

Identifying and measuring the common elements of naturalistic developmental behavioral interventions for autism spectrum disorder: Development of the *NDBI-Fi*

Autism
2020, Vol. 24(8) 2285–2297
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1362361320944011
journals.sagepub.com/home/aut



Kyle M Frost¹ , Jessica Brian², Grace W Gengoux³ ,
Antonio Hardan³, Sarah R Rieth⁴, Aubyn Stahmer⁵ ,
and Brooke Ingersoll¹

Abstract

Naturalistic developmental behavioral interventions for young children with autism spectrum disorder share key elements. However, the extent of similarity and overlap in techniques among naturalistic developmental behavioral intervention models has not been quantified, and there is no standardized measure for assessing the implementation of their common elements. This article presents a multi-stage process which began with the development of a taxonomy of elements of naturalistic developmental behavioral interventions. Next, intervention experts identified the common elements of naturalistic developmental behavioral interventions using quantitative methods. An observational rating scheme of those common elements, the eight-item *NDBI-Fi*, was developed. Finally, preliminary analyses of the reliability and the validity of the *NDBI-Fi* were conducted using archival data from randomized controlled trials of caregiver-implemented naturalistic developmental behavioral interventions, including 87 post-intervention caregiver–child interaction videos from five sites, as well as 29 pre–post video pairs from two sites. Evaluation of the eight-item *NDBI-Fi* measure revealed promising psychometric properties, including evidence supporting adequate reliability, sensitivity to change, as well as concurrent, convergent, and discriminant validity. Results lend support to the utility of the *NDBI-Fi* as a measure of caregiver implementation of common elements across naturalistic developmental behavioral intervention models. With additional validation, this unique measure has the potential to advance intervention science in autism spectrum disorder by providing a tool which cuts across a class of evidence-based interventions.

Lay abstract

Naturalistic developmental behavioral interventions for young children with autism spectrum disorder share key elements. However, the extent of similarity between programs within this class of evidence-based interventions is unknown. There is also currently no tool that can be used to measure the implementation of their common elements. This article presents a multi-stage process which began with defining all intervention elements of naturalistic developmental behavioral interventions. Next, intervention experts identified the common elements of naturalistic developmental behavioral interventions using a survey. An observational rating scheme of those common elements, the eight-item *NDBI-Fi*, was developed. We evaluated the quality of the *NDBI-Fi* using videos from completed trials of caregiver-implemented naturalistic developmental behavioral interventions. Results showed that the *NDBI-Fi* measure has promise; it was sensitive to change, related to other similar measures, and demonstrated adequate agreement between raters. This unique measure has the potential to advance intervention science in autism spectrum disorder by providing a tool to measure the implementation of common elements across naturalistic developmental behavioral intervention models.

¹Michigan State University, USA

²Bloorview Research Institute, Canada

³Stanford University, USA

⁴San Diego State University, USA

⁵University of California, Davis, USA

Corresponding author:

Kyle M Frost, Michigan State University, 69F Psychology, 316 Physics Road, East Lansing, MI 48824, USA.
Email: frostkyl@msu.edu

Given that naturalistic developmental behavioral interventions have numerous shared strategies, this may ease clinicians' uncertainty about choosing the "right" intervention package. It also suggests that there may not be a need for extensive training in more than one naturalistic developmental behavioral intervention. Future research should determine whether these common elements are part of other treatment approaches to better understand the quality of services children and families receive as part of usual care.

Keywords

autism spectrum disorders, common elements, interventions—psychosocial/behavioral, measure development

Introduction

Current best practices for the treatment of young children on the autism spectrum include interventions that integrate developmental and behavioral approaches and include caregivers in children's treatment (National Research Council, 2001; Zwaigenbaum et al., 2015). There is a growing evidence base for several such manualized interventions, broadly classified as naturalistic developmental behavioral interventions (NDBI), which supports their positive influence on children's development trajectories (Schreibman et al., 2015). These interventions embed teaching in naturalistic contexts rather than highly structured environments, and emphasize spontaneous initiation rather than repeated responding to adult-led prompts (Tiede & Walton, 2019). Despite support for the efficacy of both therapist- and caregiver-implemented NDBIs (Tiede & Walton, 2019), our knowledge of core intervention elements and treatment mechanisms in these interventions remains limited. Though NDBI developers acknowledge their individual interventions share several common elements despite differing theoretical perspectives (Schreibman et al., 2015), the extent to which the models are similar in practice has not been addressed systematically. Furthermore, researchers studying the various models do not articulate or measure these elements in the same way and often identify different components as fundamental to their approach. Thus, researchers and practitioners alike may benefit from additional clarity regarding which specific elements are most effective or necessary for improving outcomes.

Development of an intervention taxonomy, or shared set of intervention elements, can support our understanding of evidence-based interventions by providing the field with standardized language and a way to describe and compare intervention ingredients across studies (Chorpita et al., 2005; Lokker et al., 2015; McHugh et al., 2009). Identifying common elements across similar evidence-based treatments allows for a more nuanced understanding of how these treatments work. Shifting the unit of analysis from a whole treatment package to individual elements (Chorpita & Daleiden, 2009) supports the identification of potentially active ingredients of existing NDBIs (Embry & Biglan, 2008; Tate et al., 2016). Although common elements are not necessarily responsible for therapeutic change, their inclusion across multiple treatment packages suggests that they may be good candidates to consider in empirical

research (Garland et al., 2008). Accordingly, common elements of evidence-based interventions have been examined in the context of many types of behavioral treatments for children with mental health concerns, including those targeting disruptive behavior disorders (Garland et al., 2008; Kaehler et al., 2016) and parenting skills (Barth & Liggett-Creel, 2014). In addition, identifying common elements can facilitate the development of a standardized measure to better characterize similarities among treatment groups (Godfrey et al., 2007), including active treatment and treatment-as-usual control groups.

A focus on individual elements of intervention packages may also improve the measurement of treatment fidelity. Measuring treatment fidelity, or adherence to the intervention protocol, is essential for understanding how treatments work, and for interpreting the results of intervention trials (Wainer & Ingersoll, 2013). However, most reports of treatment fidelity in the literature provide summary ratings, such as overall percent adherence to the entire treatment protocol. NDBI studies rarely link fidelity of specific intervention elements directly to intervention outcomes (see Gulsrud et al., 2016 for a notable exception); therefore, it is unclear how elements contribute to improvements in child social communication. Furthermore, among NDBIs, measures of treatment fidelity used for research often remain unpublished; therefore, limited data exist regarding which strategies contribute to the overall rating. To our knowledge, NDBI intervention fidelity measures have not been examined psychometrically in a published study, which limits the understanding of their validity, reliability across short time intervals, or sensitivity to change. Without common terminology to describe intervention elements and a common measurement tool for reporting fidelity, researchers cannot easily compare intervention elements across studies. This limitation hinders our ability to understand the key elements of NDBIs associated with positive outcomes. Finally, implementation science has recently highlighted the importance of treatment fidelity for establishing and maintaining high-quality services among community providers over time (Hogue et al., 2015). Thus, the development of an NDBI fidelity tool that can guide training for community providers would be extremely helpful.

