University of Kent

Kent Academic Repository

Bunyan, Nathan, Feuerer, Clarissa and Vostanis, Athanasios (2026) *Evaluating the effects of skill-based treatment: A systematic review and meta-analysis of single-case studies.* Journal of Autism and Developmental Disorders . ISSN 0162-3257.

Downloaded from

https://kar.kent.ac.uk/116549/ The University of Kent's Academic Repository KAR

The version of record is available from

https://doi.org/10.1007/s10803-026-07547-2

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in ***Title of Journal*** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

Journal of Autism and Developmental Disorders
https://doi.org/10.1007/s10803-026-07547-2

ORIGINAL ARTICLE

Evaluating the Effects of Skill-Based Treatment: A Systematic Review and Meta-Analysis of Single-Case Studies

Nathan Bunyan[1] · **Clarissa Feuerer**[1] · **Athanasios Vostanis**[1]

Received: 13 February 2026 / Accepted: 13 September 2026
© The Author(s) 2026

Abstract

Purpose The skill-based treatment for challenging behaviour (SBT) is an intervention package that combines functional communication training with contingency-based delay and denial tolerance training. This systematic review and meta-analysis aimed to (a) investigate the application of SBT across diverse populations, setting, and target behaviours; and (b) to evaluate the effectiveness and social validity of SBT in reducing challenging behaviours.

Methods This review was pre-registered on PROSPERO (ID: CRD42024561455). Seven databases were searched. Data were extracted on characteristics of studies, participants, and applications of SBT. Literature was assessed for quality using the Evaluative Method and Single Case Design Risk of Bias Tool. SBT outcomes were quantified with Baseline-Corrected Tau (Tau-BC).

Results A total of 28 studies were reviewed, representing 99 individual SBT applications. SBT has been applied across participants with varying diagnoses, challenging behaviours, and levels of ability, though mostly in controlled settings by skilled implementers. Studies provided evidence for the social validity of SBT. Baseline-Corrected Tau indicated SBT shows good reduction in challenging behaviour ($M = 0.47$, range: −0.35–1.00). The Evaluative Method indicated the literature to be of varying quality (borderline-strong = 7%; adequate = 13%; borderline-adequate = 43%; weak = 37%), and the evidence-based practice score suggests SBT to be a potentially evidence-based practice.

Conclusion The evaluative method indicates SBT to be an emerging evidence-based practice for addressing challenging behaviour, although the majority of the literature scored lower than adequate. We make recommendations for strengthening the literature base and outline key implications for practitioners.

Keywords Meta-analysis · Skill-based treatment · Challenging behaviour · Problem behavior · Functional communication training · Schedule thinning

Introduction

Challenging behaviours can impact the safety and quality of life of the person and those around them and may result in responses that are restrictive or exclusionary (Emerson et al., 2014; Holden & Gitlesen, 2006). Increased presentation of challenging behaviours correlates with communication difficulties (e.g., Holden & Gitlesen, 2006), and research indicates that challenging behaviours often serve a communicative function (Melanson & Fahmie, 2023). Therefore, behavioural interventions are often directed towards teaching communication skills that reduce or replace the need for challenging behaviour (e.g., Carr & Durand, 1985) along with additional ecological manipulations (Gore et al., 2013).

One such intervention that has received great attention in recent years is skill-based treatment (SBT). First introduced by Hanley et al. (2014), SBT is purported to be an efficient, effective, and safe way to address challenging behaviours (Jessel et al., 2018, 2022). The SBT intervention package comprises two core components: functional communication training (FCT) and delay and denial tolerance training. FCT is one of the most widely researched behavioural interventions, with a growing number of systematic reviews and meta-analyses attesting to its efficacy in addressing challenging behaviour (Chezan et al., 2018; Gerow et al., 2018; Heath et al., 2015). In FCT, the interventionist teaches an

✉ Athanasios Vostanis
a.vostanis@kent.ac.uk

[1] Tizard Centre, University of Kent, Canterbury CT2 7NF, UK

Published online: 28 September 2026

alternative functional communication response (FCR) that fulfils the same function as the challenging behaviour (Carr & Durand, 1985). Because this FCR is initially on a dense schedule of reinforcement, the schedule must then be thinned to increase the likelihood of the outcomes being maintained under natural conditions (Hagopian et al., 2011; Muharib et al., 2019; Tiger et al., 2008). Put simply, while initially the FCR achieves its consequence every time, this is gradually reduced to a level that is feasible to sustain in the person's daily life. In SBT, the method of schedule thinning is referred to as delay and denial tolerance training. The following sections will outline the SBT process in more detail.

Preassessments

Prior to commencing with FCT, the analyst must determine the reinforcement contingency that is maintaining the challenging behaviour. To do so, behaviour analysts typically conduct a functional analysis (FA) to identify environmental events associated with challenging behaviour, including establishing operations (EOs) that evoke the target behaviour and the consequences that reinforce and maintain it. The analyst contrives test and control conditions where an evocative event (EO) is present or absent; an increased rate of challenging behaviour in a test condition indicates the behaviour's function. Most FAs in the behavioural literature have been based on the analysis first presented by Iwata et al. (1994), with separate, isolated conditions for each possible function (e.g., *attention*, *escape from demands*, or *tangible).* Most often, implementers of SBT use a type of FA called an interview-informed synthesised contingency analysis (IISCA), first introduced by Hanley et al. (2014). In an IISCA, a control condition is contrasted with a single test condition in which all putative EOs and reinforcers are combined (i.e., functions are *synthesised*). The synthesised contingency is determined from an open-ended interview with caregivers (Hanley, 2012). Where reinforcing events are indicated as occurring together (e.g., escape from demands to preferred activities), this is the contingency that is tested. However, if a synthesised contingency is not indicated, then isolated contingencies may be tested (e.g., Fiani & Jessel, 2022), meaning synthesised contingencies are a common but not essential feature of SBT.

Skill-Based Treatment

Functional Communication Training

In FCT, the functional reinforcement contingency identified in the FA is then made contingent on a replacement behaviour (the FCR). If multiple isolated contingencies are controlling the challenging behaviour, the analyst could implement SBT separately for each contingency (e.g., Kastner, 2025), but generally implementers conduct a single SBT application with an omnibus FCR that accesses all reinforcers (e.g., saying "my way"; Jessel et al., 2018). Using such an omnibus mand (i.e., request) may avoid increases in challenging behaviour when FCT is applied to one contingency but not yet to another (Ghaemmaghami et al., 2016a, 2016b). Once this initial FCR is acquired, it may then be shaped by reinforcing progressively more complex FCRs (Ghaemmaghami et al., 2018). See Online Resource 1 for a flow chart showing the responses that are reinforced in each phase of SBT.

Tolerance Response Training

In natural contexts, there will be occasions when reinforcement is denied or not available. Teaching a specific tolerance response, such as taking a breath and saying "okay", can increase tolerance for denial to reinforcement (Luczynski & Hanley, 2013). In SBT, tolerance response training takes place after the shaping of FCRs and before progressing to schedule thinning. In a proportion of trials in this phase, the FCR is given a denial cue (e.g., "not right now") and engaging in the tolerance response results in reinforcement. In the remaining trials, the FCR continues to be reinforced. This is done to maintain the strength of these responses and avoid the resurgence of challenging behaviour (Ghaemmaghami et al., 2016a, 2016b).

Schedule Thinning

In SBT, a contingency-based schedule thinning procedure is used to establish a chain of contextually appropriate behaviour between denial and accessing reinforcement (Jessel et al., 2018). In contingency-based schedule thinning, participants are required to engage in an alternative behaviour and/or refrain from engaging in challenging behaviour for the duration of the delay period (Ghaemmaghami et al., 2016a, 2016b). In other words, individuals are encouraged to willingly engage in meaningful tasks or activities before accessing reinforcement. Requiring the learner to engage in an alternative behaviour during delay periods is likely to produce greater tolerance and lower rates of challenging behaviour than simply increasing the time before providing reinforcement (Drifke et al., 2020; Ghaemmaghami et al., 2016a, 2016b). In SBT, the number, duration, or difficulty of tasks is gradually increased. Similar to tolerance response training, the FCR and tolerance response continue to be reinforced in some trials. Following schedule thinning, once a meaningful level of engagement in non-preferred activities is achieved, the skills are often transferred to other

implementers, contexts, and tasks of importance (this is referred to in research as a treatment extension).

