African American, Asian, Hispanic), though 50% of studies did not report the race or ethnicity of participants.

Characterization Methods

Seven studies reported the IQ of participants. The Test of Nonverbal Intelligence-3 (Brown et al., 1997) and Leiter International Performance Scale Revised (Roid & Miller, 2002) were most frequently cited. Most participants' IQs fell into the ranges of 55–70 ($N = 11$), 71–85 ($N = 10$), and 86–100 ($N = 8$). Three participants' IQ fell below 55 and three participants' IQs were over 100. Language or reading skills were assessed with a standardized measure within 11 studies; a wide variety of assessments were utilized across studies including Woodcock-Johnson-III Tests of Achievement (WJ-III; Woodcock et al., 2001), Battelle Developmental Inventory 2 (Newborg, 2005), Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007), and Test of Language Development (Hammill & Newcomer, 2008).

DI Curricula and Implementation

Curricula Evaluated

Language for Learning (Engelmann & Osborn, 2008a), *Corrective Reading* (Engelmann et al., 2002), and *Reading Mastery* (Engelmann & Osborn, 2008b) curricula were most frequently evaluated across studies (see Table 1). Two studies (Thompson et al., 2012, 2019) evaluated *Connecting Math Concepts* (Engelmann et al., 2012). In five studies only particular skill strands within the curriculum were implemented. Ganz and Flores (2009) evaluated skills related to teaching material identification within the *Language for Learning* curriculum (Engelmann & Osborn, 2008a). Cadette et al. (2016) focused on teaching skills related to answering “wh” questions within the *Reading Mastery* curriculum. Flores and Ganz (2007, 2009) evaluated reasoning skills within the *Corrective Reading* curriculum (Engelmann et al., 2002). Of note, Wolfe et al. (2018) implemented the full *Language for Learning* curriculum but results for only particular skill strands were graphed.

Implementation

DI was designed to be incorporated as part of a daily dose of instruction in a classroom context. All but one of the studies we reviewed reported some element of dosage (e.g., either frequency or duration), but reports were highly variable. Duration ranged from 10 min per session (Flores et al., 2016) to 90 min per session (Shillingsburg et al., 2015). In some cases, the number of exercises or lessons addressed per session were reported, rather than the duration (e.g., Thompson et al., 2012; Wolfe et al., 2018). Sessions were conducted multiple times per week in all instances in which dosage was reported and lasted from 4 weeks (e.g., Flores & Ganz, 2014) to multiple years (e.g., Kamps et al., 2016).

A one-to-one instructional format was used in seven studies (e.g., Flores et al., 2016; Frampton et al., 2020; Head et al., 2018; Infantino & Hempenstall, 2006; Shillingsburg et al., 2015; Thompson et al., 2012; Wolfe et al., 2018) and dyad format in three studies

(e.g., Flores et al., 2013, 2016; Thompson et al., 2019). Specific rationale for the selection of one-to-one, dyad, or group instruction was not consistently provided. Several authors noted that one-to-one instruction is not unusual for learners with ASD in school contexts (Shillingsburg et al., 2015; Wolfe et al., 2018). In addition, one-to-one instruction may be the only method possible for children receiving services in their homes (e.g., Infantino & Hempenstall, 2006; Wolfe et al., 2018).

Implementor

In nine of the reviewed studies a researcher (i.e., author of the study) was responsible for implementing the intervention (see Table 1). Nonauthor teachers were responsible for implementation in four studies (Flores et al., 2013; Flores & Ganz, 2014; Kamps et al., 2016; Weisberg & Savard, 1993) and paraprofessionals in two studies (Kamps et al., 2016; Flores & Ganz, 2014). In two studies (Shillingsburg et al., 2015; Wolfe et al., 2018), behavior therapists were responsible for implementation and in one study (Infantino & Hempenstall, 2006) the lead implementor was the parent.

Modifications

Other modifications to procedures were reported in 13 studies. We noted modifications across the dimensions of materials, response criteria, mastery criteria, probe/testing procedures, and intervention procedures. Modifications to probe procedures and mastery criteria were noted in studies that applied a single-subject design. DI curricula require only a placement test before instruction can begin. In contrast, single subject designs rely on steady state responding in baseline and repeated measures once an independent variable has been applied. Changes such as increased frequency of probe sessions (e.g., Cadette et al., 2016; Ganz & Flores, 2009) or use of posttests as pretests (e.g., Frampton et al., 2020; Shillingsburg et al., 2015) were frequently reported in order to demonstrate a functional relation. Several studies also utilized different material forms in conjunction with expanded probe procedures (e.g., Ganz & Flores, 2009; Thompson et al., 2012; Wolfe et al., 2018).

Some modifications were made to accommodate features of the altered instructional context or learner specific characteristics. For example, to accommodate use of one-to-one instruction, Shillingsburg et al. (2015) and Frampton et al. (2020) used pictures of peers (rather than actual peers) during instructional sessions. To accommodate participants utilizing speech generating devices, Frampton et al. (2020) changed response topography (e.g., from vocal speech to selecting or spelling on device) and intervention procedures (e.g., from vocally modeling to modeling responses on the device while modeling responses vocally).

In several studies the authors reported used of child-specific schedules of reinforcement (e.g., Frampton et al., 2020; Shillingsburg et al., 2015; Wolfe et al., 2018). Several studies specified that a correct response had to occur within 5 s of an instruction (e.g., Shillingsburg et al., 2015; Wolfe et al., 2018). These procedural elements do not necessarily represent a deviation from the standard DI procedures, perhaps best considered clarifications for instructors implementing DI procedures.