Owing to the fact that best practice in early intervention includes caregiver involvement (Wong et al., 2015), some NDBIs have been designed specifically for caregiver

delivery (Brian et al., 2016; Ingersoll & Dvortcsak, 2019), while others have been adapted to caregiver-implemented formats (e.g. Kaiser et al., 2000, 2014; Rogers et al., 2012). However, efficacy research of caregiver-implemented NDBIs has been mixed, with some studies finding significant gains (e.g. Bradshaw et al., 2017; Brian et al., 2017; Gulsrud et al., 2016), and others finding null results (e.g. Rogers et al., 2012). Reasons for the null effects remain unclear and could be due to multiple factors, such as a lack of efficacy, individual differences in treatment response, low treatment fidelity for key intervention ingredients, and/or high quality of community care received by control groups. Another factor unique to caregiver-implemented interventions is that caregivers vary in their implementation of intervention strategies both before and after training (Stahmer et al., 2017). This suggests that improving the measurement of caregiver intervention fidelity is an important avenue for understanding the efficacy of caregiver-implemented NDBIs.

Despite the similarity of key intervention techniques across NDBIs, researchers have not developed a defined set of common intervention elements or a standardized measure for assessing intervention fidelity. This project begins to address these gaps through the following goals: (1) develop a taxonomy of elements of NDBIs, (2) identify the common elements across NDBI models, (3) develop an observational rating scheme to measure the common elements, and (4) establish preliminary reliability and validity of the new measure with a sample of children with autism spectrum disorder (ASD) and their caregivers who participated in a several randomized controlled trials (RCTs) of different caregiver-implemented NDBI models. Caregiver-implemented models were strategically selected for our preliminary validation sample because, unlike trained therapists in RCTs, caregivers have great variability in their implementation of intervention techniques in control and treatment groups, thus allowing the measurement of the full range of intervention implementation.

The current study

This research comprised a multi-step process which prioritized content validity in the development and validation of an intervention-independent fidelity measure (McKenzie et al., 1999). The steps are depicted in Figure 1. Phase 1 describes the process and results of developing a broad taxonomy of NDBI techniques and the identification of NDBI common elements. Phase 2 describes the subsequent development and evaluation of the *NDBI-Fi*, an observational rating scheme for measuring adherence to the common elements of NDBIs. An observational rating scheme was selected because this approach is considered the gold standard in fidelity measurement in treatment efficacy trials given its potential for providing objective and highly specific information regarding intervention

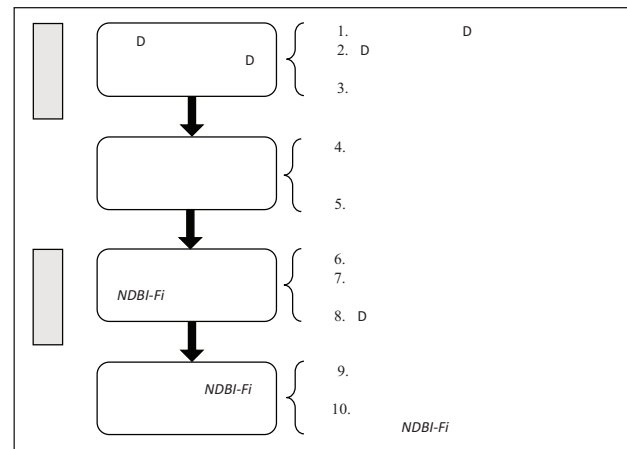


Figure 1. Method flowchart.

providers in session behavior (Hogue et al., 1996; Mowbray et al., 2003). In addition, observational ratings are more likely to be able to detect gradations in quality than indirect (e.g. therapist- or client-report) methods (Schoenwald et al., 2011), making them potentially useful as a quality improvement tool.

Phase I

Method

Intervention taxonomy. Because our aim was to develop an observational fidelity tool that could measure common elements of NDBI, we began by reviewing individual fidelity measures. We focused on therapeutic content (i.e. NDBI strategies) rather than other potentially important aspects of caregiver-implemented interventions, such as treatment techniques performed by the coach/therapist to help the parent learn and apply the therapeutic content, aspects of the therapeutic alliance, or other treatment parameters. Though the therapist/coach's skills to effectively teach caregivers are crucial in caregiver-implemented interventions, the current study focuses on the specific strategies of individual NDBI models that are directed toward the child. This is not to suggest that these other facets of the intervention, such as quality of coaching, goal setting, and duration of treatment, are not important, but rather that they do not fit within the goal of this study.

The first and last author requested published and unpublished NDBI fidelity measures from an expert panel of doctoral-level intervention developers, authors, and experts to develop a broad taxonomy of NDBI elements. Several authors of the Schreibman et al.'s (2015) paper, as well as known colleagues who have conducted RCTs of the interventions identified by Schreibman and colleagues in their seminal paper were invited by email to collaborate. Each of these interventions has been examined in a research context

and has demonstrated some evidence of efficacy as a therapist-delivered and/or caregiver-implemented intervention (Sandbank et al., 2020; Tiede & Walton, 2019). A total of 11 research teams (14 individuals; 8 interventions) were contacted. One research team did not respond. Interventions examined included *Early Achievements* (Landa et al., 2011), *Early Start Denver Model* (ESDM; Rogers & Dawson, 2010), *Enhanced Milieu Teaching* (EMT; Kaiser et al., 2000; Kaiser & Hester, 1994), *Joint Attention, Symbolic Play, Engagement, and Regulation* (JASPER; Kasari et al., 2006, 2010) *Pivotal Response Training* (PRT; Hardan et al., 2015; Schreibman & Koegel, 2005), *Project ImPACT* (Ingersoll & Dvortcsak, 2010), and *Social Antecedent-Behavior-Consequences* (ABCs) (Brian et al., 2016, 2017). While the intervention approaches used in this study do not represent a comprehensive list of all interventions that could be characterized as NDBI, those with expertise in the above interventions agreed to collaborate on this endeavor and they represent models commonly used in the literature. Furthermore, we did not examine classroom-based interventions due to the unique features of group instruction and this study's focus on parent-child interactions.

The first and last authors established a preliminary taxonomy of intervention elements by examining the content of available NDBI fidelity rating forms ($n=9$).¹ The taxonomy was inclusive of intervention-specific elements (i.e. not common across all interventions), as well as those shared among multiple interventions. The process included formally defining each of the elements based on the content of the examined fidelity forms, internally refining the taxonomy over several iterations and generating examples and non-examples for each element to further clarify the definitions. The preliminary taxonomy was then refined using an adapted Delphi method. As per the Delphi method, the expert panel representing the NDBI (identified above) received the preliminary taxonomy and provided open-ended critique and commentary; they were also encouraged to add intervention elements not included in the original taxonomy. Four individuals shared the information with an additional person on their research team to respond in addition to or in place of themselves. A total of 12 individuals responded across all of the seven identified interventions (Table 1). The internal team subsequently revised the definitions and examples, yielding a refined taxonomy of 20 unique intervention elements.

Item reduction. Next, a survey was used to obtain quantitative feedback on the refined taxonomy to reduce items to the common elements and increase the content validity of the item set. The members of our expert panel nominated survey respondents who they would consider "experts in their intervention (e.g. past grad students, qualified intervention trainers, etc.)." A total of 25 individuals were nominated, 21 of whom responded to our online survey (85%).

Table 1. Number of fidelity tools, expert panel members, and survey respondents per NDBI.

Interventions	Fidelity tools (n)	Expert panel (n)	Survey respondents (n)
Project ImPACT	2	2 ^a	4
ESDM	1	2	3
JASPER	1	2	2
Social ABCs	1	0	4
PRT	2	4	3
EMT	1	1	3
Early Achievements	1	1	2
Multiple NDBI	–	2	–

NDBI: naturalistic developmental behavioral interventions; ESDM: Early Start Denver Model; JASPER: Joint Attention, Symbolic Play, Engagement, and Regulation; PRT: Pivotal Response Training; EMT: Enhanced Milieu Teaching.

