American Journal on Intellectual and Developmental Disabilities

A Systematic Review of Maintenance in Early Childhood Autism Spectrum Disorder Research

--Manuscript Draft--

Manuscript Number:	AJIDD-D-21-00016R2
Article Type:	Comprehensive Review
Keywords:	Autism Spectrum Disorder; Early Intervention; maintenance
Corresponding Author:	Beth A Pokorski, PhD, BCBA-D NA Asheville, NC UNITED STATES
First Author:	Beth A Pokorski, PhD, BCBA-D
Order of Authors:	Beth A Pokorski, PhD, BCBA-D
	Lauren M Lejeune, PhD, BCBA
Manuscript Region of Origin:	UNITED STATES
Abstract:	Single case research designs (SCRDs) are integral to identifying evidence-based practices (EBPs) for young children with autism spectrum disorder (ASD); however, the field lacks guidance on measuring response maintenance within SCRDs. We identified 103 studies in which researchers used SCRD to investigate the maintenance of behavioral intervention outcomes for children with ASD ages 0-5. Findings include: (a) maintenance conditions across most EBP categories; (b) limited within-case replication of maintenance assessment; (c) inconsistent use of maintenance terminology; (d) varying frequencies of maintenance assessment; and (e) wide range in latency to first and last maintenance probe. Results indicate a pressing need for the regular inclusion of maintenance conditions in behavioral research to increase our understanding of programming for and assessing maintenance.

Abstract

Single case research designs (SCRDs) are integral to identifying evidence-based practices (EBPs) for young children with autism spectrum disorder (ASD); however, the field lacks guidance on measuring response maintenance within SCRDs. We identified 103 studies in which researchers used SCRD to investigate the maintenance of behavioral intervention outcomes for children with ASD ages 0-5. Findings include: (a) maintenance conditions across most EBP categories; (b) limited within-case replication of maintenance assessment; (c) inconsistent use of maintenance terminology; (d) varying frequencies of maintenance assessment; and (e) wide range in latency to first and last maintenance probe. Results indicate a pressing need for the regular inclusion of maintenance conditions in behavioral research to increase our understanding of programming for and assessing maintenance.

Keywords: autism spectrum disorder; early intervention; maintenance

A Systematic Review of Maintenance Measurement in Early Childhood Autism Spectrum Disorder Research

Autism spectrum disorder (ASD) is a developmental disability characterized by deficits in social communication and interaction and the presence of restricted, repetitive behaviors (American Psychiatric Association, 2013). Researchers in psychology, special education, applied behavior analysis (ABA), and related fields have established a large body of research focused on ameliorating these deficits and related maladaptive behaviors (e.g., challenging behavior; Steinbrenner et al., 2020). Much of this research has focused on early behavioral intervention for children with ASD from birth through 5 years old (Steinbrenner et al., 2020), traditionally evaluated using single case research design (SCRD; Barrios & Hartmann, 1988; Steinbrenner et al., 2020). SCRDs are experimental designs that have contributed to the identification of numerous evidence-based practices (EBPs) for increasing desired behaviors and decreasing challenging behavior in young children with ASD (Wong et al., 2015). However, analyses of maintenance data have typically been omitted from comprehensive EBP reviews (e.g., Wong et al., 2015; Steinbrenner et al., 2020), perhaps because less is known about using SCRD to analyze whether intervention effects maintain across time (Barrios & Hartmann, 1998; Ledford et al., 2018). This is important, given the substantial evidence indicating individuals with ASD may have difficulty maintaining skills without explicit programming (Foxx, 2008).

Leaders in the fields of ABA and special education have historically emphasized the importance of explicitly programming for response maintenance (Stokes & Baer, 1977). Maintenance of behavior change is considered an integral component of socially-valid ABA practice (Cooper et al., 2020; Kennedy, 2002), and some researchers have suggested interventions are not truly effective unless the effects endure across time (Baer et al., 1968;

Stokes & Baer, 1977). In their seminal article on the topic, Stokes and Baer (1977) described seven strategies for increasing the likelihood of response maintenance (e.g., introduce behavior to naturally maintaining contingencies, train sufficient exemplars). Importantly, Stokes and Baer cautioned against designing interventions without incorporating maintenance-supporting strategies (i.e., "train and hope"). Nonetheless, more than 40 years after they urged researchers to increase the frequency of maintenance programming and analysis, the inclusion of maintenance phases within SCRD remains rare (Ledford et al., 2018; Neely et al., 2016; Neely et al., 2018). Increased attention to this issue could extend our understanding of EBPs to include not only practices that improve outcomes, but also those that result in lasting change. This might be especially important when considering interventions targeting young children with ASD, for whom early and effective intervention may significantly improve long-term developmental outcomes (Lang et al., 2016).

Maintenance Terminology and Procedures

In addition to the infrequency of maintenance phases in SCRD, authors often use inconsistent terminology related to maintenance (noted by Cooper et al., 2020; Stokes & Baer, 1977). Stokes and Baer (1977) defined *generalization* as the occurrence of behavior under non-training conditions, including across time, and Baer et al. (1968) described *generality of behavior* as change that endures across time. More recently, Cooper et al. (2020) clarified that *response maintenance* (i.e., behavior enduring across time) can be considered a specific type of generalized behavior change and recommended that authors differentiate between the phenomenon of response maintenance (a dependent variable) and the maintenance condition in a SCRD study (an independent variable). Further, authors regularly use the term "follow-up" to describe SCRD study conditions in which researchers measure response maintenance (Barrios &

Hartmann, 1988), and authors might even interchange these terms throughout a publication. Given the absence of definitive terms and descriptions for these conditions, throughout this publication we have opted to follow the definitions for maintenance and stimulus generalization provided by Kazdin (2012). Specifically, we consider maintenance the "continuation of the behavior over time after the program has been terminated" (p. 426) and generalization the "extension of behavior changes to new situations, settings, and circumstances" (p. 426).

Beyond inconsistent terminology, there are often procedural inconsistencies between maintenance conditions. Some authors consider maintenance a condition in which intervention is no longer implemented (Kennedy, 2002; Ledford et al., 2018; Neely et al., 2018); some include conditions in which intervention is partially implemented (Cooper et al., 2020; Stokes & Baer, 1977); and others use maintenance to describe a specific condition in which the reinforcement schedule is changed from fixed to variable (i.e., intermittent reinforcement; Miltenberger, 2016) or thinned. There also may be important differences regarding for whom response maintenance is measured. For example, researchers may plan a maintenance condition to measure: (a) the primary participant's response maintenance over time or given changes to intervention; or (b) a natural implementer's response maintenance in the absence of coaching and frequent observations (i.e., fidelity of implementation over time). There is also a lack of guidance for when maintenance should occur relative to intervention and for how long it should continue (Neely et al., 2018). For example, Kennedy (2002) suggested that an intervention might be considered effective if desirable levels of behavior maintain for six months, but there is not yet validation of this suggestion. These inconsistencies in recommended terminology and procedures indicate a need for a systematic review of maintenance measurement practices in SCRD studies.