Table 1 Summary of Articles Included in Review

Authors and Year	Participants with ASD	Curricula	Design	Curriculum Span	Implementor	Quality Indicators	NAP Score or Analyses	Statistical	Effect Size
Connecting Math Concepts									
Thompson et al. (2012)	3 (6–8)	Connecting Math Concepts	Multiple probe design (MPD) across participants	Skill strands	Author	6/8 indicators; 20/22 features	P1: .60 P2: .75 P3: .71		Moderate Moderate Moderate
Thompson et al. (2019)	3 (7–9)	Connecting Math Concepts and Language for Learning	Multiple baseline design (MBD) across participants	Entirety	Author	7/8 indicators; 20/22 features	P1: .98 P2: .99 P3: .98		Strong Strong Strong
Corrective Reading: Thinking Basics, Comprehension, & Decoding									
Flores and Ganz (2007)	2 (11–14)	Corrective Reading: Thinking Basics	MPD across behaviors	Skill strands	Author	5/8 indicators; 19/22 features	P1: $M = .99$ (range: .991) P2: $M = 1$		Strong Strong Strong
Flores and Ganz (2009)	2 (12–14)	Corrective Reading: Thinking Basics	MPD across behaviors	Skill strands	Author	6/8 indicators; 19/22 features	P1–P2: $M = 1$		Strong
Flores et al. (2013)	11 (7–13)	Corrective Reading: Thinking Basics & Language for Learning	Group (DJ-CR & DJ-LL)	Entirety	Teacher	3/8 indicators; 19/24 features	$F(1, 10) = 149.25$, $p < .01$, partial $\eta^2 = .94$ $F(1, 6) = 569.38$, $p < .01$, partial $\eta^2 = .99$		Strong Strong
Head et al. (2018)	2 (10–16)	Corrective Reading: Comprehension	MPD across behaviors	Entirety	Author	6/8 indicators; 20/22 features	P1: $M = 1$ P2: $M = .98$ (range: .79–1)		Strong Strong
*Infantino and Hempenstall (2006)	1 (7)	Corrective Reading: Decoding	Case Study	Entirety	Parent	1/8 indicators; 11/22 features	Could not be calculated from reported data		NA
Language for Learning									
*Flores and Ganz (2014)	6 (4–10)	Language for Learning	Group (DI-LL compared to DJT)	Entirety	Teacher, para-professional	6/8 indicators; 22/24 features	$t(11) = -2.63$, $p = .02$ $d = .62$		Moderate
Flores et al. (2016)	1 (4)	Language for Learning	MBD across behaviors	Entirety	Author	7/8 indicators; 21/22 features	P1: $M = 1$		Strong

Table 1 (continued)

Authors and Year	Participants with ASD	Curricula	Design	Curriculum Span	Implementor	Quality Indicators	NAP Score or Statistical Analyses	Effect Size
Frampton et al. (2020)	3 (10–15)	Language for Learning	MPD across participants	Entire	Author	4/8 indicators; 18/22 features	P1–P3: 1	Strong
Ganz and Flores (2009)	2 (10)	Language for Learning	Changing criterion design	Skill strands	Author	5/8 indicators; 19/22 features	P1: .99 P2: .98	Strong Strong
Shillingsburg et al. (2015)	18 (4–12)	Language for Learning	Mixed methods	Entirety	Behavior technician	6/8 indicators; 25/28 features	P1–P18; $M = 1$	Strong
Wolfe et al. (2018)	2 (4–7)	Language for Learning	MPD across behaviors	Entirety	Behavior technician (Research assistant)	7/8 indicators; 21/22 features	P1: $M = .89$ (range: .87–.89) P2: $M = .95$ (range: .89–1)	Strong Strong
Reading Mastery								
Cadette et al. (2016)	3 (15–17)	Reading Mastery	MPD across behaviors	Skill strands	Author	6/8 indicators; 19/22 features	P1: $M = .93$ (range: .83–1) P2: $M = .93$ (range: .90–.96) P3: $M = .93$ (range: .81–1)	Strong Strong Strong
*Kamps et al. (2016)	32 (5–6)	Language for Thinking and Reading Mastery	Group	Entirety	Teacher, para-professional	5/8 indicators; 21/24 features	1) Reading Mastery Words (Intervention); $F = 9.11$, $p = .0027$, $d = 0.80$ 2) DIBELS Nonsense Word Fluency (Intervention); $F = 6.38$, $p = .0141$, $d = 0.64$ 3) Oral Reading Fluency (Time by Intervention); $F = 2.05$, not significant 4) Woodcock Reading Mastery Test (WRMT) Word Identification (Intervention); $F = 5.63$, $p = .0208$, $d = 0.60$	1) Strong 2) Medium-strong 3) NA 4) Medium 5) NA

Table 1 (continued)

Authors and Year	Participants with ASD	Curricula	Design	Curriculum Span	Implementor	Quality Indicators	NAP Score or Statistical Analyses	Effect Size
Weisberg and Savard (1993)	1 (5)	Reading Mastery	Mixed methods	Entirety	Teacher	5/8 indicators 19/22 features Across Studies:	5) WRMT Word Attack: $F = 2.44$, not significant 6) WRMT Passage Comprehension (Time by Intervention): $F = 2.27$, not significant P1: 1 (Words measure) $M = .97$ (range: .60–1)	6) NA Strong

Note: NAP (nonoverlap of all pairs) scores were calculated for intervention or treatment data only, maintenance data were not included. If participants were exposed to multiple DI interventions, scores were averaged.