^aTwo experts in Project ImPACT were lead authors.

The survey presented the text for the 20 elements from the taxonomy described above. Survey respondents rated the extent to which each element was a part of the intervention protocol in which they had expertise, using the following scale (adapted from Lawshe, 1975):

- *Essential*: This item is a component of [intervention], and it is described explicitly in the intervention manual. Interventionists use it consistently during sessions.
- *Useful but non-essential*: This item is a good clinical practice, and interventionists use it when providing [intervention], but it is not described in the intervention manual.
- *Neutral*: I would not discourage the use of this strategy when providing [intervention], but interventionists do not typically use it, and it is not described in the intervention manual.
- *Conflicting*: This item conflicts with the [intervention] intervention protocol. Intervention trainees and caregivers are discouraged from using this strategy.

This scale was selected because of its distinction between "essential" and "useful but non-essential" elements, which provided information on both manualized and non-manualized intervention elements.

Next, content validity ratios (CVRs) were calculated for each item, using the following formula $CVR = \frac{n_e - N/2}{N/2}$, where n_e is the number of survey respondents indicating a particular response, and N is the total number of respondents (Veneziano & Hooper, 1997). Survey responses from up to three individuals per intervention were used to calculate the CVR, with two intervention teams contributing only two responses. Additional responses were dropped from analysis to avoid unequal weighting of any one

Table 2. CVRs for intervention taxonomy items.

Item		Essential or useful	Essential
1	Face-to-face and on the child's level*	0.89	0.68
2	Preparing the activity space	0.89	0.37
3	Following the child's lead*	0.89	0.89
4	Imitating the child	0.37	0.05
5	Supporting turn-taking	0.79	0.26
6	Displaying positive affect and animation*	1.00	0.68
7	Engaging the child in play routines	0.79	0.16
8	Engaging the child in social routines	0.79	-0.37
9	Managing problem behavior and dysregulation	1.00	0.37
10	Modeling appropriate language*	1.00	0.58
11	Modeling gestures and joint attention	0.47	0.05
12	Modeling new play acts	0.79	0.37
13	Responding to attempts to communicate*	0.89	0.89
14	Using communicative temptations*	1.00	0.79
15	Pace and frequency of direct teaching opportunities*	0.89	0.58
16	Varying difficulty of direct teaching target	0.68	0.05
17	Using clear and appropriate teaching opportunities*	0.79	0.79
18	Providing motivating and relevant teaching opportunities*	1.00	1.00
19	Supporting a correct response using prompts*	0.68	0.37
20	Providing contingent natural and social reinforcement*	0.89	0.79

*Denotes items included in the *NDBI-Fi* measure; bold text denotes items exceeding the statistically significant cut-off of 0.42.

intervention model over another; individuals with the least amount of self-reported intervention experience were dropped. A total of 19 responses were analyzed, representing all seven NDBIs. Results were identical when all available survey data were used; we opted to present CVR results from 19 cases so that the results would not be biased toward any one intervention package.

The CVR, which quantifies consensus, was used to quantitatively evaluate the extent to which each item was characteristic of NDBIs. The published recommended cut-off for achieving statistically significant agreement with our sample size (0.42) was used to determine which items would be retained in the final measure (Lawshe, 1975; Veneziano & Hooper, 1997). CVRs were calculated for each item in two ways: (1) considering the number of respondents indicating a score of "essential" only and (2) considering the respondents who indicated a score of "essential" or "useful but non-essential." Examination of items rated as "essential" accounts for techniques specified explicitly in NDBI manuals. The addition of items rated "useful but non-essential" accounts for the fact that clinicians often draw on additional clinical skills when providing intervention beyond what is specified in a treatment manual.

Results

Intervention taxonomy. The broad taxonomy consisted of a total of 20 elements with definitions agreed on by our expert panel (Table 2; full definitions in Supplemental Material). Given the differences in terminology often used across

NDBI models, these refined definitions may be useful in translating information across research teams and in the community and better defining similarities and differences between interventions.

Item reduction. CVRs for "essential" items only and for "essential" or "useful but non-essential" elements are provided in Table 2. When considering both items rated "essential" and "useful but non-essential," all but one element of the 20 exceeded the cut-off indicating consensus across interventions. When considering only elements rated "essential," 10 of the 20 items exceeded the cut-off indicating consensus. One additional element, which referred to the use of prompting to support the child's response, was examined further and refined based on feedback from the survey respondents. Specifically, some interventions used a specific prompting hierarchy that was precluded based on the original wording of the item; therefore, the prompting item was modified to contain more generic language and was included in the final set of 11 common essential elements. Following the revision of this item, no items were rated as "conflicting" by more than one survey respondent.

Phase 2

Method

Participants. This study involved analyzing existing data from completed or ongoing treatment trials of caregiver-implemented NDBIs with children with ASD aged 7 years

Table 3. Participant demographics.

Children			Caregivers		
Child's biological sex	<i>n</i>	%	Caregiver's biological sex	<i>n</i>	%
Male	71	81.6	Male	8	9.2
Female	16	18.4	Female	79	90.8
Child's race	<i>n</i>	%	Mother's highest complete education	<i>n</i>	%
White/Caucasian	48	55.2	Graduate/Professional degree	23	26.4
Black/African American	7	8	Bachelor's degree	29	33.3
Asian/Pacific Islander	19	21.8	Associate's degree	5	5.7
Biracial/mixed race	2	2.3	High school degree/GED	21	24.1
Other	8	9.2	Did not complete high school	1	1.1
Child's ethnicity	<i>n</i>	%	Missing	8	9.2
Non-Hispanic/Latino	71	81.6	Father's highest complete education	<i>n</i>	%
Hispanic/Latino	15	17.2	Graduate/Professional degree	22	25.3
MSEL	<i>M</i>	<i>SD</i>	Bachelor's degree	14	16.1
Average AE (months)	22.1	7.6	Associate's degree	5	5.7
Chronological age (months)	37.2	13.8	High school degree/GED	12	13.8
			Did not complete high school	0	0.0
			Missing	34	39.1

MSEL: Mullen Scales of Early Learning; SD: standard deviation; AE: age equivalent.

or younger. This age range was selected to be consistent with intervention trials of NDBI. Five sites contributed videos of caregiver–child play interactions with representation from four interventions, including Project ImPACT (Ingersoll & Dvortcsak, 2010; Ingersoll et al., 2016)/ Project ImPACT for Toddlers (Stahmer et al., 2017, 2020), JASPER (Kasari et al., 2006, 2010), PRT (Hardan et al., 2015; Schreibman & Koegel, 2005), and Social ABCs (Brian et al., 2016, 2017). This study was approved by the Institutional Review Board (IRB) at Michigan State University, and sharing of videos was approved by IRBs at external study sites. All families consented for their videos to be used for research purposes. The study sample included 87 caregiver–child dyads randomized to either active treatment or control groups. Demographic information is reported in Table 3.