Maintenance Measurement within SCRD

The measurement of response maintenance in SCRD is inherently challenging, predominantly due to the difficulty of controlling for threats to internal validity (Barrios & Hartmann, 1988). Researchers who conduct SCRD studies can rule out common threats to internal validity (e.g., history, maturation, testing) through procedures such as staggering baseline lengths, collecting interobserver agreement (IOA) and procedural fidelity data, and demonstrating the effect of the independent variable at three different points in time (i.e., replication; Kratochwill et al., 2013). For example, in the withdrawal or ABAB design, researchers demonstrate experimental control over behavior if the behavior repeatedly improves when the intervention is implemented and worsens when the intervention is removed. The strength of the ABAB design is dependent on the presence of *reversible behaviors*—those behaviors for which improvements do not typically persist in the absence of intervention (e.g., challenging behavior, engagement; Ledford et al., 2018). Thus, implementing a maintenance condition to assess response maintenance in the absence of intervention is counter to this experimental logic. This paradox demonstrates one of the many challenges to maintaining internal validity during a maintenance phase. As Barrios and Hartmann asked, how do you show the treatment is responsible for changes when treatment is not being implemented?

Despite these design challenges, some authors have suggested ways to measure response maintenance in SCRD. Researchers could repeatedly introduce and withdraw a maintenance condition (Barrios & Hartman, 1988; Rusch & Kazdin, 1981) or they could include a single maintenance condition following intervention. The first option could demonstrate the effect of maintenance procedures at three different points in time (i.e., experimental control), whereas the second option would provide a single demonstration of maintenance. As an example of a design with experimental control, Ledford et al. (2018) described a multiple baseline strategy in which

researchers implement an intervention across three (or more) tiers in a time-lagged fashion, and then subsequently implement a time-lagged maintenance condition. This procedure would facilitate experimental analysis of the intervention-to-maintenance comparison. Barrios and Hartmann (1988) described an additional procedure in which researchers (a) validate an intervention in a simultaneous treatments design, then (b) replace the two interventions with two maintenance procedures (e.g., thinning vs. withdrawal) to compare outcomes between acquisition and maintenance conditions across procedures.

If a single maintenance condition is utilized, Barrios & Hartmann (1988) recommended researchers measure the behavior repeatedly, as "the single follow-up probe may serve no useful purpose" (pg. 31). Given the lack of internal validity with the application of a single condition (i.e., no experimental control is demonstrated), researchers using this method cannot conclude the treatment implemented during intervention is responsible for behavior maintenance over time. However, Barrios & Hartmann noted the clinical value of maintenance probes: Consumers can observe whether behavior continues at a desirable level, or (perhaps more importantly) whether behavior *worsens*. Such a pattern could indicate a need to reintroduce intervention (Barrios & Hartmann, 1988).

Previous Reviews of Maintenance Measurement

As previously stated, comprehensive reviews of EBPs for children with ASD have not systematically evaluated maintenance conditions of included studies (e.g., Wong et al., 2015; Steinbrenner et al., 2020). However, researchers have evaluated maintenance within reviews of specific interventions. For example, Schlosser and Lee (2000) systematically reviewed the SCRD research on augmentative and alternative communication (AAC) for individuals with disabilities. They identified 232 data series (e.g., individual participants) across 50 studies, 42 (18%) of which included maintenance conditions. The authors reported that most of the 42 series

assessed maintenance via multiple follow-up probes (85.7%) and that most comparisons involved "train and hope" approaches (91.2%); however, these results are limited by a lack of clear inclusion and coding criteria related to maintenance phases.

More recently, Neely et al. (2018) assessed generalization and maintenance of functional communication training (FCT) for individuals with developmental disabilities. They identified 19 SCRD studies in which researchers explicitly labeled and included maintenance data following intervention, 16 of which (84%) included multiple maintenance probes. Maintenance conditions occurred immediately following intervention (58%), 1-3 months later (11%), more than 3 months later (26%), and up to 2 years later (1 study). They reported 6 studies (31.6%) were methodologically rigorous (e.g., at least 3 data points per condition). Similarly, Hong and colleagues (2018) reviewed tablet-based SCRD intervention studies for individuals with ASD, focusing on the extent of generalization and maintenance measurement. They identified 39 studies meeting inclusion criteria, 16 (41.0%) of which assessed maintenance (defined as detection of a behavioral change "after completion of a targeted intervention" [p. 133]). Of these, 8 (50%) reported latency to maintenance, which ranged from 1.5 to 10 weeks post-intervention, and 2 (12.5%) met design standards (i.e., 3 or more data points, IOA collected for ≥20% of sessions). Although we located one systematic review specific to preschool-aged children with ASD (Gunning et al. 2019), authors reported generalization and maintenance data as a unit, precluding direct analysis of this condition.

Importantly, each review noted minimal methodological rigor of maintenance conditions (e.g., number of probes, collection of IOA, assessment of procedural fidelity), which reduces confident conclusions about response maintenance within these bodies of research. These results

point to the need for additional research on maintenance measurement for behavioral interventions, particularly with young children with ASD.

Purpose of This Review

Response maintenance is a crucial measure of intervention effectiveness (Baer et al., 1987; Cooper et al., 2020; Kennedy, 2002; Stokes & Baer, 1997) and should be evaluated when identifying EBPs. Although researchers have offered suggestions for both experimentally and descriptively evaluating response maintenance in SCRD, the inclusion of maintenance conditions remains rare (Ledford et. 2018). Furthermore, inconsistencies in terminology and procedural guidelines may prevent research consumers from identifying, evaluating, and replicating maintenance procedures. Although authors of previous systematic reviews have considered aspects of response maintenance within larger topical reviews, none have specifically focused on maintenance assessment or a target population. Our research questions were as follows: When considering the assessment of maintenance in behavioral research conducted with young children with ASD:

1. What patterns exist across interventions, behaviors, settings, and designs?
2. What are the features of maintenance conditions, including latency, length, intervention components, and setting?

Method

Search and Screen

The first author conducted electronic database searches via the EBSCSO*host* web platform across PsycINFO, Education Resources Information Center (ERIC), and Education Full Text databases on September 30th, 2019. We full-text queried terms related to autism (autism OR autistic OR ASD OR PDD* OR asperger* OR “childhood disintegrative disorder” OR “rett

syndrome" OR "heller's syndrome") AND early childhood ("pre-k*" OR prek* OR preschool* OR daycare OR "day care" OR "nursery school" OR "early child*" OR "young child*" OR toddler) AND intervention (interven* OR treat* OR train* OR instruct*OR program* OR teach*) AND maintenance (maintenance OR maintain* OR "follow-up" OR "follow up"). Identified studies (N = 7,894) were imported into reference management software and duplicates were removed, resulting in 7,070 studies for review (see PRISMA flow diagram; Figure 1).

Inclusion Criteria

We required each study to meet the following inclusion criteria. First, studies were required to be published in English and in a peer-reviewed academic journal. Second, we included only studies in which one or more participant was aged 0 through 5 years old and reported to have an ASD, including conditions listed in previous diagnostic criteria (e.g., PDD-NOS). We excluded studies in which children were suspected of having an ASD or reported to have "ASD-like tendencies," for example. Next, studies were required to implement an intervention with one or more qualifying participant and to analyze the effects via an experimental SCRD meeting basic What Works Clearinghouse™ (WWC) design standards (Kratochwill et al., 2013). Specifically, studies were required to (a) systematically manipulate an independent variable, (b) collect data via two or more assessors for at least 20% of sessions, demonstrating an average IOA of at least 80%, (c) include three or more attempts to demonstrate an effect over time, and (d) include a minimum of three data points—or five, for alternating treatment designs (ATD)—per experimental condition (i.e., baseline and intervention conditions but not maintenance). Any studies in which researchers used a non-experimental design (e.g., nonconcurrent multiple baseline; Gast et al., 2018) or graphed data unconventionally, precluding traditional visual analysis (e.g., cumulative ATD), were eliminated.