* = Indicates no reported modifications to DI standard procedures

Experimental Components

Experimental Design

A variety of experimental methodologies has been used to evaluate DI with learners with ASD using (see Table 1). Infantino and Hempenstall (2006) utilized a case study design, including very robust characterization methods. A multiple probe design across behaviors or participants was featured in 9 of the 10 studies that used single subject designs. Shillingsburg et al. (2015) and Weisberg and Savard (1993) used a mix of single subject and group design features by assigning participants to cohorts. All participants in both studies received the intervention, but the timing was systematically delayed to evaluate the effects of DI (Shillingsburg et al., 2015) or a variation in DI (Weisberg & Savard, 1993).

Three studies utilized a form of group design, though the nature of the designs varied. Flores et al. (2013) evaluated the effects of *Language for Learning* (Engelmann & Osborn, 2008a) or *Corrective Reading* (Engelmann et al., 2002) as applied in a school setting. The interventions were not compared to one another, nor were they compared to a control group. However, results of the study showed that both curricula led to gains in student performance when intervention was led by school personnel. Flores and Ganz (2014) compared the performance of children randomly assigned to either *Language for Learning* or discrete trial teaching during an extended school year program. Participants in the DI group outperformed participants in the discrete trial teaching (DTT) group, though only 13 total participants were included. Kamps et al. (2016) utilized more rigorous procedures to randomize assignment of 62 participants to a *Reading Mastery* (Engelmann & Osborn, 2008b) group or a treatment as usual (TAU) group in a school setting. The study spanned up to 2 years and was implemented by school personnel.

Measures

Across studies, the inclusion of procedural fidelity, reliability, and social validity measures varied. Some form of procedural fidelity measure was included in all studies, though we observed variation in the frequency of evaluation. Many studies conducted procedural fidelity checks at least weekly for each group or participant. Infantino and Hempenstall (2006) included only four evaluations of fidelity for the parent implementing the procedures in her home. However, given this was a case study, the fidelity of the implementation speaks most to the feasibility of the intervention for a nonexpert. Measures of reliability of the dependent measure were used in 12 studies, less often in cases utilizing group or mixed method approaches. Maintenance data, ranging from a single probe to multiple probes over multiple weeks, was also reported in 12 studies. Generalization was evaluated in only four of studies. For example, Thompson et al. (2012) include probes for generalization across materials and Wolfe et al. (2018) evaluated generalization across implementors. Some form of social validity assessment was included in only three of studies. In all instances, parents (e.g., Infantino & Hempenstall, 2006; Thompson et al., 2012) and teachers (e.g., Thompson et al., 2012, 2019) reported a degree of high satisfaction with the DI procedures.

Quality Indicators

No study in this review met 100% of the Cook et al., (2014) quality indicators, though three met seven of eight indicators (Flores et al., 2016; Thompson et al., 2019; Wolfe et al., 2018). The study by Infantino and Hempenstall (2006) only addressed one out of eight of the quality indicators, though this is not surprising given that it was a case study. The quality indicators most commonly satisfactorily met were description of practice (14 studies) and data analysis (14 studies). The quality indicators most commonly judged insufficient were descriptions of context and setting (eight studies), participants (five studies), and implementation fidelity (seven studies). With regard to context and settings, the feature most commonly scored as “not met” was the physical layout of the settings in which DI took place. With regard to participants, the features most commonly scored “not met” were participant race or ethnicity and method in which ASD diagnosis was obtained. With regard to implementation fidelity, the feature most commonly scored as “not met” was reporting on the consistency of procedural fidelity throughout the intervention phase.

Effect Size

NAP scores (i.e., effect sizes) for each study are reported in Table 1. The mean effect size across the 42 participants in the 10 studies using single-subject designs was .97 ($SD = .08$, 95% CI [.6–1]), corresponding to a strong treatment effect per the interpretative guidelines provided by Parker and Vannest (2009). These overall findings indicate there was little overlap between data points in the baseline and intervention or treatment phases; suggestive of a high magnitude change in behavior following DI interventions. It should be noted that NAP calculations consider only the extent of difference between phases (i.e., treatment data points higher than baseline data points): they cannot determine the extent to which the change was socially significant.

Reported effect sizes were extracted from group design studies. The particular research questions varied across the group design studies, leading to application of different forms of inferential statistical analyses and different interpretations of reported data. Flores and Ganz (2014) compared scores on the mastery test of *DI-Language for Learning* before and after exposure to either DI to DTT. Results indicated the DI group scored higher than the DTT group, reaching statistical significance with a moderate effect size ($d = .62$). Flores et al. (2013) evaluated lesson test scores for both *DI-Corrective Reading* and *DI-Language for Learning* after weeks of intervention on each respective curriculum. The changes in test scores reached significance, with strong effect sizes for both (partial $h^2 = .94$; partial $h^2 = .99$). Kamps et al. (2016) compared the effects of *Reading Mastery* to TAU reading instruction on multiple measures. The authors reported significant effects of the *Reading Mastery* intervention on three of the measures, with moderate to strong effect sizes. The authors also found that the improvements between the DI group and TAU group were significant on the same three measures, with weak to moderate effect sizes obtained (see Kamps et al., 2016). Taken together, results of group design studies report moderate to strong effect sizes for DI on evaluated dependent measures.