The Efficacy of SBT

It has now been over a decade since the initial publication demonstrating the SBT package (Hanley et al., 2014). To date, reviews and meta-analyses have focused primarily on the IISCA and its variations (Coffey et al., 2021; Layman et al., 2023; Metras & Jessel, 2021). Only a single review has been conducted looking at SBT itself (Hustyi et al., 2025). Hustyi and colleagues reviewed 21 studies applying SBT published up to September 2023. Of the 87 participants in their sample, most were male, 3 to 10 years-old, and diagnosed with autism. They found that FCRs were mostly vocal verbal, and few participants were supported to progress from omnibus to specifying FCRs. Hustyi and colleagues reported on efficacy in terms of a percentage reduction in challenging behaviour, finding an average of 98.2% reduction (range; 90–100%) across the 56 participants for whom percentage reduction was reported.

Hustyi and colleagues' scoping review gave a welcome and timely overview of the literature on SBT (Hustyi et al., 2025). However, they did not extract outcome data from the participants or compute an effect size or similar metric. They also did not assess the quality or risk of bias for the literature they reviewed. The present review will extend the literature by conducting a systematic review and meta-analysis of SBT, including assessment of research quality, risk of bias, and quantification of intervention outcomes. This systematic review and meta-analysis had two key aims:

1. to evaluate the current status of the evidence for the efficacy and social validity of SBT in reducing challenging behaviour; and
2. to investigate the application of SBT across diverse settings, populations, and target behaviours.

Methods

This review was pre-registered on PROSPERO (ID: CRD42024561455) and followed a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology (Page et al., 2021). A full protocol was not produced. See Online Resource 2 for a completed PRISMA Checklist.

Eligibility Criteria

Empirical studies were included if they were written in English, used a single case experimental design, and applied SBT in reducing challenging behaviour with a human participant. To be considered SBT, the intervention must have included the following:

1. Use of a FA to establish the function of the target behaviour.
2. Applying FCT to teach a functionally equivalent communicative response.
3. Teaching a specific tolerance response to be emitted following delay/denial cues.
4. Using a contingency-based schedule thinning procedure.

Searches

Searches of databases were conducted in June 2024 and were updated in January 2025 and August 2025 to retrieve records added after preceding searches. We searched the following databases: British Educational Index, ERIC, PubMed, PsycINFO, Scopus, and Web of Science. To include unpublished grey literature, we also searched ProQuest Dissertations and Theses. Two further grey literature databases could not be searched due to consistently returning errors (Open Access Theses and Dissertations) and outage following a cyber-attack (eThOS; British Library, 2026). Searches were limited to records published after 2013 to encompass the years since the seminal paper on SBT (Hanley et al., 2014). Search terms required either (a) words directly related to SBT (e.g., "practical functional assessment", "SBT") or (b) terms relating to FCT in combination with those relating to delay tolerance training (e.g., "delay", "tolerance", "reinforcement"). Full search strings for all databases can be found in Online Resource 3. Following initial scoping searches, searches on some databases (i.e., Web of Science, Scopus, and ProQuest Dissertations and Theses) were restricted to certain fields to reduce the quantity of irrelevant records. The search strategy was confirmed with a subject librarian prior to commencement.

We employed several additional search methods. The following journals were hand searched, reviewing issues from January 2014 onwards for any eligible studies:

1. Behavior analysis in practice
2. Behavior modification
3. Behavioral interventions
4. Behaviour analysis research and practice
5. Child and family behavior therapy
6. Education and treatment of children
7. European journal of behavior analysis
8. Journal of applied behaviour analysis
9. Journal of behavioral education
10. Journal of organisational behavior management
11. Journal of the experimental analysis of behaviour

12. Perspectives on behavior science
13. The analysis of verbal behavior.

We also searched the publications list on the Practical Functional Assessment Blog and emailed key authors to request any further unpublished works. Following full-text screening, we conducted backwards searching of references of included studies were searched and forwards searching of citations on Google Scholar. Finally, a Google Scholar alert was set up to identify studies published after searches were conducted. See Online Resource 2 for a PRISMA diagram depicting the search and screening process.

Screening

Screening was conducted in Rayyan (Ouzzani et al., 2016). After deduplication, 13,175 records remained. Two researchers independently screened all articles for inclusion across two stages. First, titles and abstracts were reviewed against the inclusion criteria and 13,127 records were excluded. The remaining 48 records were combined with those obtained through hand searches, author correspondences, and forwards and backwards searches. The resulting 64 full texts were screened for eligibility, with 34 excluded at this stage. For reasons for exclusion, see Online Resource 2. Where a dissertation and published version of the same study were identified, the dissertation was excluded but was checked for supplementary information during data extraction. The final sample comprised 30 studies. Details of included studies can be found in Online Resource 4.

Data Extraction and Synthesis

The first and second authors double-coded both stages of screening, the data extraction, quality assessment, and risk of bias assessment. The first author is a behavioural practitioner with four years' clinical experience, currently completing the UKBA(cert) Supervision and Competence Standards. The second author is a Board-Certified Behaviour Analyst with 15 years of clinical experience. Participants and SBT applications where the intervention did not meet the inclusion criteria (e.g., no tolerance training) were excluded from data analysis. Where relevant, corresponding authors were contacted for additional information. Data were collated, analysed, and visualised using Microsoft Excel.

Study and Participant Characteristics

For each study, we recorded the publication, country of origin, experimental design used, number of participants, the measurement procedures used, details of interobserver agreement and procedural fidelity, and any secondary dependent measures. The researcher also noted whether a variation of the SBT package was used (e.g., the enhanced choice model; Rajaraman et al., 2022a, 2022b). Information extracted for each participant included age, sex, ethnicity, diagnoses, and expressive and receptive communication ability. We recorded the responses included in the target response class and whether less dangerous behaviours and precursors were open to reinforcement. Expressive communication ability was extracted as reported by the authors, and was then coded as: Full Sentences, Single Words, Multiple Words/Phrases, Nonverbal, Aided Augmentative and Alternative Communication (AAC), or Unaided AAC. Aided AAC included electronic devices, picture exchanges, and communication boards. Unaided AAC included any non-speech vocalisations, hand signing, gestures, and expressions (Beukelman & Light, 2020).

Intervention Characteristics

For each SBT application, we extracted the setting, the identity and education level of the primary implementer, the reinforcement contingency applied, and details of pre-assessments (e.g., FAs and preference assessments). We recorded the form and modality of each progressive FCR and tolerance response. Modality of FCRs were coded as Vocal Verbal, Aided AAC, and Unaided AAC (Beukelman & Light, 2020). Extinction for challenging behaviour was coded as either Extinction, Extinction with Prompted FCR, Extinction with Blocking, or No Extinction. For schedule thinning, researchers extracted the criteria for providing reinforcement, the criteria for progressing the schedule thinning, and the response given when challenging behaviour occurred. We also noted whether generalisation and maintenance were tested, details of treatment extensions, the intensity (or "dosage") of the intervention, and any information on cost- or time-effectiveness. Finally, we recorded details of any measurement of social validity of the goals, procedures, and outcomes (Wolf, 1978).

Descriptive Synthesis

Data relating to study, participant, and SBT characteristics were synthesised descriptively. Categorical variables were summarised as frequencies and percentages, and continuous variables were summarised using means, medians and ranges, as appropriate. The unit of analysis varied according to the characteristic being summarised. Study-level characteristics were synthesised across the 30 included studies, participant-level characteristics across eligible participants, and intervention-level characteristics across individual SBT applications.

Intervention Outcomes

Data were extracted from graphs using WebPlotDigitizer (Rohatgi, 2021), which has shown good validity and intercoder reliability (Drevon et al., 2017). We quantified the outcomes of interventions with Baseline-Corrected Tau (Tau-BC; Tarlow, 2017). Tau-BC is a non-overlap metric that uses rank order correlation to test for a statistically significant baseline trend; where a trend is present, it adjusts for this trend before calculating the Tau value. Nonoverlap metrics generally align well with visual analysis (Alresheed & Machalicek, 2023).