Finally, we required studies to meet two inclusion criteria related to maintenance conditions. Given our goal was to report the methods used by researchers to measure response maintenance following treatment validation (per Ledford et al., 2018 and Barrios & Hartman, 1988 recommendations), we included only studies that measured the level of a child dependent variable following the conclusion of primary experimental phases. We determined treatment validation was concluded if a separate phase or time-lapsed data points were present and labeled as *maintenance* or *follow-up* in the figure or if authors used either term to describe the phase/data points in the narrative. We selected these specific terms because they are used regularly (and often interchangeably) in SCRD literature to designate post-intervention assessment (Barrios & Harmann, 1988). We excluded: (a) studies in which maintenance assessment was embedded within the design, and thus occurred during rather than following treatment validation (e.g., the majority of multiple probe [conditions variation] designs; Gast et al., 2018); (b) studies in which final data points were referred to using any terminology other than maintenance or follow-up (e.g., treatment extension, sustained use, treatment thinning); and (c) studies with indirect measures of maintenance (e.g., rating scale, parent report). We also required maintenance to be assessed with a qualifying child participant and excluded studies if maintenance was measured only for participants who did not meet inclusion criteria (e.g., implementer, child without ASD). For example, if a multiple baseline across participants design included children aged 3 to 8 years old but maintenance was not measured for any age-appropriate participants, it was excluded.

Screening

Screening occurred in two stages: (1) title and abstract screening and (2) full-text screening. At each stage, the authors created, piloted, and refined a screening tool, then independently coded practice articles until achieving an average inter-rater reliability (IRR) of

≥90% per variable. During the first stage, the first author screened the titles and abstracts of all 7,070 studies identified in database searches for any clear conflicts with inclusion criteria (e.g., adult participants, randomized control trial, literature reviews). Twenty percent (*n* = 1,437) were independently screened by a trained graduate student in special education (IRR = 94%). This resulted in 576 studies continuing to full-text screening. For the second stage, the first author screened the full text of eligible studies for the presence of all inclusion criteria, removing any in which one or more criterion was not evident (see Figure 1 for number of articles removed by category). The second author independently screened 20% of studies (*n* = 116; IRR = 92.3%). Reliability assessment occurred at regular intervals across the article set, with disagreements resolved by consensus during all stages of screening and coding. In the event a screening tool was modified, all studies screened thus far were re-evaluated for inclusion. The screening process resulted in the final inclusion of 103 studies.

Coding

The first and second author coded all studies using a codebook developed by the research team. We first co-coded three studies that were not selected for IRR assessment, then independently coded additional non-IRR studies until we reached an average agreement of ≥90% for all variables across the set of practice studies (*n* = 3 studies). Finally, we independently coded remaining studies, overlapping on 21% (*n* = 21 studies). As with the screening process, reliability assessment was distributed throughout coding with disagreements resolved through discussion until consensus was reached. Mean IRR across variables was 91.3% (*range* = 78.4%–100%); any variables with IRR of <80% were removed from data analysis.

The codebook was composed of three primary sections: (1) general study characteristics, (2) intervention and implementation characteristics, and (3) maintenance characteristics. The

majority of variables were coded at the study level. However, we coded three variables (i.e., *n* maintenance sessions and latency to first and last maintenance sessions) at the participant level, as these often differed across participants. In total, 35 distinct variables were coded.

General study characteristics included publication journal, country in which study occurred, *n* potential participants and *n* included participants, setting, implementer type (e.g., researcher, teacher), and whether a mastery criterion was set. In addition, we coded experimental design, general presence of procedural fidelity data (i.e., not specific to maintenance sessions), assessment of generalization (i.e., within different situations, settings, or circumstances than the intervention context), and presence of social validity data.

Intervention characteristics included the name of the intervention and which, if any, of Steinbrenner and colleagues' (2020) 28 EBP categories it was most aligned with (i.e., the most active component). If authors specifically reported using a behavior intervention package, we coded the intervention as such, rather than attempting to identify which component appeared most active. In studies that (a) compared two or more interventions or (b) provided participants different interventions, we recorded which (if any) EBP category *each* intervention represented. In addition, for each dependent variable we reported whether the goal was behavior acquisition or reduction and categorized according to the 13 behavior domains (e.g., cognitive, play) used by Steinbrenner and colleagues.

For maintenance, we coded the context in which it was assessed (e.g., identical to intervention context, generalization context), whether intervention was implemented in full or in-part during maintenance conditions, and the method of partial implementation, if applicable. Methods of partial implementation were coded as: (a) faded (i.e., gradual transfer of stimuli from one controlling variable to another, typically natural, stimuli); (b) thinned (i.e., gradual reduction

of reinforcement); (c) nonsystematic (i.e., natural implementer not given instructions on whether to implement intervention or told to interact as they "normally would"); or (d) "other" (i.e., idiosyncratic methods of partial implementation, such as implementing portions of the intervention during maintenance). We counted the number of maintenance sessions per participant and calculated a mean for participants whose performance was assessed across multiple behaviors or tiers. Finally, we coded how long post-intervention the first and last maintenance session occurred (i.e., latency to maintenance). We coded the number of days, weeks, or months when explicitly provided by the authors (e.g., data were collected at 2, 4, and 6 weeks). We coded the latency as "Unclear" if authors gave a general timeline (e.g., sessions occurred 3 months later) and "Not Reported" (NR) if authors did not indicate a timeline. Given authors reported this information with varying degrees of clarity (resulting in coding difficulties), we double-coded 100% of studies for these two variables, resulting in IRR of 83.8% and 78.8%, respectively. To facilitate data analysis, we then grouped data into eight latency categories as follows: < 1 wk; 1 to < 2 wk; 2 to < 4 wk; 1 to < 2 months; 2+ months; NR; Unclear.

Results

We located a total of 103 studies that met inclusion criteria. Studies were published between 1986 and 2019, with the majority published between 2000-2009 (26.2%) or 2010-2019 (68%). Studies were published in 30 peer-reviewed journals, with 31.1% published in two journals (*Journal of Autism and Developmental Disorders*, $n = 17$; *Journal of Applied Behavior Analysis*, $n = 1$; Table 1). References of included studies are available from the authors upon request.

General Study Characteristics

Of the 334 participants included in studies, 78.1% (*n* = 261) were aged 0-5 with ASD, and 66.2% (*n* = 221) met all inclusion criteria (i.e., one or more behavior measured via an experimental design with a maintenance condition)(Table 2). Studies occurred primarily in the United States (84.5%), with interventions most commonly implemented in a school/daycare setting (44.6%) or a clinic setting (29.1%). Research staff (48.5%) were the most common implementers, with naturally-occurring persons (e.g., parents, teachers) serving as implementers in 34% of studies, and both researchers and naturally-occurring persons serving as implementers in 6.8% of studies.