Measures

NDBI-Fi. The 11 quantitatively derived “essential” common elements from Phase 1 were used to develop an observational rating scheme and scoring manual for the *NDBI-Fi* measure. The measure used a macro-level rating scheme (i.e. a 1–5 rating scale) to align with many existing fidelity measures (67% of those included in this study) and to increase the likelihood that the measure would not be burdensome or costly to use. The *NDBI-Fi* manual includes practical considerations for rating, item definitions, examples and non-examples, a glossary, and descriptive anchors for assigning ratings. Of the 11 common items from Phase 1, one item specified the *frequency* of direct teaching episodes; these teaching episodes comprise a multi-step procedure based on the principles of operant conditioning with an antecedent-behavior-consequence

(ABC) structure. Four additional items focused on the *quality* of direct teaching episodes (*Clear and appropriate, Motivating and relevant, Supporting a correct response, and Providing contingent and natural reinforcement*). These were collapsed into a single item, *Quality of direct teaching*, to facilitate ease of coding and ensure that full teaching trials were being scored. Additional items include *Face-to-face and on the child's level, Following the child's lead, Displaying positive affect and animation, Modeling appropriate language, Responding to attempts to communicate, and Using communicative temptations*. Thus, the *NDBI-Fi* consists of an eight-item rating scheme (Table 4). The measure is available in the Supplemental Material and from the corresponding author.

A total of two raters, including the first author and a research assistant, piloted the rating scheme on a small set of videos to refine the descriptive rating anchors and to achieve inter-rater reliability. One rater was a graduate student with direct intervention experience in three different NDBI models, while the other rater was an undergraduate research assistant without direct intervention experience. Raters discussed scoring differences and refined items and rating anchors to improve clarity and ease of scoring. These two raters independently coded videos and held consensus meetings to discuss discrepancies in ratings until inter-rater reliability was met. Raters were considered reliable when they could rate three consecutive not previously reviewed videos according to the following criteria: (1) at least seven out of eight items were within 1 point, (2) no items were greater than 2 points apart, and (3) the average score was within 0.5 points (i.e. ± 0.25 points). The primary rater was kept blind to treatment condition for all videos; the secondary rater was kept blind to treatment

Table 4. *NDBI-Fi* item descriptions.

Item	Brief description
1. Face-to-face and on the child's level	<ul style="list-style-type: none"> • Child and adult facing each other • Child and adult on similar level
2. Following the child's lead	<ul style="list-style-type: none"> • Child and adult are both active participants in child-chosen activity
3. Positive affect and animation	<ul style="list-style-type: none"> • Adult uses positive affect • Adult matches affect to child's sensory needs
4. Modeling appropriate language	<ul style="list-style-type: none"> • Adult adjusts language to the child's developmental level
5. Responding to attempts to communicate	<ul style="list-style-type: none"> • Adult verbally responds to child's communication behaviors by repeating, clarifying, or expanding
6. Using communicative temptations	<ul style="list-style-type: none"> • Adult creates situations to elicit communication from the child followed by a brief period of expectant waiting
7. Frequency of direct teaching episodes	<ul style="list-style-type: none"> • Adult directs the child to demonstrate new or emerging skills
8. Quality of direct teaching episodes	<ul style="list-style-type: none"> • Adult uses high-quality teaching strategies (e.g. clear instructions, when child is motivated, contingent reinforcement).

Table 5. Number of videos examined per intervention across group and time point.

Intervention	Pre-intervention		Post-intervention	
	Treatment (n)		Treatment (n)	Control (n)
Project ImPACT	24		24	9
Project ImPACT for Toddlers	0		8	9
PRT	0		10	10
Social ABCs	5		6	0
JASPER	0		6	5

PRT: Pivotal Response Training; JASPER: Joint Attention, Symbolic Play, Engagement and Regulation.

condition when possible (39% of double-coded videos). One rater was involved in data collection for a subset of videos, and as such, blinding of both raters was not possible for these select cases.

Caregiver–child interaction. Videos included caregiver–child interactions from existing treatment trials. All videos involved an approximately 10-min free-play interaction between the child and the caregiver. Sites selected videos that included English-speaking participants within the treatment and control groups at random, using an online random number generator (<https://www.random.org/integer-sets>). A total of 87 post-timepoint videos were collected from five intervention trials (JASPER, Project ImPACT, Project ImPACT for Toddlers, PRT, and Social ABCs), including 54 videos of dyads who received treatment and 33 videos of control participants (i.e. waitlist or treatment-as-usual; Table 5). In addition, 29 pre–post video pairs from two of the sites (Project ImPACT and Social ABCs) were used to examine sensitivity to change.

Established NDBI fidelity. Caregiver treatment adherence using the established fidelity measure for each intervention was available for 76 post-treatment videos (representing Project ImPACT, Project ImPACT for Toddlers, PRT, and Social ABCs). Because intervention fidelity forms utilized different scales (Table 6), scores were

transformed as necessary so that all fidelity ratings were on the same scale (with a minimum score of 1 and a maximum score of 5).

Mullen Scales of Early Learning. The Mullen Scales of Early Learning (MSEL; Mullen, 1995) is a standardized cognitive assessment with four domains that evaluate skills in the domains of visual reception, fine motor, expressive language, and receptive language. The MSEL was administered for all five intervention trials at the study sites. Age equivalent scores across all the four MSEL domains were averaged to obtain an overall estimate of child's developmental level.

Analysis plan. An exploratory factor analysis was used to evaluate the dimensionality of the *NDBI-Fi*, and Cronbach's alpha was subsequently used to evaluate internal consistency. In addition, two raters coded a total of 52 videos (60%) from three sites. Intraclass correlations (ICCs) were used to evaluate the agreement between coders on individual items as well as overall score. The ICC is the preferred metric for this type of scale; furthermore, it incorporates the magnitude of disagreement into the metric, yielding a more precise estimate of reliability than the metrics of all-or-nothing agreement (Hallgren, 2012). A single-measure, two-way mixed design based on absolute agreement was used.

Table 6. Characteristics of established NDBI fidelity measures.

Intervention	<i>n</i> items (subscales)	Rating scale	Type of coding	Example items
PRT ^a	8 (3)	0–1	Interval (1 min)	<ul style="list-style-type: none"> • Child choice • Clear opportunity
PRT ^b	6 (0)	0–1	Interval (2 min)	<ul style="list-style-type: none"> • Contingent reinforcement • Natural reinforcement
Project ImPACT ^c	29 (5)	1–5	Global	<ul style="list-style-type: none"> • Expands on child's language • Uses clear and appropriate prompts
Project ImPACT for Toddlers ^a	19 (7)	1–5	Global	<ul style="list-style-type: none"> • Makes comments and avoids questions • Adjusts levels of prompts based on child responding
Social ABCs	10 (0)	0–1	Interval (1 min)	<ul style="list-style-type: none"> • Positive emotion • Shared control

PRT: Pivotal Response Training.

^aUniversity of California–San Diego site.

^bStanford University site.

^cMichigan State University site.

To address concurrent validity, an independent samples *t*-test was used to determine if caregivers who received training differed from those who did not at the post-intervention timepoint. We hypothesized that caregivers in the active study treatment groups across trials would receive a significantly higher *NDBI-Fi* rating at the end of the treatment phase than caregivers in control groups.

Convergent and discriminant validity were examined using Pearson correlation to test the relationship between the *NDBI-Fi* and relevant constructs. We expected that overall ratings for the *Established NDBI Fidelity* would be significantly correlated with the *NDBI-Fi Average Rating* with a medium to large effect size. Next, we expected that the *NDBI-Fi* would not be related (i.e. a small effect size, $r < 0.2$) to child factors, such as child's chronological age or child's developmental age equivalent, which might impact parent–child interactions.