We calculated Tau-BC using the Single Case Design Effect Size calculator (Pustejovsky et al., 2024) in R (R Core Team, 2023). Each A–B pair (baseline–intervention pairing) from participant graphs was entered into the calculator. For the intervention phase, we included from the first datapoint of the intervention to the last datapoint of maintenance to encompass all phases of SBT. Tau-BC was then computed for each pair before averaging across A–B pairs and SBT applications to produce a single Tau-BC value, a 95% confidence interval, and a standard error for each participant. The weighting used for aggregation was the inverse of the variance. A Tau score less than 0.20 indicates a small change, between 0.20 and 0.60 is a moderate change, between 0.60 and 0.80 is a large change, and greater than 0.80 indicates a very large change. If a study measured challenging behaviour and precursor behaviour separately, they were summed before calculating Tau-BC to give a measure comparable to other studies that provided only a combined value.

Quality Assessment/Risk of Bias

We assessed the quality of studies using the Evaluative Method (Reichow, 2011). With this tool, studies are scored either high, acceptable, or unacceptable against six primary indicators (Participant Characteristics, Dependent Variable, Independent Variable, Baselines, Visual Analysis, and Experimental Control). Secondary indicators (IOA, Kappa, Blind Raters, Fidelity, Generalisation and Maintenance, and Social Validity) are scored as either present or not present. From these scores is derived an overall quality rating: strong, adequate, or weak. Two further ratings, borderline strong and borderline adequate, were added to increase sensitivity (Cowell et al., 2024; Tomlinson et al., 2018). The Evaluative Method also allows for calculating an evidence-based practice *Z* score. This was conducted following the quality assessment to indicate the extent to which SBT can be considered an evidence-based practice. Details of how overall scores and Z-scores are calculated can be found in Online Resource 5.

The Single-Case Design Risk of Bias Tool (Reichow et al., 2018) was applied to each study. The tool translates the types of risk of bias from the Cochrane risk of bias tool (Higgins et al., 2016) into those relevant to single-case designs: selection bias, performance bias, detection bias, and other risk of bias. Each type of bias contains individual items that are scored as high risk, unclear risk, or low risk.

Interrater Agreement

Disagreements were resolved until the overall agreement was 100%. Before resolving disagreements, the mean agreement was 93% (range: 74%–100%) for the data extraction, 81% (range: 50%–100%) for quality assessment, and 78% (range: 41%–97%) for risk of bias assessment. Online Resource 6 details interrater agreement scores.

Results

Study Characteristics

The final sample included 27 published peer-reviewed studies and 3 dissertations. Journals of publications were: Journal of Applied Behaviour Analysis (n = 8), Behavioral Interventions (n = 7), Behavior Analysis in Practice (n = 4), Clinical Case Studies (n = 2), Education and Treatment of Children (n = 2), Journal of Autism and Developmental Disorders (n = 2), European Journal of Behaviour Analysis (n = 1), and Research and Practice for Persons with Severe Disabilities (n = 1). Almost all studies originated from the United States of America (n = 26; 87%). However, two USA-based studies had participants in India (Rajaraman et al., 2024), Western Canada, and Bahrain (Metras et al., 2024). The median number of participants was 3 (range: 1–25). The majority of included studies were conducted by researchers connected to the originators of SBT, but 11 studies were authored by researchers from unrelated research groups (i.e., Boyle et al., 2020; Cervi et al., 2024; Chan, 2021; Edelstein et al., 2021; Ferguson et al., 2020; Kastner, 2025; Lundy et al., 2021; Quiroz et al., 2024; Santa Maria, 2025; Strand & Eldevik, 2017; Taylor et al., 2018). Four studies applied SBT within the enhanced choice model (Metras et al., 2024; Rajaraman et al., 2022a, 2022b, 2024; Staubitz et al., 2022) and four used telehealth. All studies demonstrated experimental control by applying the reinforcement contingency progressively to different alternative communication responses, described as using "the logic of a changing criterion design", usually embedded within a multiple baseline design.

Participant Characteristics

After omitting participants whose SBT did not meet the inclusion criteria, there were a total of 107 participants in the sample. Results for participant characteristics can be seen in Table 1. Around three-quarters of participants were male (n = 81; 76%), and the significant majority (n = 103, 96%) were aged under 14 years. Of the 41 participants for whom ethnicity was reported, most (n = 23; 56%) were white. The majority of participants were autistic (n = 88, 80%), and 22% had an intellectual disability. Most participants had some level of vocal verbal language ability (n = 74, 69%). Receptive communication levels were only able to be coded for 14 participants.

Target Behaviours and FCRs

Most participants' target responses were a compound class including multiple topographies of challenging behaviour. For details of response types included in these classes, see Table 1. Less severe forms of behaviour were open to reinforcement (i.e., an open response class) for 72% of participants (n = 77). Most applications included some shaping of FCRs. Vocal verbal was the most common modality across FCRs (71% for initial FCRs and 61% for terminal FCRs). Aided AAC was more common (22% for initial FCRs and 22% for terminal FCRs) than unaided AAC (1% for initial FCRs and 0% for terminal FCRs). Terminal FCRs were more likely to include unaided AAC with a primary vocal verbal (15%) or aided AAC response (2%; e.g., requiring the participant to also put their hands on their lap or orient to speaker while emitting the main FCR component). Tolerance responses were also most likely to be of vocal verbal modality (40%). Only two tolerance responses used aided AAC (Coffey, 2019; Fruchtman et al., 2025). Most non-vocal tolerance responses were unaided AAC (11%; e.g., nodding the head or clasping hands together). Detail of modalities for each FCR are shown in Online Resource 7.

SBT Characteristics

There were 111 SBT applications across the 107 participants. Results for SBT characteristics are shown in Table 2. Most applications (n = 105; 95%) applied contingencies derived from IISCAs (or variations thereof), with the majority applying a synthesized reinforcement contingency (n = 95; 86%). Only 15 applications applied a preference assessment, with most relying on preferences identified through the functional assessment interviews. Four studies applied SBT within an enhanced choice model (representing 10 applications). Terminal schedules averaged 7.6 min (range: 0.5–60 min) or 24 tasks (range: 1–73), although these should

Table 1 Participant Characteristics

Characteristic	N%
Age	
0–4	29%
5–9	54%
10–14	13%
15–19	3%
20+	1%
Sex/Gender	
Male	76%
Female	16%
Unclear/not provided	8%
Ethnicity	
White	21%
Black	7%
Latinx	7%
Arab	1%
Asian	1%
Unclear/not provided	62%
Expressive Communication	
Full sentences	30%
Multiple words	24%
Single words	15%
Nonvocal	13%
Aided AAC	7%
Multiple	2%
Unclear/not provided	8%
Receptive communication	
Multistep instructions	12%
Single-step instructions	1%
Unclear/not provided	87%
Diagnosis[a]	
Autism	82%
Intellectual disability	22%
ADHD	16%
No diagnoses	12%
Generalised anxiety disorder	2%
Depression	1%
Down syndrome	1%
Dyspraxia	1%
Foetal alcohol exposure	1%
Fragile X	1%
Global developmental delay	1%
Obsessive compulsive disorder	1%
Sensory processing disorder	1%
Tourette's syndrome	1%
Unclear/not provided	3%
Target Response Class Included[a]	
Aggression	76%
Elopement	16%
Property destruction	24%
Sexually inappropriate behaviour	1%
SIB	36%
Tantrums	32%
Vocalisations	61%
Unclear/not provided	6%
Response class	

Table 1 (continued)

Characteristic	N%
Open response class	72%
Closed response class	28%
FCRs	
Shaping of FCRs	95%
Single FCR	5%

AAC alternative and augmentative communication, *ADHD* attention deficit hyperactivity disorder, *FCR* functional communication response. Some percentages may not total 100% due to rounding

[a]Multiple-coded variable, so percentages may total more than 100

be interpreted with caution because schedule thinning was highly individualised and generally included variation of task difficulty as well as quantity or duration. SBT was most often implemented in outpatient clinics and by therapists, graduate students, or clinicians (e.g., BCBAs). Most applications were of lesser intensity, with fewer than 20 h per week allocated to SBT. Where cost and time data were available, total hours averaged 68.78 h (range: 23–167 h) and total costs averaged $5,984.91 (range: $2,975–8,650).