As seen in Table 3, the majority of experimental designs were multiple baseline (51.5%) and multiple probe (19.4%), most often across participants (56.2% of multiple baseline/probe studies). Most studies (91.7%) targeted acquisition behaviors, including social behavior (22.2%), communication targets (18%), pre-academic/academic behavior (16%), and joint attention (13.2%). No studies intervened on motor targets or vocational skills. Over half of studies (58.3%) required participants meet a mastery criterion prior to progressing to the maintenance condition, 41.8% reported generalization data, and 46.6% assessed the social validity of the treatment. Regarding study quality, procedural fidelity was reported in 69.9% of studies

Intervention Characteristics

Twenty-three of the 28 EBPs identified by Steinbrenner and colleagues (2020) were present in our review (Table 4). Those represented most frequently included Discrete Trial Training (DTT), Video Modeling, AAC, Naturalistic Interventions, and Prompting, each of which comprised 11.5%–12.5% of studies. Categories that were not represented included Cognitive/Behavioral Instructional Strategies, Direct Instruction, Self-Management, Sensory Integration, and Task Analysis.

Maintenance Characteristics

As seen in Table 5, included studies were divided on how they referred to the maintenance condition: (a) maintenance (42.7%), (b) follow-up (34%), or (c) both terms across figures and text (23.3%). Most assessed maintenance within the same context as intervention (e.g., setting, implementer, materials; 71.8% of studies), with few doing so in a generalization context only (7.8%). In 61.2% of studies, intervention was not implemented during the maintenance condition, such that it was identical to the baseline/control condition. However, 23.3% of authors reported implementing some aspects of intervention in the maintenance condition and 10.7% did not clearly indicate if intervention was implemented.

The average number of sessions across participants and studies was 3.8 (*range* = 1–37 sessions; *mode* = 2), with 61.5% of participants having three or more maintenance probes. As depicted in Figure 2, the latency to the first maintenance session ranged from 1 day to 6 months, with the largest percentage categorized as 1 month to less than 2 months (27.9%). Similarly, the latency to the last maintenance session ranged from 1 day to over 7 years (Lucyshyn et al. 2007), with the highest percentage occurring from 2 weeks up to 4 weeks (23.7%). A portion of authors reported unclear latency information for the first or last maintenance session (10.4% and 6.2%, respectively) or did not report latencies (13.8% for first session; 11.8% for last session). Note that the latency to last session was not applicable (coded as NA) for studies that assessed maintenance within a single session (18.01%).

Discussion

The purpose of this review was to systematically analyze maintenance procedures in behavioral intervention research for children with ASD aged 0 to 5 years old. To our knowledge, this was the first systematic review to focus on maintenance procedures specifically, and the first

to examine maintenance for a specific population rather than for an intervention. We identified 103 studies in which researchers experimentally evaluated the effect of behavioral interventions on child-level outcomes and included a condition described as maintenance or follow-up. Several patterns emerged in our results related to characteristics of research procedures, interventions, and maintenance conditions.

Perhaps most notably, our results highlight the dearth of high-quality SCRD studies that have assessed response maintenance for young children with ASD. During full-text screening, the presence of a maintenance condition was our final criterion. Of the 202 articles that met all other criteria (e.g., participants 0-5 years old with ASD; experimental SCRD meeting WWC standards) only 103 (51.0%) included a maintenance condition for one or more participant. In other words, only half of the relevant research clearly measured response maintenance. These results align closely with those of Neely et al. (2018), who found only half of their study set included a maintenance condition (51.4%; n = 19 studies). However, we were encouraged to find that most of the studies included in this review were published in the last decade (68%), which suggests the frequency of maintenance assessment in SCRD may be increasing.

Independent and Dependent Variables

Of the 23 practices designated as EBPs by Steinbrenner and colleagues (2020), the majority (n = 16; 69.6%) were represented three or more times in our study set, 7 (30.4%) were represented fewer than three times, and 5 (21.7%) were not represented (Table 4). This suggests a disconnect between the current evidence supporting some EBPs and ABA standards stating interventions are not truly effective unless their effects endure (Baer et al., 1968; Stokes & Baer, 1977). However, readers should note that not all studies included in Steinbrenner and colleagues' review met our inclusion criteria (e.g., group designs, older participants). It is probable that

additional studies in their report assessed maintenance, or that studies measured response maintenance but used terminology we excluded (e.g., treatment extension, sequential withdrawal). In contrast, 5 EBPs were represented 12 or more times in this review, indicating a more robust knowledge of how their effects maintain over time. Future researchers might consider if characteristics of specific EBPs (e.g., intervention setting, target behavior) lend themselves to maintenance assessment more readily than characteristics of lesser-represented EBPs.

Regarding dependent variables, the large majority of outcomes targeted were skill acquisition (e.g., communication), whereas only 12 of the 144 outcomes (8.33%) were behaviors targeted for reduction (e.g., challenging behavior). Although this may be an artifact of the predominance of SCRD studies targeting behavior acquisition (72.2%) versus behavior reduction (27.8%), as identified by Shadish and Sullivan, 2011, it remains notable. Given the presence of maladaptive behavior in a learner's repertoire can restrict them from certain environments or prevent them from acquiring adaptive skills (Favell & Lovaas, 1987), it is imperative that interventions focused on behavior reduction produce results that maintain for a socially-significant period of time (Cooper et al., 2020). The limited research assessing maintenance of behavior reduction interventions indicates a critical need in the field.

Methodological Features

Commonalities among methodological procedures were evident within this review. Most researchers applied multiple baseline or multiple probe designs (73 of 103 studies; 70.9%). This is encouraging, as the measurement of response maintenance via multiple baseline designs (as described by Ledford et al., 2018) is an ideal choice for experimentally assessing maintenance of irreversible behaviors. However, only 12 studies in our set (11.7%) used a multiple baseline or

probe design with time-lagged maintenance conditions containing three or more data points per tier. Further, although more than half of studies in this review (61.5%) included three or more sessions in their maintenance condition, most did not replicate effects as necessary to demonstrate a functional relation (Kratochwill et al., 2013). Hence, despite the fact that many studies in our review (and in Neely et al., 2018; Schlosser et al., 2000) followed the recommendation of Barrios and Hartmann (1988) regarding the adequate number of maintenance probes, few permitted conclusive evidence of maintenance over time.

Approximately 30% (*n* = 31) of included studies did not report procedural fidelity data, which is an integral component of high-quality SCRD research (Ledford et al., 2018). Approximately 50% (*n* = 55) of studies did not directly assess social validity, although we acknowledge maintenance assessment might be considered a natural indicator of social validity (Kennedy, 2002). There remains a pressing need for researchers to demonstrate methodologically rigorous and socially-valid procedures within SCRD research. Providing this evidence will foster greater confidence in intervention outcomes, and consequently maintenance outcomes, resulting in more meaningful implications for future practice.

Sixty studies in our set (58.3%) reported training to a mastery criterion, which is notable given a study by Fuller and Fineup (2018) provided preliminary evidence relating training to mastery and response maintenance. The results of their small-*n* study indicated that behavior maintained at a level slightly below the mastery criterion, such that training to a higher criterion (e.g., 90%) might result in more robust maintenance over time. This suggests: (a) interventions in which a rigorous mastery criterion is not met prior to moving to maintenance trials may not result in sustained effects; and thus (b) researchers might consider requiring participants to meet a rigorous mastery criterion prior to terminating the intervention condition.