For a measure, such as the *NDBI-Fi* to be useful in the context of intervention research, it must capture change in parent behaviors as they learn intervention techniques. To evaluate the sensitivity of the *NDBI-Fi* in capturing change in this context, the available subset of videos of the same dyads' pre- and post-training was rated. This analysis only included dyads in treatment conditions, though the structure and intensity of training offered to caregivers were likely different across sites. A paired sample *t*-test was used to assess for significant change in caregiver use of techniques from pre- to post-training. We expected that, on average, caregivers would score significantly higher on the *NDBI-Fi* after participating in the intervention.

Results

The *NDBI-Fi* average score ($M=3.28$, $SD=0.75$) was adequately normally distributed (Figure 2), with skewness of -0.13 ($SE=0.26$) and kurtosis of -0.84 ($SE=0.51$). Some individual items deviated from normality according to skewness and kurtosis values (Table 7), including a low-frequency

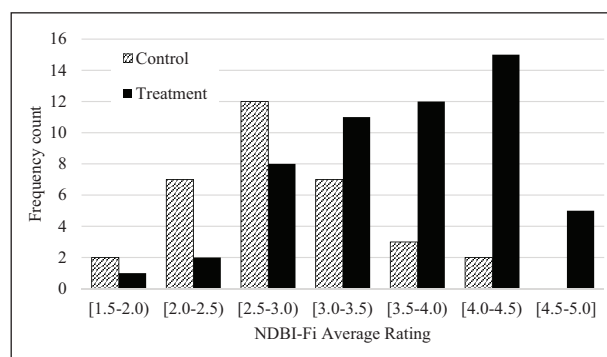


Figure 2. Frequency distribution for the *NDBI-Fi* average rating for treatment cases and control cases.

behavior with positive skew (6. Communicative Temptations) and some high-frequency behaviors with negative skew (e.g. 7. Frequency of Direct Teaching). An exploratory factor analysis of all post-timepoint *NDBI-Fi* ratings was conducted using principal axis factoring. Two factors were extracted with eigenvalues greater than 1 (3.43, 1.12); however, the scree plot demonstrated a clear “elbow” at factor two, suggesting a one-factor solution fits the data best.

Reliability. The eight *NDBI-Fi* items as a scale yielded a Cronbach's alpha of 0.80, thereby demonstrating good internal consistency. Inter-item correlations ranged from 0.11 to 0.65. The single-measures ICC for the *NDBI-Fi* average rating demonstrated excellent reliability (Cicchetti, 1994). Individual item ICCs ranged from 0.33 to 0.82 (Table 6); two items had poor to fair reliability, four items had good reliability, and two items had excellent reliability (Cicchetti, 1994).

Concurrent validity. An independent samples *t*-test was used to compare post-timepoint ratings for caregivers in the active study treatment groups ($n=54$) and control groups ($n=33$). Caregivers who received training ($M=3.56$,

Table 7. Mean, standard deviation, normality, and reliability of *NDBI-Fi* items at post-intervention.

<i>NDBI-Fi</i> item	<i>M</i>	<i>SD</i>	Skewness		Kurtosis		ICC
			Statistic	<i>SE</i>	Statistic	<i>SE</i>	
1. Face-to-Face	2.81	1.34	0.28	0.26	-1.10	0.51	0.82
2. Follow child's lead	3.58	1.36	-0.60	0.26	-0.93	0.51	0.63
3. Positive affect	3.64	1.28	-0.54	0.26	-0.93	0.51	0.78
4. Modeling language	3.17	1.06	-0.17	0.26	-0.87	0.51	0.61
5. Responding to communication	3.28	1.08	-0.12	0.26	-0.63	0.51	0.52
6. Communicative temptations	1.82	1.15	1.22	0.26	0.29	0.51	0.70
7. Frequency of direct teaching	4.05	1.00	-1.24	0.26	1.50	0.51	0.74
8. Quality of direct teaching	3.93	0.89	-0.60	0.26	-0.24	0.52	0.33
Average score	3.28	0.75	-0.13	0.26	-0.84	0.51	0.80

SD: standard deviation; *ICC*: intraclass correlation; *SE*: standard error.

SD=0.69) received higher *NDBI-Fi* average ratings than caregivers in the study control groups on average, with a large effect size ($M=2.81$, $SD=0.62$), $t(85)=5.09$, $p<0.001$, $d=1.12$. However, there was overlap in the frequency distributions of trained and untrained caregivers, with some untrained caregivers demonstrating high fidelity and some trained caregivers demonstrating low fidelity (Figure 2) at the end of the treatment phase.

Convergent and discriminant validity. A Pearson correlation showed that the *NDBI-Fi* average rating correlated significantly with individual established intervention fidelity with a large effect size ($r=0.60$, $p<0.001$). As expected, caregivers who performed the interventions at higher fidelity also received higher ratings on the *NDBI-Fi*. Pearson correlations revealed that the *NDBI-Fi* average rating did not significantly correlate with either developmental level ($r=0.21$, $p=0.06$) or child's chronological age at the start of the study ($r=0.01$, $p=0.92$).

Sensitivity to change. Caregivers who received intervention training scored significantly higher at post-intervention on the *NDBI-Fi* average rating ($M=3.56$, $SD=0.61$) than at pre-intervention ($M=2.89$, $SD=0.72$), $t(28)=4.22$, $p<0.001$, $d=1.00$.

Discussion

Various *NDBI*s for young children with ASD have been independently developed and validated. While researchers acknowledge common elements across these treatments (Schreibman et al., 2015), this study represents the first attempt to evaluate the extent to which experts systematically agree that individual elements are shared across manualized treatment packages. In addition, we present preliminary validation data of a unique measure designed to capture caregiver implementation of common intervention techniques across five *NDBI* trials.

Development of the *NDBI-Fi* began with the creation of a taxonomy of *NDBI* intervention techniques (see

Supplemental Material). This collaborative effort yielded a list of 20 defined elements, refined by expert clinical scientists representing seven different *NDBI*s, with accompanying examples and non-examples to illustrate them. Subsequent findings identified 11 "essential" common elements shared across *NDBI*s. These included elements, such as being face-to-face and on the child's level, following the child's lead, modeling language, positive affect and animation, responding to the child's attempts to communicate, using communicative temptations, and the frequency and quality of direct teaching episodes. Furthermore, these elements were examined across five trials of four different *NDBI*s to validate an intervention-independent fidelity measure. The *NDBI-Fi* demonstrated adequate psychometric properties, as well as preliminary evidence of convergent validity and sensitivity to change. Results suggested that the reliability of some items was limited, and attempts should be made to improve these items or adjust coding practices to support higher reliability. In particular, the inter-rater reliability for the *Quality of direct teaching* item suggests the need for further refinement. Although evidence is preliminary, it is our hope that the ongoing development of this measure will help spark innovative research that cuts across interventions by providing a mechanism for measuring the implementation of common elements of *NDBI*s during intervention trials.

The *NDBI-Fi* item development process revealed that many clinical best practices are shared among *NDBI*s but not necessarily included across all *NDBI* treatment manuals and fidelity forms. This was indicated by a discrepancy in the number of items for which there was consensus while examining "essential" ratings only (i.e. items explicitly described in the intervention manual; $n=11$) as compared to a combination of "essential" and "useful but non-essential" ratings (i.e. items implemented but not manualized; $n=19$). This result suggests that eight of the broad items are commonly implemented while delivering *NDBI*s regardless of whether these practices are defined in their treatment manuals or fidelity measures. The presence of these common practices may compromise direct comparison of

different NDBIs and obscure our understanding of which techniques promote improvement in child outcomes.