Intervention Outcomes

We computed Tau-BC values for 59 participants, of whom 81% were autistic (n = 48) and 68% were male (n = 40). Values could not be calculated for the remaining 48 participants. Reasons included graphed data not being available, at least one phase having fewer than three datapoints (preventing testing for a trend), or missing data. Details of Tau-BC values and reasons for exclusion for all participants are given in Online Resource 9. For most participants (n = 25), Tau values indicated a moderate change (Tau = 0.20–0.60). Eight participants showed a very large change (Tau > 0.80) and 15 showed a large change (Tau = 0.60–0.80). A small change (Tau < 0.20) was measured for eleven participants. When interpreting these figures, it should be noted that 42 participants had wide 95% CIs (i.e., with a width of 0.5–1), suggesting a wide margin of error.

Social Validity

Social validity was reported in 23 studies (see Online Resource 4 for details). In all instances, social validity was assessed by questionnaire, which represented 73% of participants (n = 78). Questionnaires were completed by caregivers (n = 65), members of the home care team (n = 16), or school staff (n = 6). Social validity questionnaires were also completed by two individuals who received SBT, with positive results in both instances (Diego in Leyman et al., 2025; Adam in Taylor et al., 2018). Wherever questionnaires asked respondents to rate the social validity of procedures (n = 73) and outcomes (n = 75), results were positive. Questions relating to the social validity of goals were less common: only six studies included them in the questionnaires, with positive responses in all instances.

Research Quality

Quality Assessment

Following the application of the Evaluative Method (Reichow, 2011), eleven studies were rated weak, thirteen studies were rated borderline-adequate, four studies were rated adequate, and two studies borderline-strong. Details of overall ratings and indicators for each study can be found in Online Resource 8. Six studies rated adequate or above were included in calculating the evidence-based practice Z score, representing 23 participants for whom visual analysis identified positive intervention outcomes. Of these participants, 83% were male (n = 19) and 78% were autistic (n = 18). The resulting Z score was 46, indicating SBT as a probable evidence-based practice.

Risk of Bias

Risk of bias ratings for the overall sample are summarised in Fig. 1. See the supplementary materials for details of ratings for individual studies. The study with the most risk of bias domains rated as low risk was Rajaraman et al., (2022a, 2022b; 7 of 9 indicators). Following that, three studies had six domains rated as low risk of bias (Lundy et al., 2021; Rose & Beaulieu, 2019; Staubitz et al., 2022). All the studies were scored as high risk on blinding of participants and personnel and blinding of outcome assessment. In half of the studies (n = 15), participants received training to emit FCRs prior to intervention measurement sessions, which involved prompting FCRs and contingent access to the functional reinforcers identified through the FA. This was considered high risk of bias for selective outcome reporting, as participants have already achieved non-baseline rates of FCRs and challenging behaviour before measurement begins, thereby establishing an inflated immediacy of effect.

Discussion and Conclusion

This systematic review and meta-analysis reviewed literature on SBT since the seminal publication of Hanley et al. (2014). In the decade since its introduction, there has been a sustained interest in SBT, with publications across eight journals, four countries, and from a variety of research teams. We extracted and analysed data from 30 studies, most of which were rated as weak using the Evaluative Method (Reichow et al., 2008). These studies represented 111 SBT

applications across 107 participants. The analysis sought to address two key research aims. The following sections will discuss the results in terms of these two aims and explore their implications for practice and research.

Research Aim 1: Evaluating the Evidence for the Efficacy and Social Validity of SBT in Reducing Challenging Behaviour

The analysis of intervention outcomes indicates SBT was effective in reducing challenging behaviour for the individuals in this review. The average duration of 68.78 h and average estimated cost of $5,984.91 indicate SBT may be a cost-effective intervention. The Tau-BC scores generally indicated a clinically significant reduction in challenging behaviour, although most applications had wide confidence intervals. Considering only those without wide 95% confidence intervals, eight showed a very large change, two a large change, six a moderate change, and one a small change. However, it is important to note that only three of these applications were from studies with a quality rating of adequate or above.

The results of the quality assessment indicate SBT to be a potentially evidence-based practice, although most studies were of weak quality. However, we identified that the Evaluative Method can be overly sensitive in how its indicators are scored, increasing the likelihood that a study will score poorly. For example, a study in which some participants' data are unstable or overlap across phases will score poorly on the Baseline or Visual Analysis indicators, even if all other participants had good outcomes and stable data. This does not necessarily relate to the study quality and may actually indicate a strong design with a lower risk of bias (e.g., a consecutive controlled case series where participants' data were published even if they were not as positive). Having a single primary indicator rated as unacceptable will mean the study cannot score above borderline-adequate. The risk of bias tool also highlighted possible indicators of bias. Taken together, the results of the quality assessments suggest caution in considering SBT as evidence-based at this time. Limitations to external validity are not uncommon in single-case research designs (Horner et al., 2005). While research in single-case designs should, in general, aim to improve this area (Burns, 2012), this meta-analysis found specific gaps within the SBT literature. Some actions researchers could take to improve the quality of the literature base include reporting inclusion criteria, providing clear operational definitions for target behaviours, reporting demographic information for participants and interventionists, measuring procedural fidelity, reporting the percentage of sessions within each phase for which interobserver agreement and fidelity were collected, considering the use

Table 2 Characteristics of SBT Applications

Variable	*N*%
Experimental Analysis	
IISCA	82%
Latency-based IISCA	4%
Performance-based IISCA	6%
Single-session IISCA	3%
Isolated FA	5%
Mand analysis	1%
Other Pre-assessments	
Preference assessment	14%
Setting	
Outpatient clinic	60%
Home	23%
Special education school	15%
Mainstream school	1%
Primary Implementer	
Therapist	39%
Graduate student	17%
Clinician/BCBA	13%
Parent	7%
Researcher	5%
Direct care staff	3%
Parent + clinician	1%
Unclear/not provided	15%
Implementer's Education Level	
Undergraduate	50%
Postgraduate	5%
Pre-undergraduate	2%
Unclear/not provided	43%
Contingencies	
Synthesised	86%
Attention + Escape	10%
Attention + Tangible	10%
Escape + Tangible	37%
Attention + Escape + Tangible	29%
Isolated	14%
Attention	2%
Escape	4%
Tangible	9%
Days per Week	
Up to 1	1%
Up to 2	4%
Up to 3	10%
Up to 4	11%
Up to 5	30%
Unclear/not provided	45%
Hours per Week	
Up to 20	23%
Greater than 20	24%
Unclear/not provided	48%
Extinction and FCR Prompting	
Extinction	83%
Extinction only	12%
Extinction + prompt FCR	59%
Extinction + block CB	5%
Unclear if prompted	8%

Table 2 (continued)

Variable	*N*%
No extinction	9%
Unclear/not provided	8%
ST – Criteria for Reinforcement	
Response based	65%
Time based	18%
Both	17%
ST – Response to CB	
Reset response requirement	37%
Did not reset requirement	23%
Unclear	41%

IISCA interview informed synthesised contingency analysis, *BCBA* Board Certified Behaviour Analyst, *DRA* differential reinforcement of alternative behaviour, *DRO* differential reinforcement of other behaviour. Some percentages may not total 100% due to rounding, *ST* Schedule Thinning

[a]Multiple-coded variable so percentages may total more than 100

of blinding, and conducting maintenance and generalisation tests. Additionally, including pre-session teaching data in measurement would give a more accurate reflection of participants' learning and avoid exaggerating the immediacy of effect (e.g., Rose & Beaulieu, 2019).

A high proportion of the studies in this review reported data on social validity (73%), which is a particular strength of SBT research compared with behaviour analytic research more generally (Leif et al., 2024). Overall, when social validity was reported, stakeholders found SBT acceptable and were pleased with the outcomes. Additionally, some studies incorporated other strategies which promote social validity, such as parent involvement in goal setting, parents being present or observing sessions, and children choosing whether to participate in SBT sessions. Involving stakeholders in goal setting can help increase social validity (Vollmer & Pendergrass, 2025). Where parents were observing and able to terminate sessions at will, they did not do so, which indicates that they felt comfortable with the intervention. Social validity data were rarely gathered from the individuals receiving SBT. Although children actively choosing to engage in SBT sessions could be interpreted as a sign of finding the intervention acceptable or even enjoyable (e.g., Rajaraman et al., 2022a, 2022b). While these data were not included as evidence toward social validity in this review, they offer more insight into social validity and acceptability of SBT, particularly where researchers measured other indicators such as engagement or interaction (Vollmer & Pendergrass, 2025). An avenue for further research would be including other indicators of social validity than just questionnaires (Barton et al., 2018), such as employing qualitative methodologies with various groups including, among others, autistic adults (Baiden et al., 2025), parents (Guinness et al., 2024), and the individuals receiving support (Hanley, 2010).