Fewer than half of included studies (*n* = 43; 41.8%) reported generalization data, which we found surprising given the relationship between generalization and maintenance in the field of ABA. Although maintenance can be considered generalization across time (Stokes & Baer, 1977), and in some cases the two conditions are indistinguishable, generalization also occurs across a number of additional dimensions (e.g., setting, stimuli). It is thus notable that the majority of authors in our review only assessed generalization across the dimension of time. In contrast, 7.8% of authors assessed maintenance in a generalization setting only, effectively obviating any distinction between the two conditions. This lack of conditional clarity was noted as a challenge by Schlosser and Lee (2000) as well, perhaps indicating further consideration regarding the distinction between generalization and maintenance and how best to measure each within SCRD research.

Characteristics of Maintenance Conditions

Our results indicated inconsistencies in the use of maintenance terminology and procedures, as discussed by Cooper and colleagues (2020). Authors referred to conditions that used similar procedures as maintenance, follow-up, and sometimes both across figures and text (24.2% of studies). This use of multiple terms is problematic when related to a common understanding of each term. For example, might follow-up refer to a condition in which no intervention is provided, whereas maintenance might indicate continued, but reduced, support (e.g., thinning, prompt fading)? Procedurally, most authors reported fully withdrawing intervention during maintenance conditions (61.2%), yet intervention was fully implemented in others (4.9%). Some researchers partially withdrew intervention (23.3% of studies), often in ways that were nonsystematic (e.g., teacher's choice to implement; 9 studies) or idiosyncratic (i.e., coded as "Other;" 9 studies). Various maintenance procedures may be appropriate when

considering the behavior, reinforcement schedule, instructional strategy, and context. Nonetheless, clearer operational definitions may help: (a) consumers distinguish between procedural distinctions, and (b) the field identify distinctions in outcomes following the use of various procedures.

Another noteworthy finding of our review was the wide variability in both the number and latency of maintenance probes (Table 5 & Figure 2). Although this study is the first (of which we aware) to measure the latency to both the first *and* last maintenance probe, other researchers (e.g., Hong et al., 2018; Neely et al., 2018) reported a similar range of latencies to initial maintenance probes. In addition, there was no clear association between the number of data points and the latency of probes. For example, one author collected 16 maintenance probes across 3 weeks (Noell et al., 2000) and another collected 5 probes over 7 years (Lucyshyn et al. 2007). Likewise, we found no patterns between the number or latency of probes and target behaviors. However, our data are confounded by the ambiguous descriptions of maintenance provided by many authors. Some authors precisely described the timing of maintenance probes (e.g., "We assessed Brandon's maintenance for skills taught with the teaching interaction procedure 56, 106, and 107 days after he reached mastery criterion;" Kassardjian et al., 2014, p. 2337), whereas others were vague. This pattern was also noted by Hong and colleagues (2018), who reported only 50% of studies provided codable data on latency to maintenance. We encourage future researchers are to describe the exact period of time, in relation to intervention, during which maintenance is measured. This could aid in clarifying current maintenance assessment practices, which in turn might result in recommendations for the field.

Limitations

There were several limitations to this review. Regarding the search and screen process, we did not conduct ancestral or hand searches. This might have resulted in the omission of literature meeting inclusion criteria, potentially impacting results. However, as the purpose of this review was to summarize general characteristics related to maintenance of interventions for young children with ASD, we did not intend to conduct an exhaustive review. In addition, we did not double-screen all studies for inclusion during either stage of screening, as is considered best practice (Polanin et al., 2019), but instead opted for a single screener with frequent IRR checks by a second screener. Although this decision may have resulted in the exclusion of relevant literature, we hypothesize it resulted in a minimal reduction of studies, an outcome considered more probable when screening is conducted by an experienced reviewer, such as the first author (Waffenschmidt et al., 2019).

We also identified limitations related to our maintenance inclusion criteria and coding variables. First, we did not require maintenance sessions to be conducted within the same context (e.g., location, implementer, stimuli) as intervention; thus, some studies analyzed effects of contextual variable alteration in addition to time (i.e., generalization; Stokes & Baer, 1977). Second, we required authors to specifically label conditions as maintenance or follow-up, which may have eliminated studies that included a maintenance condition but referred to it using an alternate term (e.g., return to baseline, post-treatment). Third, we did not evaluate if researchers programmed for maintenance (and if so, how). Although we initially coded variables related to maintenance programming, we did not include them in our analysis given low IRR for both (73% and 76.3%, respectively). Gunning et al. (2019) reported a similar level of reliability for this variable (75%), while other reviews did not specifically report IRR for this variable (Hong et al., 2018; Neely et al., 2018; Schlosser & Lee, 2000). This indicates a need for greater clarity in

reporting study procedures used to foster response maintenance. Fourth, we did not code effects of experimental or maintenance conditions, precluding the analysis of whether outcomes maintained across time. Given that most studies used non-experimental follow-up probes, such an analysis would have been greatly limited. Follow up analyses and exploratory studies that specifically address these issues will be essential in targeting this research gap.

Finally, although we applied Steinbrenner and colleagues' (2020) intervention and behavior categories to our studies, we adapted some coding procedures. The authors reported they: (a) "frequently coded [studies] into multiple intervention categories when multiple EBPs were present" (p. 30), and (b) "reclassified [behavior intervention packages] into each of the individual EBPs" (p. 30). Given these descriptions were not replicable, we used Wong et al. (2014) guidelines by coding studies into only one intervention category and coding "Behavior Intervention Package" for studies with multi-component interventions. Consequently, the results of this review may not be directly comparable to other reviews that apply this framework.

Recommendations for Future Research

The results of this review point to multiple recommendations for researchers and the fields of special education and ABA on a whole. First, we recommend that researchers who conduct intervention studies collect maintenance data whenever possible. Although assessing maintenance experimentally is preferable, we believe there is value in both experimental and descriptive maintenance data. Either application should include three or more maintenance probes to detect patterns in responding. We agree with Schlosser and Lee (2000), who note that follow-up probes conducted in novel settings can lead to confusion regarding what phenomenon is begin assessed. Thus, we recommend researchers assess generalization continuously

throughout intervention (Ledford et al., 2018) and reserve post-intervention probes for the assessment of maintenance.

We urge authors to use consistent and clear terminology when describing maintenance conditions and outcomes. We suggest that the terms maintenance and follow-up should be reserved for conditions in which intervention is withdrawn. For post-treatment conditions during which intervention is programmed or might be implemented (e.g., if natural implementers can choose), we recommend researchers use technological terms such as "sustained use," "treatment extension," or "partial withdrawal." Regardless of the terminology used, researchers should clearly describe whether any aspects of intervention were present and whether the intervention included specific procedures intended to support maintenance. In addition, researchers should precisely describe maintenance conditions, including the length of time between: (a) intervention and maintenance phases, (b) sessions within maintenance phases, and (c) the first and last maintenance sessions.