While we found mean-level group differences in *NDBI-Fi* scores between caregivers with and without training, our data also demonstrated variability within these groups, with some untrained caregivers demonstrating the use of several NDBI strategies and some trained caregivers demonstrating the limited use of strategies. This highlights the fact that many of these intervention strategies are also natural parenting techniques that families may use to some extent without training. However, the extensiveness of implementation likely varies across families. In future research, it will be important to consider how *change* in caregiver fidelity of implementation relates to child improvement, in addition to standard between-group comparisons. In practice, this finding has implications for the use of stepped-care models in caregiver-implemented interventions for ASD (Phaneuf & McIntyre, 2011; Wainer & Ingersoll, 2015). Caregivers who do not intuitively use many of these strategies may have the most to gain from training and may require a higher level of support to be successful. However, caregivers who do intuitively use some NDBI strategies may benefit from less intensive training, or training targeting other areas of need.

Finally, research in implementation science has documented barriers to providing evidence-based interventions in the community for social services more broadly (Osterling & Austin, 2008; Pagoto et al., 2007) and for ASD interventions specifically (Brookman-Frazee et al., 2016; Pickard et al., 2016; Suhrheinrich et al., 2020; Wood et al., 2015). Research suggests that practitioners have concerns about the use of packaged treatment manuals, perhaps due to the perceived inflexibility of treatment manuals, or difficulty knowing which treatment manual(s) to use at what time. This study demonstrates that NDBIs have numerous shared strategies, which may alleviate clinicians' uncertainty about choosing the "right" intervention package. It also suggests that there may not be a need for extensive training in more than one NDBI, given the demonstrated overlap across treatment models.

Limitations and future directions

This report constitutes a preliminary validation of the *NDBI-Fi*. Future research should attempt to evaluate this measure across additional NDBIs. Analyses using a greater number of caregiver-child interaction videos and intervention models would allow for a more rigorous assessment of the validity of the measure. This study did not account for the dose and duration of intervention due to limited space and a focus on evaluating the *NDBI-Fi* measure; however, this may be possible in future research. While we found preliminary evidence that the measure was sensitive to change from pre- to post-intervention among caregivers who received NDBI training, a group by time interaction would be a more rigorous test. Particularly given that some

caregivers obtained high scores without training, a more in-depth assessment of change is warranted, including shorter term changes and changes that may occur without intervention. Comparing the *NDBI-Fi* and other fidelity measures in terms of their sensitivity to change would be useful to better understand this issue. Furthermore, data on inter-rater reliability suggest that while training someone without direct intervention experience in rating the *NDBI-Fi* can be achieved, it yields reliability estimates that are acceptable but could be improved.

In the item reduction stage of Phase 1 of this study, we selected a pool of respondents hand-picked by NDBI developers. We did so because these individuals are intimately familiar with the interventions as they are meant to be delivered. However, in the community, these interventions may be delivered alongside other treatments or merged with other types of treatment elements not considered part of NDBIs. This group of individuals could not speak to how community providers may use these elements, or whether these elements are parts of other types of interventions as well. Future research should attempt to clarify if and how often the techniques we identified are utilized in different intervention approaches, such as more structured approaches based on applied behavior analysis, or those used in special education and speech-language pathology. Understanding the extent to which NDBI intervention elements are part of other treatment approaches is important for understanding what exactly comprises "usual care" early intervention services. Such work is essential for refining our understanding of what constitutes NDBIs as a class of interventions and how they are distinct from other practices in early intervention. However, the *NDBI-Fi* was not designed to evaluate the full breadth of intervention techniques found in other types of interventions and cannot be used to evaluate the quality of such services. Nonetheless, it is our hope that the iterative development process of this measure may prove useful for characterizing other types of treatments as well.

This study was limited to examining common elements used across a selection of NDBIs for young children with ASD, and we consider it the first step in an ongoing process of better characterizing and measuring this class of early interventions. It is important to reiterate that these common elements are not necessarily the most important or "active" ingredients responsible for child change; they are simply items that were common across several manualized treatment packages. While NDBIs are acknowledged to have key similarities, it is not known whether they also share active ingredients or exert change in unique ways. Identification of common elements is necessary to determine the unique features of individual interventions as well. To develop the science of NDBIs and better understand the active ingredients and mechanisms through which they exert change, researchers will need to build upon this and other work to understand the full range of treatment elements that comprise these complex interventions. Furthermore, understanding how these treatments work will require the design

of creative experimental studies that can examine the causal relationship between implementing specific treatment techniques and child outcomes. For example, single-case experimental designs or group designs (e.g. dismantling trials, factorial experiments), which systematically examine the effects of these common elements, could reveal which, if any, of them contributes to observed changes in child social communication (Collins et al., 2014; Guidi et al., 2018; Gulsrud et al., 2016; Ward-Horner & Sturmey, 2010). Measurement tools which cut across intervention models are necessary for advancing this goal.

Acknowledgements

The authors thank the families whose participation in research made this work possible. In addition, they thank Kaylin Russell, who dedicated so much of her time to coding for this project. This project was completed with the expert input and fidelity rating forms of numerous experts in ASD intervention (listed in alphabetical order) and their trainees and colleagues: Susan Bryson, Geraldine Dawson, Helen Flanagan, Ann Kaiser, Connie Kasari, So Hyun Kim, Catherine Lord, Rebecca Landa, Mendy Minjarez, Jennifer Nietfeld, Sally Rogers, Stephanie Shire, and Isabel Smith.

Author contributions

K.M.F. contributed to the study conceptualization, measure development, data coding, analysis, interpretation, and manuscript preparation. J.B., G.W.G., A.H., S.R.R., and A.S. contributed to data analysis, interpretation, and manuscript preparation. B.I. contributed to study conceptualization, measure development, analysis, interpretation, and manuscript preparation.

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: B.I. receives royalties from the sale of one of the manuals used in the research. Royalties are donated to the research.

Funding




The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the following grants: Autism Speaks #5773, PI: Hardan; R21DC01368902, PI: Hardan; R01MH081757, PI: Lord; US Department of Education Grant: R324A140005, PI: Stahmer; Autism Speaks Canada ASCFS-2013-13, PI: Brian; Craig Foundation and the Stollery Children's Hospital Foundation, PI: Brian; Congressionally Directed Medical Research Programs W81XWH-10-1-0586, PI: Ingersoll; Health Resources and Services Administration—Maternal and Child Health Bureau R40MC27704, PI: Ingersoll. The content and conclusions are those of the authors and should not be construed as the official position or policy of nor should any endorsements be inferred by HRSA or the US Government.

Ethical approval

This study was approved by the Institutional Review Board (IRB) at Michigan State University and sharing of data was approved by

IRBs at external study sites. All participants consented for their data to be used for research purposes.

ORCID iDs

Kyle M Frost  <https://orcid.org/0000-0001-7941-9485>
 Grace W Gengoux  <https://orcid.org/0000-0002-6942-9513>
 Aubyn Stahmer  <https://orcid.org/0000-0002-1596-9848>

Supplemental material

Supplemental material for this article is available online.