Research Aim 2: Investigating the Application of SBT Across Diverse Settings, Populations, and Target Behaviours

Our review identified several features of the literature on SBT which may be of interest to practitioners and researchers. SBT was mostly conducted by skilled implementers (e.g., therapists) in clinic settings. Additionally, as most researchers were connected to the originators of SBT, there are likely to be procedural similarities. Recipients of SBT were generally male, autistic, under the age of 9, and were able to speak in full sentences or multiple word phrases. While some studies implemented SBT in schools, with parents as implementers, with participants with other diagnoses, and with AAC users, they represent fewer overall applications of SBT. Only two participants were over 18. Consequently, the current evidence base reflects a relatively narrow range of participants, implementers, settings, and research teams. This limits confidence in the extent to which findings will generalise to less-represented populations and contexts. Practitioners should consider these limitations when looking to implement SBT within broader contexts and participant groups. If implementing SBT within these groups or contexts, practitioners may need to adapt procedures, implementation time may vary, or SBT may not be effective. While publications on a new intervention often come from a smaller pool of researchers, the field would benefit if more independent teams conducting SBT published their findings (Cook et al., 2015). Researchers could further strengthen the literature by focusing on these less-represented populations and contexts, with particular attention to any adjustments that may support more effective implementation.

While the overall structure of SBT remains the same (as outlined in the introduction), the literature reviewed shows that implementers often adapt the specific procedures to individualise SBT. This individualisation aligns with the concept of evidence-based practice not as a prescriptive adherence to a protocol, but as a decision-making process where practitioners use clinical expertise to select empirically evidenced procedures, with relevance to the needs and preferences of the individual (Contreras et al., 2022).

Adaptations were made in three broad components of SBT. For example, in some cases, specified mands were taught, but most applications used an omnibus mand. Specific mands are beneficial in that they enable a learner to better communicate their wants and needs to a variety of listeners, whereas an omnibus mand is less likely to be reinforced by unfamiliar communication partners (Ward et al., 2021). However, Ward and colleagues found that learning an omnibus mand does not preclude learners from later acquiring specified mands. While almost all reviewed applications

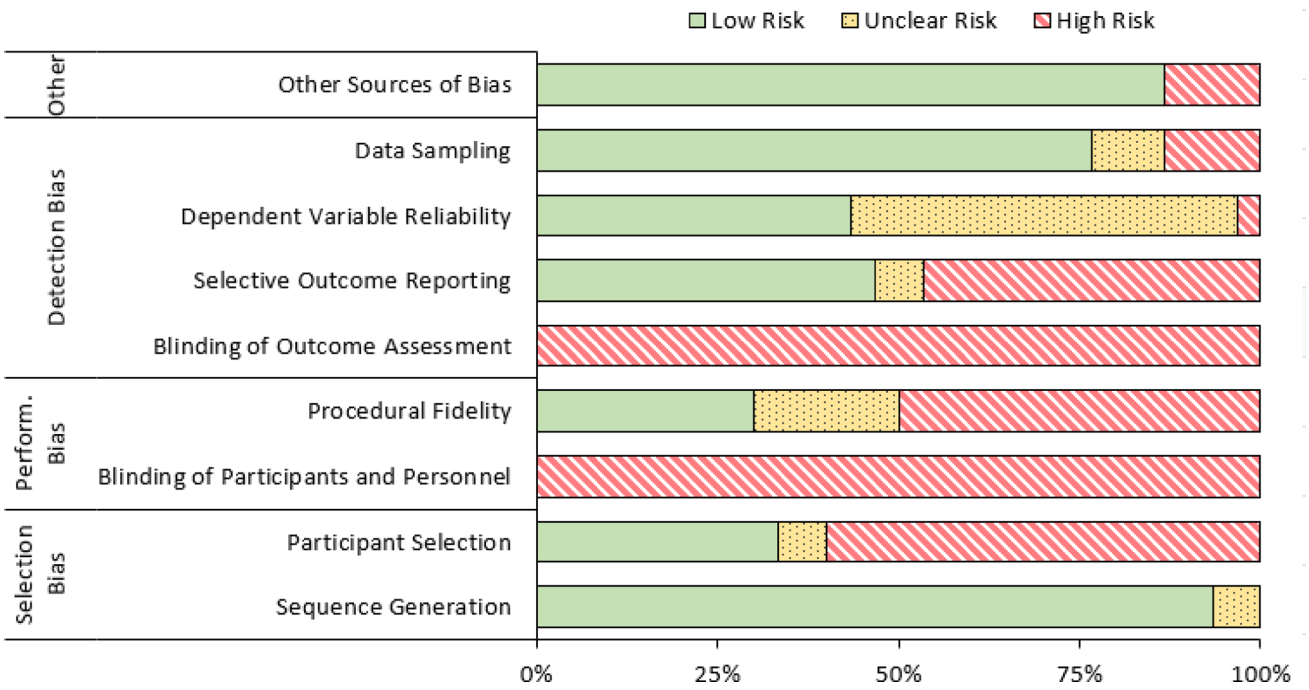

Fig. 1 Risk of bias across the reviewed literature. Perform = performance

shaped FCRs, we identified no cases where omnibus mands were shaped into specific mands. Researchers should continue to explore how the choice of omnibus or specified mands as well as the timing of shaping impacts the efficacy of interventions and long-term outcomes. Clinicians should carefully consider these matters when selecting and shaping FCRs to balance fast and functional skill acquisition, as is common practice when implementing Functional Communication training (Corr et al., 2025; Ghaemmaghami et al., 2021; Tiger et al., 2008; Weber et al., 2024).

Implementers also frequently varied in their application of schedule thinning. Most applications employed a response-based criterion for accessing reinforcement, with only 20 using purely time-based criteria. This is congruent with research indicating that response-based schedule thinning produces better tolerance for delayed reinforcement and lower rates of challenging behaviour (Drifke et al., 2020; Ghaemmaghami et al., 2016a, 2016b). When determining whether to progress to the next phase of schedule thinning, most studies combined (a) their criterion for accessing reinforcement with (b) a criterion for challenging behaviour and (c) the number of trials/sessions across which these criteria needed to be met. For example, Lundy et al. (2021) progressed the delay when the participant achieved (a) 100% independent responding at each response requirement level with (b) no challenging behaviour (c) across two consecutive sessions. For clinicians, these variations highlight the need for individualised support. However, studies give little-to-no guidance on why criteria were chosen or how these influenced learners' progress. Considering the research into schedule thinning that is already available to support practitioner decision making (Hagopian et al., 2011; Kranak & Brown, 2024), future studies should explore these variations in more detail as they relate to SBT. Valuable areas of research may be to explore which schedules are most effective, how quickly they should be thinned, which combinations of criteria best suit different learner profiles, and how different schedules potentially influence acquisition, progress, and engagement.

Applications often varied in ways that attempted to reduce the potential aversiveness of procedures and increase safety. For example, applications varied in their use of full extinction, extinction with prompting of replacement behaviour, extinction with blocking of challenging behaviour, and some applications achieved a reduction in challenging behaviour without using extinction at all (e.g., Rajaraman et al., 2024; Staubitz et al., 2022). In some cases, extinction was applied only to nondangerous challenging behaviour (Fruchtman et al., 2025) or challenging behaviour was differentially reinforced (e.g., received a lesser quality of reinforcement than communication responses; Boyle et al., 2020). Further, the majority of applications reinforced precursors or less severe behaviours. By limiting the application of extinction and reinforcing less severe behaviours, participants are less likely to be exposed to aversive conditions and may be less likely to engage in more harmful behaviours (Warner et al., 2020). Such considerations align with current movements toward trauma-assumed and compassionate care,

assent-based practices, and procedures such as "kind extinction" (e.g., Breaux & Smith, 2023; Leyman et al., 2025; Tarbox et al., 2023). It would be remiss not to highlight that some studies reset the reinforcement criteria in schedule thinning which can be conceptualised as response cost and could be aversive to participants. However, there were no data to demonstrate an increase in challenging behaviour when this occurred. Clinicians implementing SBT should consider carefully how they will respond to challenging behaviour to achieve effective outcomes while minimising aversiveness for the learner. Researchers could further explore how strategies to reduce distress and discomfort for participants impacts outcomes in SBT.