Discussion

Summary of Results

We identified 16 studies meeting our inclusion criteria addressing a total of seven DI curricula implemented in a variety of settings by implementers of varying expertise. Our review indicated that no study included all of the quality indicators identified by the CEC (2014), though we identified three studies meeting seven of eight indicators (Flores et al., 2016; Thompson et al., 2019; Wolfe et al., 2018). Across studies, the “not met” features were almost entirely related to omitted or unclear study elements (e.g., participant race/ethnicity, method of determining ASD diagnosis, method of training implementers, distribution of procedural fidelity data collection across phases, exact dosage of the intervention). These features were not included in the Steinbrenner et al. (2020) review, which accounts for the obtained differences. It is also important to note that 50% of the studies evaluated in our review were published on or before 2014, when the CEC guidelines were made available. With consideration of the overall scope of the DI literature, quality indicators, and demonstrations of effect, our overall assessment echoes that of Steinbrenner et al. that there is evidence for the efficacy of DI with learners with ASD. However, there remains much to be done to increase the certainty of this evidence and progression into adoption in clinical settings.

Many of the studies we reviewed consisted of single subject designs focused on evaluating the efficacy of various DI curricula (see Table 1). However, as indicated by Smith et al. (2007), evaluations of efficacy are only the first step in the life cycle of research, progressing to manualization, randomized controlled trials (RCT), and ultimately to community effectiveness. It is fortunate that DI programs are already manualized and our review identified two RCTs in which DI was compared to a TAU control group (Kamps et al., 2016; Shillingsburg et al., 2015). This progression towards community effectiveness is promising, though additional research is needed along the steps Smith et al. (2007) highlighted to bring DI to wider adoption in clinical settings. Thus, we propose that future research on DI with children with ASD progress iteratively within and across curricula, moving from strands to entire curricula to randomized controlled trials. By advancing along these steps, gaps in the research can be identified and addressed with the appropriate research tools, leading towards adoption in clinical settings. To this end, we will discuss future directions in research aimed at gaps identified by our review while also highlighting how practitioners may utilize what is currently known about DI in their practice with their clients with ASD.

Recommendations for Research

Our review identified multiple studies focusing on particular skill strands using single subject designs (e.g., Cadette et al., 2016; Flores & Ganz, 2007, 2009; Ganz & Flores, 2009; Thompson et al., 2012). These demonstrations are critical contributions as there is relatively little behavior analytic literature to guide instruction on complex skills such as time-telling (covered in *Connecting Math Concepts*) and analogical reasoning (covered in *Corrective Reading: Thinking Basics*). Furthermore, a focused demonstration with a particular population or novel curriculum can serve as a foundation for a more comprehensive evaluation (Smith et al., 2007).

For curricula for which efficacy has been established with skill strands, research should pivot into evaluating the entire curricula with an aim towards achieving a dose resembling typical educational practices (e.g., Kamps et al., 2016). For example, Thompson et al. (2012) used single subject designs to evaluate the skill strand of telling time within *Connecting Math Concepts*. Then, in Thompson et al. (2019) the full *Connecting Math Concepts* curricula was evaluated. For curricula with multiple single subject evaluations on a curriculum in its entirety, the next most appropriate step may be an evaluation at a larger dose utilizing a randomized control (Smith et al., 2007). For example, the *Corrective Reading* curricula have multiple demonstrations with single subject designs examining both skill strands and the full curricular span, but no evaluations have been conducted using group design with randomized control (see Table 1).

Our quality appraisal indicated that diagnostic status was inconsistently reported across studies and information from formal diagnostic evaluations not consistently available. This outcome is perhaps to be expected as much of the literature to date was conducted in the context of special education services. But it leaves open the possibility that students were included under an educational category of ASD that did not actually have a psychological diagnosis of ASD. This inconsistency coupled with inconsistent reporting of cognitive or language assessments, suggests we still have much to learn about the participants for whom this approach is efficacious. In addition to diagnostic details, reporting on specific prerequisite skills may also be critical to advance this line of research. For example, Thompson et al. (2019) found that the participants in their study struggled to engage in unison responding during *Connecting Math Concepts* group instruction. We recommend as a preliminary step that future studies include reports from diagnostic and standardized testing, in addition to characteristics such as frequency and severity of challenging problem behavior and responsiveness to group instruction. If negative outcomes correlate with specific participant variables (e.g., stereotypy), this will guide the development of future research questions and matched remedial supports.

Our quality appraisal indicated that the participants' race and ethnicity were inconsistently reported. Of the participants for whom these demographics were reported, our results indicate that most participants were Caucasian/white, males, under the age of 13, and had a vocal-verbal language repertoire. The greater the representation of participants from varying genders, races, ethnicities, abilities, social economic status, the stronger our assessment of external validity (Pierce et al., 2014). Our review indicates that DI has been evaluated with only seven participants with ASD between the ages of 14 and 17 and zero participants older than 17. We also found that only three participants with ASD that utilized AAC have been represented in DI research (Frampton et al., 2020). To address generality and ensure representation across groups it may be insufficient to simply cast a wide net and hope for diversity. To extend this literature into wider-scale adoption we may need to specifically seek out underrepresented members of the ASD population for participation in research (for an example, see Cariveau et al., 2019). If differences are detected, these results may drive further investigations and matched supports. Comprehensive reporting of participant demographics will also increase the overall quality appraisal of this line of research (Cook et al., 2014).