Note

1. For two of the interventions (Project ImPACT and PRT), we obtained fidelity forms from two different research teams.

References

- Barth, R. P., & Liggett-Creel, K. (2014). Common components of parenting programs for children birth to eight years of age involved with child welfare services. *Children and Youth Services Review, 40*, 6–12.
- Bradshaw, J., Koegel, L. K., & Koegel, R. L. (2017). Improving functional language and social motivation with a parent-mediated intervention for toddlers with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*(8), 2443–2458. <https://doi.org/10.1007/s10803-017-3155-8>
- Brian, J. A., Smith, I. M., Zwaigenbaum, L., & Bryson, S. E. (2017). Cross-site randomized control trial of the Social ABCs caregiver-mediated intervention for toddlers with autism spectrum disorder: RCT of caregiver-mediated early ASD intervention. *Autism Research, 10*(10), 1700–1711. <https://doi.org/10.1002/aur.1818>
- Brian, J. A., Smith, I. M., Zwaigenbaum, L., Roberts, W., & Bryson, S. E. (2016). The Social ABCs caregiver-mediated intervention for toddlers with autism spectrum disorder: Feasibility, acceptability, and evidence of promise from a multisite study. *Autism Research, 9*(8), 899–912.
- Brookman-Frazee, L., Stahmer, A., Chlebowski, C., Suhrheinrich, J., & Aarons, G. (2016, December). *Characterizing implementation processes and influences in two community effectiveness trials for autism* [Conference session]. 9th Annual Conference on the Science of Dissemination and Implementation, Washington, DC, United States.
- Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology, 77*(3), 566–579. <https://doi.org/10.1037/a0014565>
- Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2005). Identifying and selecting the common elements of evidence based interventions: A distillation and matching model. *Mental Health Services Research, 7*(1), 5–20.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290.
- Collins, L. M., Trail, J. B., Kugler, K. C., Baker, T. B., Piper, M. E., & Mermelstein, R. J. (2014). Evaluating individual intervention components: Making decisions based on the results of a factorial screening experiment. *Translational Behavioral*

- Medicine*, 4(3), 238–251. <https://doi.org/10.1007/s13142-013-0239-7>
- Embry, D. D., & Biglan, A. (2008). Evidence-based kernels: Fundamental units of behavioral influence. *Clinical Child and Family Psychology Review*, 11(3), 75–113. <https://doi.org/10.1007/s10567-008-0036-x>
- Garland, A. F., Hawley, K. M., Brookman-Frazee, L., & Hurlburt, M. S. (2008). Identifying common elements of evidence-based psychosocial treatments for children's disruptive behavior problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47(5), 505–514. <https://doi.org/10.1097/CHI.0b013e31816765c2>
- Godfrey, E., Chalder, T., Ridsdale, L., Seed, P., & Ogen, J. (2007). Investigating the active ingredients of cognitive behaviour therapy and counselling for patients with chronic fatigue in primary care: Developing a new process measure to assess treatment fidelity and predict outcome. *British Journal of Clinical Psychology*, 46(Pt 3), 253–272. <https://doi.org/10.1348/014466506x147420>
- Guidi, J., Brakemeier, E.-L., Bockting, C. L. H., Cosci, F., Cuijpers, P., Jarrett, R. B., Linden, M., Marks, I., Peretti, C. S., Rafanelli, C., Rief, W., Schneider, S., Schnyder, U., Sensky, T., Tomba, E., Vazquez, C., Vieta, E., Zipfel, S., Wright, J. H., & Fava, G. A. (2018). Methodological recommendations for trials of psychological interventions. *Psychotherapy and Psychosomatics*, 87(5), 276–284. <https://doi.org/10.1159/000490574>
- Gulsrud, A. C., Helleman, G., Shire, S., & Kasari, C. (2016). Isolating active ingredients in a parent-mediated social communication intervention for toddlers with autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 57(5), 606–613. <https://doi.org/10.1111/jcpp.12481>
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34. <https://doi.org/10.20982/tqmp.08.1.p023>
- Hardan, A. Y., Gengoux, G. W., Berquist, K. L., Libove, R. A., Ardel, C. M., Phillips, J., Frazier, T. W., & Minjarez, M. B. (2015). A randomized controlled trial of pivotal response treatment group for parents of children with autism. *Journal of Child Psychology and Psychiatry*, 56(8), 884–892. <https://doi.org/10.1111/jcpp.12354>
- Hogue, A., Dauber, S., Lichvar, E., Bobek, M., & Henderson, C. E. (2015). Validity of therapist self-report ratings of fidelity to evidence-based practices for adolescent behavior problems: Correspondence between therapists and observers. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(2), 229–243. <https://doi.org/10.1007/s10488-014-0548-2>
- Hogue, A., Liddle, H. A., & Rowe, C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy: Theory, Research, Practice, Training*, 33(2), 332–345. <https://doi.org/10.1037/0033-3204.33.2.332>
- Ingersoll, B., & Dvortcsak, A. (2010). *Teaching social communication to children with autism: A practitioner's guide to parent training and a manual for parents*. Guilford Press.
- Ingersoll, B., & Dvortcsak, A. (2019). *Teaching social communication to children with autism and other developmental delays: The project ImPACT guide to coaching parents and the project ImPACT manual for parents* (2nd ed.). Guilford Press.
- Ingersoll, B., Wainer, A. L., Berger, N. I., Pickard, K. E., & Bonter, N. (2016). Comparison of a self-directed and therapist-assisted telehealth parent-mediated intervention for children with ASD: A pilot RCT. *Journal of Autism and Developmental Disorders*, 46(7), 2275–2284. <https://doi.org/10.1007/s10803-016-2755-z>
- Kaehler, L. A., Jacobs, M., & Jones, D. J. (2016). Distilling common history and practice elements to inform dissemination: Hanf-model BPT programs as an example. *Clinical Child and Family Psychology Review*, 19(3), 236–258. <https://doi.org/10.1007/s10567-016-0210-5>
- Kaiser, A. P., Hancock, T. B., & Nietfeld, J. P. (2000). The effects of parent-implemented enhanced milieu teaching on the social communication of children who have autism. *Early Education & Development*, 11(4), 423–446. https://doi.org/10.1207/s15566935eed1104_4
- Kaiser, A. P., & Hester, P. P. (1994). Generalized effects of enhanced milieu teaching. *Journal of Speech, Language, and Hearing Research*, 37(6), 1320–1340.
- Kasari, C., Freeman, S., & Paparella, T. (2006). Joint attention and symbolic play in young children with autism: A randomized controlled intervention study: Joint attention and symbolic play in young children with autism. *Journal of Child Psychology and Psychiatry*, 47(6), 611–620. <https://doi.org/10.1111/j.1469-7610.2005.01567.x>
- Kasari, C., Gulsrud, A. C., Wong, C., Kwon, S., & Locke, J. (2010). Randomized controlled caregiver mediated joint engagement intervention for toddlers with autism. *Journal of Autism and Developmental Disorders*, 40(9), 1045–1056. <https://doi.org/10.1007/s10803-010-0955-5>
- Kasari, C., Lawton, K., Shih, W., Barker, T. V., Landa, R., Lord, C., Orlich, F., King, B., Wetherby, A., & Senturk, D. (2014). Caregiver-mediated intervention for low-resourced preschoolers with autism: An RCT. *Pediatrics*, 134, e72–e79. <https://doi.org/10.1542/peds.2013-3229>
- Landa, R. J., Holman, K. C., O'Neill, A. H., & Stuart, E. A. (2011). Intervention targeting development of socially synchronous engagement in toddlers with autism spectrum disorder: A randomized controlled trial: RCT of social intervention for toddlers with ASD. *Journal of Child Psychology and Psychiatry*, 52(1), 13–21. <https://doi.org/10.1111/j.1469-7610.2010.02288.x>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575.
- Lokker, C., McKibbin, K. A., Colquhoun, H., & Hempel, S. (2015). A scoping review of classification schemes of interventions to promote and integrate evidence into practice in healthcare. *Implementation Science*, 10(1), Article 27. <https://doi.org/10.1186/s13012-015-0220-6>
- McHugh, R. K., Murray, H. W., & Barlow, D. H. (2009). Balancing fidelity and adaptation in the dissemination of empirically-supported treatments: The promise of transdiagnostic interventions. *Behaviour Research and Therapy*, 47(11), 946–953. <https://doi.org/10.1016/j.brat.2009.07.005>
- McKenzie, J. F., Wood, M. L., Kotecki, J. E., Clark, J. K., & Brey, R. A. (1999). Establishing content validity: Using qualitative and quantitative steps. *American Journal of Health Behavior*, 23(4), 311–318.

- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 26*, 315–340.
- Mullen, E. M. (1995). *Mullen scales of early learning*. AGS Circle Pines.
- National Research Council. (2001). *Educating children with autism*. National Academies Press.
- Osterling, K. L., & Austin, M. J. (2008). The dissemination and utilization of research for promoting evidence-based practice. *Journal of Evidence-Based Social Work, 5*(1–2), 295–319.
- Pagoto, S. L., Spring, B., Coups, E. J., Mulvaney, S., Coutu, M., & Ozakinci, G. (2007). Barriers and facilitators of evidence-based practice perceived by behavioral science health professionals. *Journal of Clinical Psychology, 63*(7), 695–705.
- Phaneuf, L., & McIntyre, L. L. (2011). The application of a three-tier model of intervention to parent training. *Journal of Positive Behavior Interventions, 13*(4), 198–207. <https://doi.org/10.1177/1098300711405337>
- Pickard, K. E., Kilgore, A. N., & Ingersoll, B. (2016). Using community partnerships to better understand the barriers to using an evidence-based, parent-mediated intervention for autism spectrum disorder in a medicaid system. *American Journal of Community Psychology, 57*(3–4), 391–403.
- Rogers, S. J., & Dawson, G. (2010). *Early start Denver model for young children with autism: Promoting language, learning, and engagement*. Guilford Press.
- Rogers, S. J., Estes, A., Lord, C., Vismara, L., Winter, J., Fitzpatrick, A., Guo, M., & Dawson, G. (2012). Effects of a brief Early Start Denver Model (ESDM)-based parent intervention on toddlers at risk for autism spectrum disorders: A randomized controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(10), 1052–1065. <https://doi.org/10.1016/j.jaac.2012.08.003>
- Sandbank, M., Bottema-Beutel, K., Crowley, S., Cassidy, M., Dunham, K., Feldman, J. I., Crank, J., Albarran, S. A., Raj, S., Mahbub, P., & Woynaroski, T. G. (2020). Project AIM: Autism intervention meta-analysis for studies of young children. *Psychological Bulletin, 146*(1), 1–29. <https://doi.org/10.1037/bul0000215>
- Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(1), 32–43. <https://doi.org/10.1007/s10488-010-0321-0>
- Schreibman, L., Dawson, G., Stahmer, A. C., Landa, R., Rogers, S. J., McGee, G. G., Kasari, C., Ingersoll, B., Kaiser, A. P., Bruinsma, Y., McNerney, E., Wetherby, A., & Halladay, A. (2015). Naturalistic developmental behavioral interventions: Empirically validated treatments for autism spectrum disorder. *Journal of Autism and Developmental Disorders, 45*(8), 2411–2428. <https://doi.org/10.1007/s10803-015-2407-8>
- Schreibman, L., & Koegel, R. L. (2005). Training for parents of children with autism: Pivotal responses, generalization, and individualization of interventions. In E. D. Hibbs & P. S. Jensen (Eds.), *Psychosocial treatments for child and adolescent disorders: Empirically based strategies for clinical practice* (pp. 605–631). American Psychological Association.
- Stahmer, A. C., Brookman-Frazee, L., Rieth, S. R., Stoner, J. T., Feder, J. D., Searcy, K., & Wang, T. (2017). Parent perceptions of an adapted evidence-based practice for toddlers with autism in a community setting. *Autism, 21*(2), 217–230. <https://doi.org/10.1177/1362361316637580>
- Stahmer, A. C., Rieth, S. R., Dickson, K. S., Feder, J., Burgeson, M., Searcy, K., & Brookman-Frazee, L. (2020). Project ImPACT for Toddlers: Pilot outcomes of a community adaptation of an intervention for autism risk. *Autism, 24*, 617–632. <https://doi.org/10.1177/1362361319878080>
- Suhrheinrich, J., Rieth, S. R., Dickson, K. S., & Stahmer, A. C. (2020). Exploring associations between inner-context factors and implementation outcomes. *Exceptional Children, 86*, 155–173.
- Tate, D. F., Lytle, L. A., Sherwood, N. E., Haire-Joshu, D., Matheson, D., Moore, S. M., Loria, C. M., Pratt, C., Ward, D. S., Belle, S. H., & Michie, S. (2016). Deconstructing interventions: Approaches to studying behavior change techniques across obesity interventions. *Translational Behavioral Medicine, 6*(2), 236–243. <https://doi.org/10.1007/s13142-015-0369-1>
- Tiede, G., & Walton, K. M. (2019). Meta-analysis of naturalistic developmental behavioral interventions for young children with autism spectrum disorder. *Autism, 23*, 2080–2095. <https://doi.org/10.1177/1362361319836371>
- Veneziano, L., & Hooper, J. (1997). A method for quantifying content validity of health-related questionnaires. *American Journal of Health Behavior, 21*(1), 67–70.
- Wainer, A. L., & Ingersoll, B. (2013). Intervention fidelity: An essential component for understanding ASD parent training research and practice. *Clinical Psychology: Science and Practice, 20*(3), 335–357.
- Wainer, A. L., & Ingersoll, B. (2015). Increasing access to an ASD imitation intervention via a telehealth parent training program. *Journal of Autism and Developmental Disorders, 45*(12), 3877–3890.
- Ward-Horner, J., & Sturmey, P. (2010). Component analyses using single-subject experimental designs: A review. *Journal of Applied Behavior Analysis, 43*(4), 685–704. <https://doi.org/10.1901/jaba.2010.43-685>
- Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., Brock, M. E., Plavnick, J. B., Fleury, V. P., & Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disorders, 45*(7), 1951–1966. <https://doi.org/10.1007/s10803-014-2351-z>
- Wood, J. J., McLeod, B. D., Klebanoff, S., & Brookman-Frazee, L. (2015). Toward the implementation of evidence-based interventions for youth with autism spectrum disorders in schools and community agencies. *Behavior Therapy, 46*(1), 83–95.
- Zwaigenbaum, L., Bauman, M. L., Choueiri, R., Kasari, C., Carter, A., Granpeesheh, D., Mailloux, Z., Smith Roley, S., Wagner, S., Fein, D., Pierce, K., Buie, T., Davis, P. A., Newschaffer, C., Robins, D., Wetherby, A., Stone, W. L., Yirmiya, N., Estes, A., . . . Natowicz, M. R. (2015). Early intervention for children with autism spectrum disorder under 3 years of age: Recommendations for practice and research. *Pediatrics, 136*, S60–S81. <https://doi.org/10.1542/peds.2014-3667E>