Limitations

There were several limitations to this meta-analysis. First, only one grey literature database was able to be searched, which may mean other unpublished studies were missed which could have reduced the impact of any publication bias. We did not formally assess risk of publication bias, which is generally considered best practice in meta-analysis for group designs (Page et al., 2018, 2024). Publication bias is rarely assessed in single case design meta-analyses and there is a lack of consensus on how to do so (Dowdy et al., 2022; Vannest et al., 2018). While emerging evidence suggests that publication bias in single-case research may be relatively modest, its potential influence cannot be ruled out (Dowdy et al., 2020). Grey literature was included in this review with a view to reducing the risk from publication bias. Good outcomes were observed in the three unpublished dissertations that were included, and the participant from Coffey (2019) that was not in the published version showed a moderate Tau-BC effect. To address this limitation further, future reviews could examine publication status as potential moderator by comparing effect sizes between published and grey literature (Shadish et al., 2014).

Second, it is possible that our inclusion criteria may have led to the exclusion of relevant data (e.g., participants who did not progress to schedule thinning). Additionally, our study included variations on the SBT procedure that may have differential outcomes (Rajaraman et al., 2022a, 2022b). An avenue for further exploration may be to conduct a review with a slightly wider definition of SBT and then conduct a moderator analysis to explore the influence of each component on the outcomes (Dowdy et al., 2021). This would allow the inclusion and comparison of applications in some studies that were excluded from the present analysis.

A third limitation relates to the use of Tau-BC to measure intervention effects. Tau-BC quantifies only an absence of overlap between phases and not the magnitude of effect (Carter, 2013; Moeyaert et al., 2018). Additional measures of level change could have been incorporated to quantify the magnitude of the intervention effect (e.g., Landman et al., 2024). Furthermore, there was no aggregation between cases to produce study-level or overall effect sizes, as is common practice in group design meta-analysis (Moeyaert et al., 2018). Tau-BCs creator advocates computing omnibus effect sizes with Tau-BC (Tarlow, 2017); however, others advise against using overlap indices for such aggregations (Moeyaert et al., 2018). We did not calculate effect sizes across cases due to the within-case nature of single case designs. Nevertheless, participant-level effect sizes require more effort for readers to interpret (though notably less than visual analysis of individual participant graphs; Dowdy et al., 2021).

Conclusion

Taken together, the findings of this review show emerging evidence of the efficacy of SBT. There is a need for further high-quality studies evidencing the effectiveness of SBT, and there are some areas and populations for which evidence is particularly lacking. Researchers should consider carefully the recommendations for improving quality assessment scores when conducting further research. The results of this review suggest that practitioners may consider SBT as a potentially evidence-based practice. However, the status of the research on the procedure on its own does not indicate for its use in a particular clinical case. Evidence-based practice for a scientist practitioner consists of using clinical judgement to select the procedural elements and variations that align with the best research evidence while prioritising the needs and preferences of the individual (Contreras et al., 2022; Peterson et al., 2016). Thus, practitioners should consider SBT as a general template, considering the evidence for a particular population or context, and then tailor the intervention to the specific needs of their clients.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10803-026-07547-2.

Acknowledgments Thank you to Emma Mires-Richards for advising on the search strategy. Thank you also to Dr Adithyan Rajaraman, Dr Johanna Staubitz, and Dr Joshua Jessel for their responses to our questions around SBT and SBT research.

Author Contributions Conceptualisation: NB, CF, AV. Data collection: NB, CF. Data analysis: NB. Writing – original draft: NB. Writing – review & editing: NB, CF, AV.

Funding No funding was received in connection with this work.

Data Availability Additional data are available through the Figshare open-access data repository at https://doi.org/10.6084/m9.figshare.c.8

735842.

Declarations

Conflict of interest The first author declares no conflicts of interest. The second author declares an interest as an independent practitioner who implements and delivers training courses on skill-based treatment. The third author declares no conflict of interest.

Ethical Approval This study did not involve human participants and therefore no ethical approval was required.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Alresheed, F., & Machalicek, W. (2023). A systematic comparison of non-overlap metrics and visual analysis in single-case experimental design. *Evidence-Based Communication Assessment and Intervention, 17*(1), 59–77. https://doi.org/10.1080/17489539.2023.2200813

Baiden, K. M. P., Williams, Z. J., Schuck, R. K., Dwyer, P., & Wang, M. (2025). The social validity of behavioral interventions: Seeking input from autistic adults. *Journal of Autism and Developmental Disorders, 55*(4), 1172–1186. https://doi.org/10.1007/s10803-024-06297-3

Barton, E. E., Meadan-Kaplansky, H., & Ledford, J. R. (2018). Independent variables, fidelity, and social validity. In J. R. Ledford & D. L. Gast (Eds.), *Single case research methodology: Applications in special eduction and behavioral sciences* (3rd ed.). Routledge.

Beukelman, D. R., & Light, J. C. (2020). *Augmentative and alternative communication: Supporting children and adults with complex needs* (5th ed.). Paul H. Brookes.

Boyle, M. A., Bacon, M. T., Brewer, E. M., Carton, S. M., & Gaskill, L. A. (2020). Evaluating a treatment without extinction for elopement maintained by access to stereotypy. *Journal of Applied Behavior Analysis, 53*(3), 1531–1541. https://doi.org/10.1002/jaba.682

Breaux, C. A., & Smith, K. (2023). Assent in applied behaviour analysis and positive behaviour support: Ethical considerations and practical recommendations. *International Jouranl of Developmental Disabilities, 69*(1), 111–121. https://doi.org/10.1080/20473869.2022.2144969

British Library. (2026, 27 January). *What's currently available*. Retrieved 8 February 2026 from https://bl.libguides.com/currently-available

Burns, M. K. (2012). Meta-analysis of single-case design research: Introduction to the special issue. *Journal of Behavioral Education, 21*(3), 175–184. https://doi.org/10.1007/s10864-012-9158-9

Carr, E. G., & Durand, V. M. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis, 18*(2), 111–126. https://doi.org/10.1901/jaba.1985.18-111

Carter, M. (2013). Reconsidering overlap-based measures for quantitative synthesis of single-subject data: What they tell us and what they don't. *Behavior Modification, 37*(3), 378–390. https://doi.org/10.1177/0145445513476609

Cervi, R., Gruber, D. J., & Napolitano, D. A. (2024). Implementing skill-based treatment within a classroom setting for an adolescent with autism. *Behavioral Interventions, 39*(2), Article e1989. https://doi.org/10.1002/bin.1989

Chan, M. M. Y. (2021). *How to train school staff efficiently using a behavioural skill training package: A replication of the practical functional assessment model to inform skills-based treatment in primary school children with autism spectrum disorder* [Unpublished master's thesis]. University of Aukland.