Our review found that the majority of reviewed studies included some modification to standard DI procedures. Perhaps this is not surprising, given the emphasis on individualization in special education (e.g., Odom et al., 2010) and applied behavior analysis (e.g., Baer, 2005; Baer et al., 1968). These modifications can be viewed as both a strength and a limitation of the DI literature. As a strength, these modifications suggest that small changes to the curricula can be applied and efficacious outcomes still obtained; expanding appeal to practitioners requiring flexibility to serve learners of diverse needs. However, it is a slippery slope as too many changes may result in substantial drift from core procedures. As studies with modifications are replicated, the modifications may be unknowingly replicated as well. Drift from the guiding principles may lead to a packaged intervention no longer recognizable as DI. As a protection against this possibility, we suggest in future studies that authors note all modifications to standard procedures and provide rationale to clarify the decision for the reader (e.g., to support the design, to support the learner's needs). With more consistent reporting, descriptions, and evaluations of outcomes, future studies may then evaluate the necessity of particular modifications to determine under which conditions they are best deployed. These modifications may eventually serve as a foundation of decision-making model that may support implementers as they deliver DI to children with ASD without expert support.

Our review indicated that measures such as maintenance, reliability, and procedural fidelity were commonly included within reviewed studies. However, we noted inconsistent reporting on the distribution of procedural fidelity data collection within and across phases, which limited the assessment of study quality according to the CEC guidelines (Cook et al., 2014). Future studies should continue to incorporate these critical study elements with an eye towards the goal of community effectiveness (Smith et al., 2007). In practice contexts, the durability, acceptability, and feasibility of approaches are critical to promote actual use (Cook & Odom, 2013). With the aim of adoption in clinical settings, it may be important for future DI studies to include different forms of these standard measures that are more aligned with conditions of practice when serving children with ASD. In practice, oversight by experts has been reported as intermittent (Love et al., 2009). Future studies should evaluate the maintenance of procedural fidelity by nonexpert implementers when expert support is faded. In practice, the frequency of data collection ranges from each trial to only a subset of daily trials (Love et al., 2009). Future studies should also evaluate if intermittent evaluations of performance, such as lesson posttests, are sufficient for making decisions to progress or repeat exercises with learners with more idiosyncratic needs. In practice, assessments of generalization and maintenance are common features of programming (Love et al., 2009). Generalization to relevant contexts should continue to be investigated with DI curricula. For example, does fluency using a full sentence when labeling during *Language for Learning* (Engelmann & Osborn, 2008a) exercises generalize when tacting objects on a walk with a caregiver? Social validity measures should be incorporated to evaluate whether parents and teachers report favorably about DI and whether they choose to continue to use it when expert support is withdrawn. Addressing these critical areas may serve to bring research towards the conditions of practice that it is

hoped will accelerate adoption into the settings in which ABA services are provided.

Recommendations for Practice

Our review suggests DI may be a useful contribution to the practice of applied behavior analytic services for children with ASD. The consistently strong demonstrations of effect suggest that this approach is a powerful behavior change intervention that may be feasible in practice settings (Hyman et al., 2020; Odom et al., 2010). For clinicians wishing to utilize DI as a component of their practice, first and foremost, we suggest they enhance their own competence in the area through supervision or professional development (Brodhead et al., 2018). In addition to reading research articles (see References section) and attending continuing education events, clinicians may access a wealth of resources through NIFDI. The NIFDI (2021) website offers a variety of online resources that may be useful for this purpose including webinars, guides, and example materials. In addition, NIFDI offers training, coaching, and support by experts in DI.

Next, clinicians should consider focusing on the specific curricula that have been empirically evaluated with children with ASD across multiple studies (see Table 1). Our review suggests the curricula with the most empirical support (by considering number of studies and participant numbers) are *Language for Learning*, *Corrective Reading*, and *Reading Mastery*. For curricula evaluated with few children with ASD (e.g., *Connecting Math Concepts* and *Language for Thinking*) or no children with ASD (e.g., *Expressive Writing* [Engelmann & Silbert, 1983]), we suggest clinicians proceed with caution. Because DI curricula rely on shared core procedures, the risks associated with utilizing an underevaluated curricula may be minimal, as long as clinicians exercise appropriate judgement with respect to prerequisite skills and closely monitor performance. The selection of a curriculum should be matched to the individual's unique educational or treatment goals. Scope and sequence guides are embedded in many of the DI materials. These guides, along with review of the specific required responses as shown in the presentation books, can be helpful in determining what responses will be established through the curriculum. These responses can then be matched to target goals to assist a clinician in determining what proportion of an individual's educational or clinical services may be addressed through DI.

From our own clinical experience and the examples pulled from this literature, we suggest clinicians consider a progressive evaluation of DI with each individual served. Following the general model established in the research literature, this may begin with an evaluation of carefully selected and matched skill strands (e.g., Cadette et al., 2016; Flores & Ganz, 2007, 2009; Ganz & Flores, 2009; Thompson et al., 2012) with ongoing data collection to detect barriers or error patterns. Frampton et al. (2020) included measures such as participant affect, repetitions per exercise, and frequency of session terminations that may be useful to this end. Close evaluation of these data will allow clinicians to make modifications, if necessary, and evaluate their effects on an individual basis. As a successful model of implementation is established, the scope of the DI exercises could expand and progress toward use of the entire curricula. In some contexts, one-on-one instruction may be the only model available (e.g., in home

therapy). When possible we suggest clinicians actively progress towards paired or small group instruction to gain the maximum benefit of the DI approach. The procedures used by Thompson et al. (2019) serve as a strong example of a means by which this goal can be achieved.