The results of this review also demonstrate a clear need for additional guidance on variables related to maintenance. Specifically, our field would benefit from improved guidelines related to how and how often to assess response maintenance. Determining these recommendations will require additional research analyzing these variables across a range of EBPs. In addition, future standards for high-quality research (e.g., WWC Standards) might include the measurement of maintenance as a quality indictor for intervention studies, as they typically do IOA and procedural fidelity. Further, journal editors and reviewers could highlight the importance of maintenance conditions when drafting submission guidelines and reviewing manuscripts. These changes have the potential to increase maintenance assessment, which in turn might result in a more robust understanding of the long-term effects of behavioral interventions.

Conclusion

Response maintenance is an integral aspect of effective behavioral intervention. However, our findings demonstrate infrequent, inconsistent, and unclear reporting of procedures for assessing response maintenance in behavioral research, indicating a need for improvement in this area. Despite ample research on EBPs for young children with ASD, the long-term effects of these interventions remain unclear. Continued efforts are needed to improve our understanding of how to promote and assess response maintenance to better support those we treat.

References

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. doi: 10.1901/jaba.1968.1-91

Barrios, B. A., & Hartmann, D. P. (1988). Recent developments in single-subject methodology: Methods for analyzing generalization, maintenance, and multicomponent treatments. *Progress in Behavior Modification*, *22*, 11-47.

Cooper, J. O, Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (3rd ed.). Pearson Education, Inc.

Gast, D. L., Lloyd, B. P., & Ledford, J. R. (2018). Multiple baseline and multiple probe designs. In Ledford J. R., & Gast D. L. (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (3rd ed., pp. 239-281). Routledge. doi: 10.4324/9781315150666

Favell, J. E., & Lovaas, O. I. (1987). Protection for clients undergoing aversive/restrictive interventions. *Education and Treatment of Children*, *10*(4), 311-325.

Foxx, R. M. (2008). Applied behavior analysis treatment of autism: The state of the art. *Child and Adolescent Psychiatric Clinics of North America, 17*(4), 821–834. doi: 10.1016/j.chc.2008.06.007

Fuller, J. L., & Fienup, D. M. (2018). A preliminary analysis of mastery criterion level: Effects on response maintenance. *Behavior Analysis in Practice, 11*(1), 1-8. doi: 10.1007/s40617-017-0201-0

Gunning, C., Holloway, J., Fee, B., Breathnach, Ó., Bergin, C. M., Greene, I., & Bheoláin, R. N. (2019). A systematic review of generalization and maintenance outcomes of social skills

intervention for preschool children with autism spectrum disorder. *Review Journal of Autism and Developmental Disorders, 6*(2), 172-199. doi: 10.1007/s40489-019-00162-1

Hong, E. R., Kawaminami, S., Neely, L., Morin, K., Davis, J. L., & Gong, L. Y. (2018). Tablet-based interventions for individuals with ASD: Evidence of generalization and maintenance effects. *Research in Developmental Disabilities, 79*, 130-141. doi: 10.1016/j.ridd.2018.01.014

Kassardjian, A., Leaf, J. B., Ravid, D., Leaf, J. A., Alcalay, A., Dale, S., Tsuji, K., Taubman, M., Leaf, R., McEachin, J., & Oppenheim-Leaf, M. L. (2014). Comparing the teaching interaction procedure to social stories: A replication study. *Journal of Autism and Developmental Disorders, 44*(9), 2329-2340. doi:10.1007/s10803-014-2103-0

Kazdin, A. E. (2012). *Behavior modification in applied settings*. Waveland Press.

Kennedy, C. H. (2002). The maintenance of behavior change as an indicator of social validity. *Behavior Modification, 26*(5), 594-604. doi: 10.1177/014544502236652

Kennedy, C. H. (2005). Social validity. In Kennedy, C. H., *Single-case designs for educational research.* (pp. 218-233). Pearson Education, Inc.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*(1), 26-38. doi: 10.1177/0741932512452794

Lang, R., Hancock, T. B., & Singh, N. N. (Eds.). (2016). *Early intervention for young children with Autism Spectrum Disorder*. Springer.

Ledford, J. R., Lane, J. D., & Tate, R. (2018). Evaluating quality and rigor in single case research. In Ledford J. R., & Gast D. L. (Eds.), *Single case research methodology:*

Applications in special education and behavioral sciences (3rd ed., pp. 239-281). Routledge. doi: 10.4324/9781315150666

Lucyshyn, J. M., Albin, R. W., Horner, R. H., Mann, J. C., Mann, J. A., & Wadsworth, G. (2007). Family implementation of positive behavior support for a child with autism: Longitudinal, single-case, experimental, and descriptive replication and extension. *Journal of Positive Behavior Interventions, 9*(3), 131-150. doi: 10.1177%2F10983007070090030201

Miltenberger, R. G. (2016). *Behavior modification: Principles and procedures* (6th Ed.). Cengage Learning.

Neely, L. C., Ganz, J. B., Davis, J. L., Boles, M. B., Hons, E. R., Ninci, J., & Gilliland, W. D. (2016). Generalization and maintenance of functional living skills for individuals with autism spectrum disorder: A review and meta-analysis. *Review Journal of Autism and Developmental Disorders, 3*(1), 37-47. doi: 10.1007/s40489-015-0064-7

Neely, L., Garcia, E., Bankston, B., & Green, A. (2018). Generalization and maintenance of functional communication training for individuals with developmental disabilities: A systematic and quality review. *Research in Developmental Disabilities, 19,* 116-129. doi: pallas2.tcl.sc.edu/10.1016/j.ridd.2018.02.002

Noell, G. H., Roane, H. S., VanDerHeyden, A. M., Whitmarsh, E. L., & Gatti, S. L. (2000). Programming for the generalization of communication to the classroom following assessment and training outside of the classroom. *School Psychology Review*, *29*(3), 429-442. doi: 10.1080/02796015.2000.12086028

Polanin, J. R., Pigott, T. D., Espelage, D. L., & Grotpeter, J. K. (2019). Best practice guidelines for Abstract screening large- evidence systematic reviews and meta- analyses. *Research Synthesis Methods, 10*(3), 330-342. doi: 10.1002/jrsm.1354

Rusch, F. R., & Kazdin, A. E. (1981). Toward a methodology of withdrawal designs for the assessment of response maintenance. *Journal of Applied Behavior Analysis, 14*(2), 131-140. doi: 10.1901%2Fjaba.1981.14-131

Schlosser, R. W., & Lee, D. L. (2000). Promoting generalization and maintenance in augmentative and alternative communication: A meta-analysis of 20 years of effectiveness research. *Augmentative and Alternative Communication, 16*(4), 208–226. doi: 10.1080/07434610012331279074

Shadish, W.R., Sullivan, K.J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavioral Research Methods, 43(4),* 971–980. doi: 10.3758/s13428-011-0111-y

Steinbrenner, J. R., Hume, K., Odom, S. L., Morin, K. L., Nowell, S. W., Tomaszewski, B., Szendry, S., McIntyre, N. S., Yücesoy-Özkan, S., & Savage, M. N. (2020). *Evidence-based practices for children, youth, and young adults with autism.* The University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Institute, National Clearinghouse on Autism Evidence and Practice Review Team. Retrieved from https://ncaep.fpg.unc.edu/sites/ncaep.fpg.unc.edu/files/imce/documents/EBP%20Report%202020.pdf

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*(2), 349-367. doi: 10.1901/jaba.1977.10-349