Chezan, L. C., Wolfe, K., & Drasgow, E. (2018). A meta-analysis of functional communication training effects on problem behavior and alternative communicative responses. *Focus on Autism and Other Developmental Disabilities, 33*(4), 195–205. https://doi.org/10.1177/1088357617741294

Coffey, A. L. (2019). *The interview-informed synthesized functional analysis (IISCA): A replication and evaluation of function-based treatment* [Doctoral dissertation]. Endicott College. https://www.proquest.com/openview/b2eb3436e0bbf3fb23702c2b4f56dc85/1

Coffey, A. L., Shawler, L. A., Jessel, J., Bain, T., Nye, M., & Dorsey, M. F. (2021). Generality of the practical functional assessment and skill-based treatment among individuals with autism and mental health disorders. *Behavioral Interventions, 36*(1), 298–314. https://doi.org/10.1002/bin.1755

Contreras, B. P., Hoffmann, A. N., & Slocum, T. A. (2022). Ethical behavior analysis: Evidence-based practice as a framework for ethical decision making. *Behavior Analysis in Practice, 15*(2), 619–634. https://doi.org/10.1007/s40617-021-00658-5

Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., & Test, D. W. (2015). CEC's standards for classifying the evidence base of practices in special education. *Remedial and Special Education, 36*(4), 220–234. https://doi.org/10.1177/0741932514557271

Corr, F., Rispoli, M., & Welker, N. P. (2025). A mega-review of functional communication training for students with disabilities in educational settings. *Journal of Behavioral Education*. https://doi.org/10.1007/s10864-025-09598-4

Cowell, R., Vostanis, A., & Langdon, P. E. (2024). Increasing face mask wearing in autistic individuals using behavior analytic interventions: A systematic review and meta-analysis. *Journal of Autism and Developmental Disorders, 54*(11), 4061–4080. https://doi.org/10.1007/s10803-023-06128-x

Dowdy, A., Hantula, D. A., Travers, J. C., & Tincani, M. (2022). Meta-analytic methods to detect publication bias in behavior science research. *Perspectives on Behavior Science, 45*(1), 37–52. https://doi.org/10.1007/s40614-021-00303-0

Dowdy, A., Peltier, C., Tincani, M., Schneider, W. J., Hantula, D. A., & Travers, J. C. (2021). Meta-analyses and effect sizes in applied behavior analysis: A review and discussion. *Journal of Applied Behavior Analysis, 54*(4), 1317–1340. https://doi.org/10.1002/jaba.862

Dowdy, A., Tincani, M., & Schneider, W. J. (2020). Evaluation of publication bias in response interruption and redirection: A meta-analysis. *Journal of Applied Behavior Analysis, 53*(4), 2151–2171. https://doi.org/10.1002/jaba.724

Drevon, D., Fursa, S. R., & Malcolm, A. L. (2017). Intercoder reliability and validity of webplotdigitizer in extracting graphed data. *Behavior Modification, 41*(2), 323–339. https://doi.org/10.1177/0145445516673998

Drifke, M. A., Tiger, J. H., & Lillie, M. A. (2020). DRA contingencies promote improved tolerance to delayed reinforcement during FCT compared to DRO and fixed-time schedules. *Journal of Applied Behavior Analysis, 53*(3), 1579–1592. https://doi.org/10.1002/jaba.704

Edelstein, M. L., Becraft, J. L., Gould, K., & Sullivan, A. (2021). Evaluation of a delay and denial tolerance program to increase appropriate waiting trained via telehealth. *Behavioral Interventions, 37*(2), 383–396. https://doi.org/10.1002/bin.1855

Emerson, E., Blacher, J., Einfeld, S., Hatton, C., Robertson, J., & Stancliffe, R. J. (2014). Environmental risk factors associated with the persistence of conduct difficulties in children with intellectual disabilities and autistic spectrum disorders. *Research in Developmental Disabilities, 35*(12), 3508–3517.

Ferguson, J. L., Leaf, J. A., Cihon, J. H., Milne, C. M., Leaf, J. B., McEachin, J., & Leaf, R. (2020). Practical functional assessment: A case study replication and extension with a child diagnosed with autism spectrum disorder. *Education and Treatment of Children, 43*(2), 171–185. https://doi.org/10.1007/s43494-020-00015-1

Fiani, T., & Jessel, J. (2022). Practical functional assessment and behavioral treatment of challenging behavior for clinically based outpatient services: A consecutive case series evaluation. *Education and Treatment of Children, 45*(2), 211–230. https://doi.org/10.1007/s43494-022-00071-9

Fruchtman, T., Jessel, J., Pan, B., McLeod, S., & Rajaraman, A. (2025). The performance-based IISCA can inform effective and socially meaningful skill-based treatment. *Behavior Analysis in Practice, 18*, 921–940. https://doi.org/10.1007/s40617-024-01036-7

Gerow, S., Hagan-Burke, S., Rispoli, M., Gregori, E., Mason, R., & Ninci, J. (2018). A systematic review of parent-implemented functional communication training for children with ASD. *Behavior Modification, 42*(3), 335–363. https://doi.org/10.1177/0145445517740872

Ghaemmaghami, M., Hanley, G. P., & Jessel, J. (2016a). Contingencies promote delay tolerance. *Journal of Applied Behavior Analysis, 49*(3), 548–575. https://doi.org/10.1002/jaba.333

Ghaemmaghami, M., Hanley, G. P., & Jessel, J. (2021). Functional communication training: From efficacy to effectiveness. *Journal of Applied Behavior Analysis, 54*(1), 122–143. https://doi.org/10.1002/jaba.762

Ghaemmaghami, M., Hanley, G. P., Jessel, J., & Landa, R. (2018). Shaping complex functional communication responses. *Journal of Applied Behavior Analysis, 51*(3), 502–520. https://doi.org/10.1002/jaba.468

Ghaemmaghami, M., Hanley, G. P., Jin, S. C., & Vanselow, N. R. (2016b). Affirming control by multiple reinforcers via progressive treatment analysis. *Behavioral Interventions, 31*(1), 70–86. https://doi.org/10.1002/bin.1425

Gore, N. J., McGill, P., Toogood, S., Allen, D., Hughes, J. C., Baker, P., Hastings, R. P., Noone, S. J., & Denne, L. D. (2013). Definition and scope for positive behavioural support. *International Journal of Positive Behavioural Support, 3*(2), 14–23. https://www.ingentaconnect.com/content/bild/ijpbs/2013/00000003/00000002/art00003

Guinness, K. E., Atkinson, R. S., & Feil, E. G. (2024). Evaluating social validity to inform intervention development: Qualitative analysis of caregiver interviews. *Behavior Analysis in Practice, 17*(3), 870–879. https://doi.org/10.1007/s40617-023-00899-6

Hagopian, L. P., Boelter, E. W., & Jarmolowicz, D. P. (2011). Reinforcement schedule thinning following functional communication training: Review and recommendations. *Behavior Analysis in Practice, 4*(1), 4–16. https://doi.org/10.1007/bf03391770

Hanley, G. P. (2010). Toward effective and preferred programming: A case for the objective measurement of social validity with recipients of behavior-change programs. *Behavior Analysis in Practice, 3*(1), 13–21. https://doi.org/10.1007/bf03391754

Hanley, G. P. (2012). Functional assessment of problem behavior: Dispelling myths, overcoming implementation obstacles, and developing new lore. *Behavior Analysis in Practice, 5*(1), 54–72. https://doi.org/10.1007/bf03391818

Hanley, G. P., Jin, C. S., Vanselow, N. R., & Hanratty, L. A. (2014). Producing meaningful improvements in problem behavior of children with autism via synthesized analyses and treatments. *Journal of Applied Behavior Analysis, 47*(1), 16–36. https://doi.org/10.1002/jaba.106

Heath, A. K., Ganz, J. B., Parker, R., Burke, M., & Ninci, J. (2015). A meta-analytic review of functional communication training across mode of communication, age, and disability. *Review Journal of Autism and Developmental Disorders, 2*(2), 155–166. https://doi.org/10.1007/s40489-014-0044-3

Higgins, J. P. T., Sterne, J. A. C., Savović, J., Page, M. J., Hróbjartsson, A., Boutron, I., Reeves, B., & Eldridge, S. (2016). A revised tool for assessing risk of bias in randomized trials. In J. Chandler, J. McKenzie, I. Boutron, & V. Welch (Eds.), *Cochrane Methods*. Cochrane Database of Systematic Reviews. https://doi.org/10.1002/14651858.CD201601

Holden, B., & Gitlesen, J. (2006). A total population study of challenging behaviour in the county of Hedmark, Norway: Prevalence, and risk markers. *Research in Developmental Disabilities, 27*(4), 456–465. https://doi.org/10.1016/j.ridd.2005.06.001

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165–179. https://doi.org/10.1177/001440290507100203

Hustyi, K. M., Logue, J. J., & Hall, S. S. (2025). Skill-based treatment for challenging behavior in autism spectrum disorder: A scoping review of treatment characteristics and outcomes. *Research in Autism Spectrum Disorders, 119*, Article 102523. https://doi.org/10.1016/j.rasd.2024.102523

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*(2), 197–209. https://doi.org/10.1901/jaba.1994.27-197

Jessel, J., Ingvarsson, E. T., Metras, R., Kirk, H., & Whipple, R. (2018). Achieving socially significant reductions in problem behavior following the interview-informed synthesized contingency analysis: A summary of 25 outpatient applications. *Journal of Applied Behavior Analysis, 51*(1), 130–157. https://doi.org/10.1002/jaba.436