Some clinicians may find the cost of DI materials to be a barrier to adoption, especially if they are serving children with a wide range of needs requiring multiple curricula or if children served are in multiple locations (e.g., home and school settings). In these instances, it may be advisable to start with curricula that will be most immediately utilized for the greatest number of children served and conducting a cost–benefit analysis over months of use. In this analysis, elements such as time saved making materials, time saved in ongoing training of implementers, and time saved in program development by clinicians should be weighed. These benefits may offset the upfront costs as the materials are incorporated into practice on an enduring basis. In instances where children served are in different locations, DI may be accommodated to remote options through video meeting technology and use of digital materials (NIFDI, 2021). NIFDI provided additional support on transitioning to remote DI to meet the needs of students and children during the COVID-19 pandemic, though more research is necessary to determine the feasibility and efficacy of these modifications on a long-term basis.

Limitations and Conclusions

We acknowledge that this review has several limitations that must be considered. Though our reliability scores indicate we were internally consistent in our assessment of study quality, some interpretation of study features is explicitly required according to the CEC (Cook et al., 2014) guidelines. We echo Mitchell et al.'s (2017) sentiment that the dichotomous reporting of “met” or “not met” with respect to CEC (Cook et al., 2014) standards may yield an overly conservative assessment of study quality. We noted instances in which many components of a feature were addressed, but absent one component we adhered to recommendations and precedent from prior research to score as “not met” (Mitchell et al., 2017; Weston et al., 2018). Taken together, it is possible future teams of reviewers may interpret reported features in a different manner and reach different conclusions regarding study quality. Our findings may also be limited by inclusion of only “established” DI programs (i.e., listed on the NIFDI [2021] website). Future reviews may consider including a broader scope to incorporate studies that utilized instructional design steps and teaching procedures based on DI with new curricula or materials (e.g., Ragnarsdóttir, 2007).

The research path of DI with learners with ASD illustrates the unique contributions of many types of research. When working in concert with one another, the limitations of one study may inspire the methods of the next study. For behavior analysts serving children with ASD, familiarity with DI is simply a first step. We hope this review helps to demystify the DI approach and shine a spotlight on the work that has been done evaluating this practice with learners with ASD. As more behavior analysts utilize DI within this population, together we can address the challenges of adoption and implementation.

References

(Studies included in the review are marked with an asterisk [*].)

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.
- Baer, D. M. (2005). Letters to a lawyer. In W. L. Heward, T. E. Heron, N. A. Neef, S. M. Peterson, D. M. Sainato, G. Cartledge, R. Gardner III, L. D. Peterson, S. B. Hersh, & J. C. Dardig (Eds.), *Focus on behavior analysis in education: Achievements, challenges, and opportunities* (pp. 3–30) Merrill/Pearson.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*(1), 91–97. <https://doi.org/10.1901/jaba.1968.1-91>.
- Barbash, S. (2012). *Clear teaching: With direct instruction, Siegfried Engelmann discovered a better way of teaching*. Education Consumers Foundation
- Bereiter, C., & Engelmann, S. (1966). Observations on the use of direct instruction with young disadvantaged children. *Journal of School Psychology, 4*(3), 55–62. [https://doi.org/10.1016/0022-4405\(66\)90008-2](https://doi.org/10.1016/0022-4405(66)90008-2).
- Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*(2), 163–197.
- Brodhead, M. T., Quigley, S. P., & Wilczynski, S. M. (2018). A Call for Discussion About Scope of Competence in Behavior Analysis. *Behavior Analysis in Practice, 11*, 424–435. <https://doi.org/10.1007/s40617-018-00303-8>.
- Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). *Test of nonverbal intelligence* (3rd ed). PRO-ED.
- *Cadette, J. N., Wilson, C. L., Brady, M. P., Dukes, C., & Bennett, K. D. (2016). The effectiveness of direct instruction in teaching students with autism spectrum disorder to answer “Wh-” questions. *Journal of Autism & Developmental Disorders, 46*(9), 2968–2978. <https://doi.org/10.1007/s10803-016-2825-2>
- Cariveau, T., Shillingsburg, M. A., Alamoudi, A., Thompson, T., Bartlett, B., Gillespie, S., & Seahill, L. (2019). A structured intervention to increase response allocation to instructional settings for children with autism spectrum disorder. *Journal of Behavioral Education, 29*, 699–716. <https://doi.org/10.1007/s10864-019-09340-x>.
- Cook, B., Buysse, V., Klingner, J., Landrum, T., McWilliam, R., Tankersley, M., & Test, D. (2014). Council for Exceptional Children: Standards for evidence-based practices in special education. *Teaching Exceptional Children, 46*(6), 206.
- Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children, 79*(3), 135–144. <https://doi.org/10.1177/001440291307900201>.
- Critchfield, T. S., Doepke, K. J., Epting, L. K., Becirevic, A., Reed, D. D., Fienup, D. M., Kreamsreiter, J. L., & Ecott, C. L. (2017). Normative emotional responses to behavior analysis jargon or how not to use words to win friends and influence people. *Behavior Analysis in Practice, 10*(2), 97–106. <https://doi.org/10.1007/s40617-016-0161-9>.
- Dunn, L., & Dunn, D. (2007). *Peabody picture vocabulary test* (4th ed). Pearson.
- Engelmann, S., Becker, W. C., Carnine, D., & Gersten, R. (1988). The direct instruction follow through model: Design and outcomes. *Education & Treatment of Children, 11*(4), 303–317.
- Engelmann, S., Haddox, P., Hanner, S., & Osborn, J. (2002). *Corrective reading thinking basics*. Science Research Associates/McGraw-Hill.
- Engelmann, S., Engelmann, O., Carnine, D., & Kelly, B. (2012). *SRA connecting math concepts: Comprehensive edition*. McGraw-Hill Education.
- Engelmann, S., & Osborn, J. (2008a). *Language for learning*. Science Research Associates.
- Engelmann, S., & Osborn, J. (2008b). *Reading mastery: Signature edition*. Science Research Associates/McGraw-Hill.
- Engelmann, S., & Silbert, J. (1983). *Expressive writing I*. Science Research Associates.
- *Flores, M. M., & Ganz, J. B. (2007). Effectiveness of direct instruction for teaching statement inferences, use of facts, and analogies to students with developmental disabilities and reading delays. *Focus on Autism & Other Developmental Disabilities, 22*, 244–251.
- *Flores, M. M., & Ganz, J. B. (2009). Effects of direct instruction on the reading comprehension of students with autism and developmental disabilities. *Education & Training in Developmental Disabilities, 44*(1), 39–53. <https://doi.org/10.1177/10883576070220040601>.
- *Flores, M. M., & Ganz, J. B. (2014). Comparison of direct instruction and discrete trial teaching on the curriculum-based assessment of language performance of students with autism. *Exceptionality: A Special Education Journal, 22*(4), 191–204. <https://doi.org/10.1080/09362835.2013.865533>