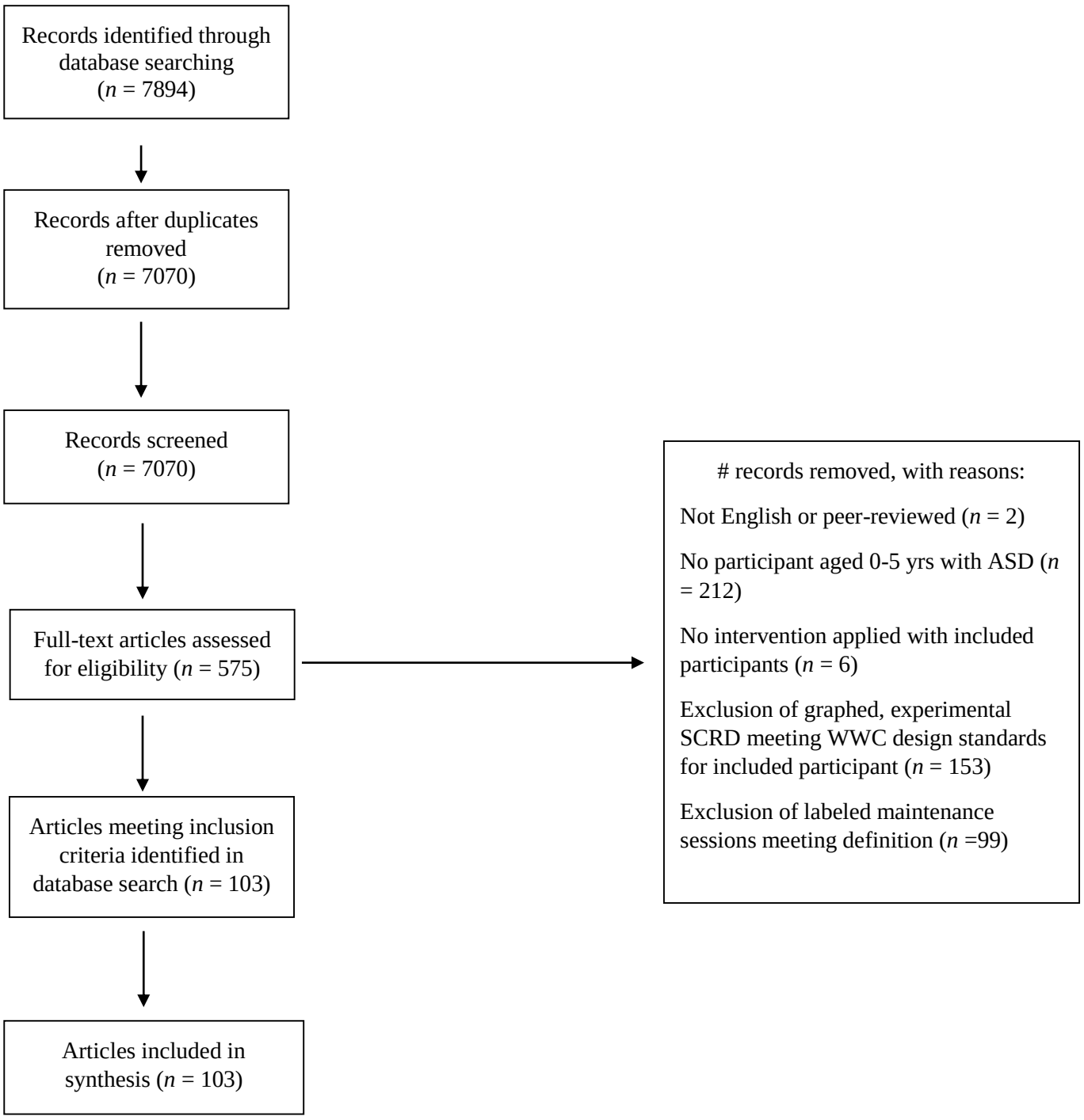
Waffenschmidt, S., Knelangen, M., Sieben, W., Bühn, S., & Pieper, D. (2019). Single screening

versus conventional double screening for study selection in systematic reviews: a methodological systematic review. *BMC medical research methodology*, *19*(1), 1-9. doi: 10.1186/s12874-019-0782-0

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., Brock, M. E., Plavnick, J. B., Fluery, 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. doi: 10.1007/s10803-014-2351-z

Figure 1

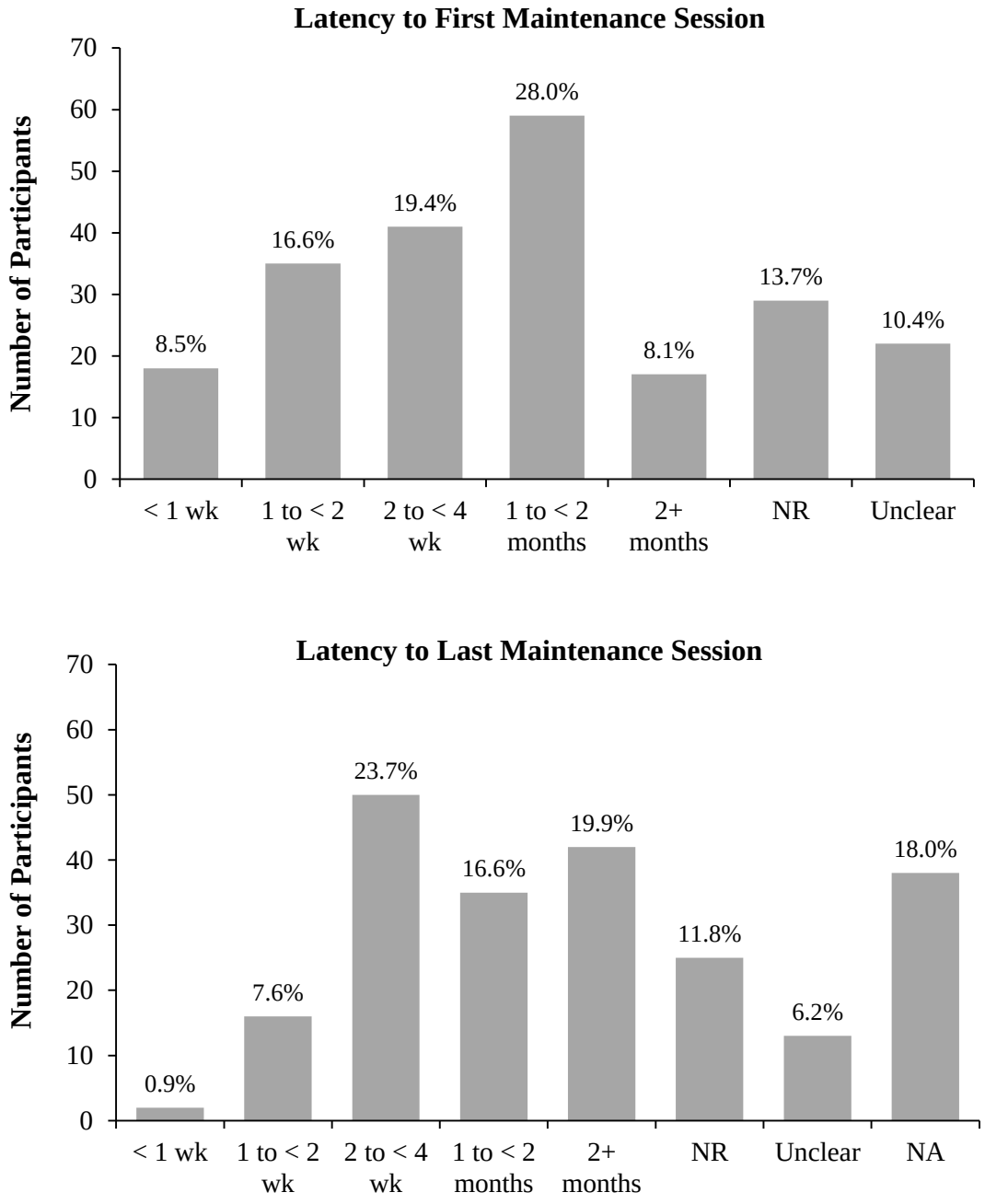
PRISMA Flow Diagram



*Adapted from "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," by D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, & PRISMA Group, 2009. *PLoS Medicine*, *6*, e1000097.