- *Flores, M. M., Nelson, C., Hinton, V., Franklin, T. M., Strozier, S. D., Terry, L., & Franklin, S. (2013). Teaching reading comprehension and language skills to students with autism spectrum disorders and developmental disabilities using direct instruction. *Education & Training in Autism & Developmental Disabilities, 48*(1), 41–48.
- *Flores, M. M., Schweck, K. B., & Hinton, V. (2016). Teaching language skills to preschool students with developmental delays and autism spectrum disorder using language for learning. *Rural Special Education Quarterly, 35*(1), 3–12. <https://doi.org/10.1177/875687051603500102>
- *Frampton, S. E., Shillingsburg, M. A., & Simeone, P. J. (2020). Feasibility and preliminary efficacy of direct instruction for individuals with autism utilizing speech-generating devices. *Behavior Analysis in Practice, 13*(3), 1–11. <https://doi.org/10.1007/s40617-020-00412-3>
- *Ganz, J. B., & Flores, M. M. (2009). The effectiveness of direct instruction for teaching language to children with autism spectrum disorders: Identifying materials. *Journal of Autism & Developmental Disorders, 39*, 75–83. <https://doi.org/10.1007/s10803-008-06002-2>
- Goldstein, H. (1983). Recombinative generalization: Relationships between environmental conditions and the linguistic repertoires of language learners. *Analysis & Intervention in Developmental Disabilities, 3*(4), 279–293. [https://doi.org/10.1016/0270-4684\(83\)90002-2](https://doi.org/10.1016/0270-4684(83)90002-2).
- Hammill, D., & Newcomer, P. (2008). *Test of oral language development, Intermediate version (4th ed)*. Pro-Ed.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- *Head, C. N., Flores, M. M., & Shippen, M. E. (2018). Effects of direct instruction on reading comprehension for individuals with autism or developmental disabilities. *Education & Training in Autism & Developmental Disabilities, 53*(2), 176–191.
- Heward, W. L., Alber, S. R., & Konrad, M. (2006). *Exceptional children: An introduction to special education*. Pearson Education/Merrill/Prentice Hall.
- Hyman, S. L., Levy, S. E., Myers, S. M., & AAP Council on Children with Disabilities, Section on Developmental & Behavioral Pediatrics. (2020). Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics, 145*(1), 1–64. <https://doi.org/10.1542/peds.2019-3447>.
- *Infantino, J., & Hempenstall, K. (2006). Effects of a decoding program on a child with Autism Spectrum Disorder. *Australasian Journal of Special Education, 30*(2), 126–144. <https://doi.org/10.1017/S1030011200025513>
- *Kamps, D., Heitzman-Powell, L., Rosenberg, N., Mason, R., Schwartz, I., & Romine, R. S. (2016). Effects of reading mastery as a small group intervention for young children with ASD. *Journal of Developmental & Physical Disabilities, 28*(5), 703–722. <https://doi.org/10.1007/s10882-016-9503-3>
- LaFrance, D. L., & Tarbox, J. (2020). The importance of multiple exemplar instruction in the establishment of novel verbal behavior. *Journal of Applied Behavior Analysis, 53*(1), 10–24. <https://doi.org/10.1002/jaba.611>.
- Love, J. R., Carr, J. E., Almason, S. M., & Petursdottir, A. I. (2009). Early and intensive behavioral intervention for autism: A survey of clinical practices. *Research in Autism Spectrum Disorders, 3*(2), 421–428. <https://doi.org/10.1016/j.rasd.2008.08.008>.
- Mitchell, B. S., Adamson, R., & McKenna, J. W. (2017). Curbing our enthusiasm: An analysis of the check-in/check-out literature using the council for exceptional children’s evidence-based practice standards. *Behavior Modification, 41*(3), 343–367. <https://doi.org/10.1177/0145445516675273>.
- National Institute for Direct Instruction (NIFDI). (2021). About the programs. <https://www.nifdi.org/programs/about-the-programs.html>
- Newborg, J. (2005). *Battelle developmental inventory*. Riverside Publishing Company.
- Odom, S. L., Collet-Klingenberg, L., Rogers, S. J., & Hatton, D. D. (2010). Evidence-based practices in interventions for children and youth with autism spectrum disorders. *Preventing School Failure: Alternative Education for Children & Youth, 54*(4), 275–282. <https://doi.org/10.1080/10459881003785506>.
- O’Neill, R. E. (1990). Establishing verbal repertoires: Towards the application of general case analysis and programming. *Analysis of Verbal Behavior, 8*, 113–126. <https://doi.org/10.1007/BF03392852>.
- Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single-case research: nonoverlap of all pairs. *Behavior Therapy, 40*, 357–367. <https://doi.org/10.1016/j.beth.2008.10.006>.
- Pierce, N. P., O’Reilly, M. F., Sorells, A. M., Fragale, C. L., White, P. J., Aguilar, J. M., & Cole, H. A. (2014). Ethnicity reporting practices for empirical research in three autism-related journals. *Journal of Autism & Developmental Disorders, 44*(7), 1507–1519. <https://doi.org/10.1007/s10803-014-2041-x>.
- Ragnarsdóttir, G. A. (2007). Teaching an Icelandic student with autism to read by combining direct instruction and precision teaching. *European Journal of Behavior Analysis, 8*(2), 289–304. <https://doi.org/10.1080/15021149.2007.11434290>.
- Roid, G. H., & Miller, L. J. (2002). *Leiter-R: Leiter international performance scale-revised*. Stoelting.