Figure 2

Latency to Maintenance Sessions



Note. NA = not applicable (i.e., only 1 maintenance session available); NR = not reported

Table 1

Journals of Study Publication

Journal	Number	Percentage
Autism	1	0.97%
Behavior Analysis in Practice	1	0.97%
Behavior Modification	1	0.97%
Behavioral Development	1	0.97%
Behavioral Disorders	1	0.97%
Behavioral Interventions	3	2.91%
Behaviour Change	1	0.97%
Early Education and Development	1	0.97%
Education and Training in Autism and Developmental Disabilities	4	3.88%
Education and Treatment of Children	5	4.85%
Educational Sciences: Theory & Practice	2	1.94%
Developmental Neurorehabilitation	2	1.94%
Focus on Autism and Other Developmental Disabilities	6	5.83%
International Journal of Psychology and Psychological Therapy	1	0.97%
Journal of Applied Behavior Analysis	15	14.56%
Journal of Autism and Developmental Disorders	17	16.50%
Journal of Behavioral Education	1	0.97%
Journal of Developmental and Physical Disabilities	8	7.77%
Journal of Communication Disorders	1	0.97%
Journal of Early Intervention	3	2.91%
Journal of Positive Behavior Interventions	3	2.91%
Journal of Special Education Technology	1	0.97%
Language, Speech, and Hearing Services in Schools	1	0.97%
Remedial and Special Education	1	0.97%
Research and Practice for Persons with Severe Disabilities	3	2.91%
Research in Autism Spectrum Disorders	9	8.74%
Research in Developmental Disabilities	2	1.94%
School Psychology Review	1	0.97%
Topics in Early Childhood Special Education	7	6.80%

Note. There were 103 total studies included in the review.

Table 2

Participants, Implementers, and Settings of Included Studies

	Number	Percentage
Participants[a]		
Total	334	100%
Aged 0-5 with ASD[b]	261	78.1%
Met review criteria[b]	221	66.1%
Implementer		
Natural	35	34.0%
Research staff	50	48.5%
Both	7	6.8%
Not reported or Not clear	11	10.7%
Country		
Within US	87	84.5%
Outside of US or NR	16	15.5%
Setting		
Clinic	30	29.1%
Community	1	1.0%
Home	17	16.5%
School/Daycare	46	44.7%
Multiple	9	8.7%

Note. US = United States; NR = Not reported

[a]Total participants across studies. All other totals represent number of studies.

[b]Percentages calculated from total number of participants. All other percentages calculated from total number of studies (n = 103).

Table 3

Characteristics of Designs and Dependent Variables of Included Studies

Characteristic	Number	Percentage
Design		
ABAB	5	4.9%
Adapted alternating treatments	8	7.8%
Alternating treatments	11	10.7%
Multiple baseline	53	1.0%
Multiple probe	20	51.5%
Parallel treatments	3	19.4%
Other	3	1.9%
Desired Outcome[a]		
Acquisition	132	91.7%
Reduction	12	8.3%
Domain/Instructional Outcome[a]		
Adaptive/Self-help	8	5.6%
Challenging/Interfering behaviors	11	7.6%
Cognitive	1	0.7%
Communication	26	18.1%
Joint attention	19	13.1%
Mental health	1	0.7%
Motor	0	--
Play	17	11.8%
Pre-academic/Academic	23	16.0%
Self-determination	0	--
School readiness	6	4.2%
Social	32	22.2%
Vocational	0	--
Additional Data		
Mastery Criterion	60	58.3%
Generalization	43	41.8%
Treatment Fidelity	72	69.9%
Social Validity	48	48.6%

Note. Domain categories from Steinbrenner et al., 2020. MB = multiple baseline, MP = multiple probe

[a]Multiple domains were coded per study when applicable (n = 144).

Table 4

Primary Evidence-Based Practice Implemented

EBP Category	Number	Percentage
Antecedent-Based Interventions (ABI)	1	1.0%
Augmentation & Alternative Communication (AAC)	12	11.5%
Behavioral Momentum Intervention (BMI)	1	1.0%
Cognitive/Behavioral Instructional Strategies (CBIS)	0	—
Differential Reinforcement of Alternative, Incompatible, or Other Behavior (DR)	4	3.9%
Direct Instruction (DI)	0	—
Discrete Trial Teaching (DTT)	13	12.5%
Exercise and Movement (EXM)	1	1.0%
Extinction (EXT)	1	1.0%
Functional Behavior Assessment (FBA)	1	1.0%
Functional Communication Training (FCT)	2	1.9%
Modeling (MD)	3	2.9%
Music-Mediated Intervention (MMI)	1	1.0%
Naturalistic Intervention (NI)	12	11.5%
Parent-Implemented Intervention (PII)	7	6.7%
Peer-Based Instruction and Intervention (PBII)	8	7.7%
Prompting (PP)	12	11.5%
Reinforcement (R)	7	6.7%
Response Interruption/Redirection (RIR)	2	1.9%
Self-Management (SM)	0	—
Sensory Integration® (SI)	0	—
Social Narratives (SN)	5	4.8%
Social Skills Training (SST)	3	2.9%
Task Analysis (TA)	0	—
Technology-Aided Instruction and Intervention (TAII)	5	4.8%
Time Delay (TD)	8	7.7%
Video Modeling (VM)	13	12.5%
Visual Supports (VS)	3	2.9%
Behavior Intervention Package (BIP)	5	4.8%
None (NA)	2	1.9%

Note. Categories from Steinbrenner et al., 2020. EBP = evidence-based practice; DV = dependent variables

Table 5

Characteristics of Maintenance Conditions

By Study[a]	Number	Percentage
Condition label		
Maintenance	44	42.7%
Follow up	35	34.0%
Both	24	23.3%
Assessment context		
Identical to intervention	74	71.8%
Generalization only	8	7.8%
Both	21	20.4%
Intervention implemented		
In full	5	4.9%
In part:	24	23.3%
Faded	*2*	*1.9%*
Thinned	*4*	*3.9%*
Nonsystematic	*9*	*8.7%*
Other	*9*	*8.7%*
None	63	61.2%
Unclear	11	10.7%
By Participant[b]	**Number**	**Percentage**
Number of sessions		
Mean	3.81	—
Range	1 - 37	—
$\geq$ 3 sessions	136	61.5%

Note. [a] Total number of studies (N = 103); [b] Total number included participants (N = 221)