- Schopler, E., Reichler, R. J., & Rocher-Renner, B. (1988). *The childhood autism rating scale (CARS)*. Western Psychological Services.
- *Shillingsburg, M. A., Bowen, C. N., Peterman, R. K., & Gayman, M. D. (2015). Effectiveness of the direct instruction language for learning curriculum among children diagnosed with autism spectrum disorder. *Focus on Autism & Other Developmental Disabilities*, 30, 44–56. <https://doi.org/10.1177/1088357614532498>
- Smith, T., Scahill, L., Dawson, G., Guthrie, D., Lord, C., Odom, S., Rogers, S., & Wagner, A. (2007). Designing research studies on psychosocial interventions in autism. *Journal of Autism & Developmental Disorders*, 37, 354–366. <https://doi.org/10.1007/s10803-006-0173-3>.
- Steinbrenner, J. R., Hume, K., Odom, S. L., Morin, K. L., Nowell, S. W., Tomaszewski, B., Szendrey, S., McIntyre, N. S., Yücesoy-Özkan, Ş., & Savage, M. N. (2020). *Children, youth, and young adults with autism*. National Clearinghouse on Autism Evidence and Practice.
- Stockard, J., Wood, T. W., Coughlin, C., & Rasplica Khoury, C. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half century of research. *Review of Educational Research*, 88(4), 479–507. <https://doi.org/10.3102/0034654317751919>.
- *Thompson, J.; Wood, C. L.; Test, D. W.; & Cease-Cook, J. (2012). Effects of direct instruction on telling time by students with autism. *Journal of Direct Instruction*, 12, 1–12.
- *Thompson, J. L., Wood, C. L., Preston, A., & Stevenson, B. (2019). Teaching unison responding during small-group direct instruction to students with autism spectrum disorder who exhibit interfering behaviors. *Education & Treatment of Children*, 42(1), 1–23. <https://doi.org/10.1353/etc.2019.0001>
- Vannest, K.J., Parker, R.I., Gonen, O., & Adiguzel, T. (2016). *Single case research: Web-based calculators for SCR analysis*. (Version 2.0) [Web-based application]. Texas A&M University. Retrieved August 4, 2020 from <http://www.singlecaseresearch.org>
- Watkins, C., & Slocum, T. (2004). The components of direct instruction. *Journal of Direct Instruction*, 3, 75–110.
- *Weisberg, P., & Savard, C. F. (1993). Teaching preschoolers to read: Don't stop between the sounds when segmenting words. *Education & Treatment of Children*, 16(1), 1–18.
- Weston, R., Hodges, A., & Davis, T. N. (2018). Differential reinforcement of other behaviors to treat challenging behaviors among children with autism: A systematic and quality review. *Behavior Modification*, 42(4), 584–609. <https://doi.org/10.1177/0145445517743487>.
- *Wolfe, K., Rispoli, M., Taylor, L., & Drasgow, E. (2018). Investigating generalization difficulties during instruction in language for learning. *Advances in Neurodevelopmental Disorders*, 2(1), 75–85. <https://doi.org/10.1007/s41252-017-0049-x>
- Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., Brock, M. E., Plavnick, J. B., Flury, V. P., & Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism & Developmental Disorders*, 45, 1951–1966. <https://doi.org/10.1007/s10803-014-2351-z>.
- Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock Johnson III*. Riverside.