Jessel, J., Rosenthal, D., Hanley, G. P., Rymill, L., Boucher, M. B., Howard, M., Perrin, J., & Lemos, F. M. (2022). On the occurrence of dangerous problem behavior during functional analysis: An evaluation of 30 applications. *Behavior Modification*. https://doi.org/10.1177/01454455211010698

Kastner, K. M. (2025). An evaluation of probabilistic delay and denial training: Durability against treatment relapse. *Journal of Applied Behavior Analysis, 58*(2), 344–362. https://doi.org/10.1002/jaba.70007

Kranak, M. P., & Brown, K. R. (2024). Updated recommendations for reinforcement schedule thinning following functional communication training. *Behavior Analysis in Practice, 17*(1), 87–106. https://doi.org/10.1007/s40617-023-00863-4

Landman, W., Bogaerts, S., & Spreen, M. (2024). Typicality of level change (TLC) as an additional effect measure to NAP and Tau-U in single case research. *Behavior Modification, 48*(1), 51–74. https://doi.org/10.1177/01454455231190741

Layman, L. N., Dufrene, B. A., Ackley, M. M., Weaver, C. M., Schneider, D. E., LaBrot, Z. C., Taylor, C. N., Rahaman, J. A., Tawney, K. N., Hart, T., & Olmi, D. J. (2023). Interview-informed synthesized contingency analyses on challenging problem behavior: A single-case meta-analysis. *Review Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s40489-023-00357-7

Leif, E. S., Kelenc-Gasior, N., Bloomfield, B. S., Furlonger, B., & Fox, R. A. (2024). A systematic review of social-validity assessments in the Journal of Applied Behavior Analysis: 2010–2020. *Journal of Applied Behavior Analysis, 57*(3), 542–559. https://doi.org/10.1002/jaba.1092

Leyman, A., MacDowell, P., & Jessel, J. (2025). *Evaluating the long-term efficacy of a trauma-informed approach to addressing challenging behavior in the home*. Advance online publication. https://doi.org/10.1177/15407969251319

Luczynski, K. C., & Hanley, G. P. (2013). Prevention of problem behavior by teaching functional communication and self-control skills to preschoolers. *Journal of Applied Behavior Analysis, 46*(2), 355–368. https://doi.org/10.1002/jaba.44

Lundy, E., Healy, O., Ramey, D., Carolan, T., Dempsey, R., & Holloway, J. (2021). Evaluating the utility of interview-informed synthesized contingency analyses in informing the treatment of problem behavior among children with autism spectrum disorder. *European Journal of Behavior Analysis*. https://doi.org/10.1080/15021149.2021.1981752

Melanson, I. J., & Fahmie, T. A. (2023). Functional analysis of problem behavior: A 40-year review. *Journal of Applied Behavior Analysis, 56*(2), 262–281. https://doi.org/10.1002/jaba.983

Metras, R. L., Hanley, G. P., & Carbone, M. J. (2024). Distance-based collaborations for assessing and treating challenging behavior. *Journal of Autism and Developmental Disorders, 54*(10), 3587–3604. https://doi.org/10.1007/s10803-023-06085-5

Metras, R., & Jessel, J. (2021). Adaptations of the interview-informed synthesized contingency analysis. *Journal of Applied Behavior Analysis, 54*(3), 877–881. https://doi.org/10.1002/jaba.849

Moeyaert, M., Zimmerman, K. N., & Ledford, J. R. (2018). Synthesis and meta-analysis of single case research. In J. R. Ledford & D. L. Gast (Eds.), *Single case research methodology* (3rd ed.). Routledge.

Muharib, R., Alrasheed, F., Ninci, J., Walker, V. L., & Voggt, A. P. (2019). Thinning schedules of reinforcement following functional communication training for children with intellectual and developmental disabilities: A meta-analytic review. *Journal of Autism and Developmental Disorders, 49*(12), 4788–4806. https://doi.org/10.1007/s10803-019-04191-x

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews, 5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M. J., Higgins, J. P. T., & Sterne, J. A. C. (2024). Chapter 13: Assessing risk of bias due to missing evidence in a meta-analysis. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions*. Cochrane. https://training.cochrane.org/handbook

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews, 10*(1), Article 89. https://doi.org/10.1186/s13643-021-01626-4

Page, M. J., McKenzie, J. E., & Higgins, J. P. T. (2018). Tools for assessing risk of reporting biases in studies and syntheses of studies: A systematic review. *British Medical Journal Open, 8*(3), Article e019703. https://doi.org/10.1136/bmjopen-2017-019703

Peterson, C. B., Becker, C. B., Treasure, J., Shafran, R., & Bryant-Waugh, R. (2016). The three-legged stool of evidence-based practice in eating disorder treatment: Research, clinical, and patient perspectives. *BMC Medicine, 14*(1), 69. https://doi.org/10.1186/s12916-016-0615-5

Pustejovsky, J. E., Chen, M., Grekov, P., & Swan, D. M. (2024). *SingleCaseES: A calculator for single-case effect size indices* (Version 0.7.3) [R Package]. https://jepusto.github.io/SingleCaseES/

Quiroz, L. S., Boyd, J. A., Pollard, J. S., & Hall, S. S. (2024). Evaluating practical functional assessment and skill-based treatment for severe challenging behavior via telehealth with novice dual parent implementers: A pilot investigation. *Clinical Case Studies, 24*(1), 59–77. https://doi.org/10.1177/15346501241297363

R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.3.2). R Foundation for Statistical Computing. https://www.R-project.org

Rajaraman, A., Hanley, G. P., Gover, H. C., Ruppel, K. W., & Landa, R. K. (2022a). On the reliability and treatment utility of the practical functional assessment process. *Behavior Analysis in Practice, 15*(3), 815–837. https://doi.org/10.1007/s40617-021-00665-6

Rajaraman, A., Hanley, G. P., Gover, H. C., Staubitz, J. L., Staubitz, J. E., Simcoe, K. M., & Metras, R. (2022b). Minimizing escalation by treating dangerous problem behavior within an enhanced choice model. *Behavior Analysis in Practice, 15*(1), 219–242. https://doi.org/10.1007/s40617-020-00548-2

Rajaraman, A., Whelan, C. J., Jessel, J., & Gover, H. C. (2024). Promoting safety while addressing dangerous behavior via telehealth: A clinical case investigation serving the family of an autistic adolescent living in India. *Clinical Case Studies*. https://doi.org/10.1177/15346501241243103

Reichow, B. (2011). Development, procedures, and application of the evaluative method for determining evidence-based practices in autism. *Evidence-based practices and treatments for children with autism* (pp. 25–39). Springer US. https://doi.org/10.1007/978-1-4419-6975-0_2

Reichow, B., Barton, E. E., & Maggin, D. M. (2018). Development and applications of the single-case design risk of bias tool for evaluating single-case design research study reports. *Research in Developmental Disabilities, 79*, 53–64. https://doi.org/10.1016/j.ridd.2018.05.008

Reichow, B., Volkmar, F. R., & Cicchetti, D. V. (2008). Development of the evaluative method for evaluating and determining evidence-based practices in autism. *Journal of Autism and Developmental Disorders, 38*(7), 1311–1319. https://doi.org/10.1007/s10803-007-0517-7

Rohatgi, A. (2021). *WebPlotDigitizer* (Version 4.5). https://automeris.io/WebPlotDigitizer/

Rose, J. C., & Beaulieu, L. (2019). Assessing the generality and durability of interview-informed functional analyses and treatment. *Journal of Applied Behavior Analysis, 52*(1), 271–285. https://doi.org/10.1002/jaba.504

Santa Maria, K. (2025). *The impact of varying treatment integrity levels on skill-based treatment outcomes [Doctoral dissertation]*. Capella University.

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology, 52*(2), 123–147. https://doi.org/10.1016/j.jsp.2013.11.005

Staubitz, J. L., Staubitz, J. E., Pollack, M. S., Haws, R. A., & Hopton, M. (2022). Effects of an enhanced choice model of skill-based treatment for students with emotional/behavioral disorders. *Journal of Applied Behavior Analysis, 55*(4), 1306–1341. https://doi.org/10.1002/jaba.952

Strand, R. C. W., & Eldevik, S. (2017). Improvements in problem behavior in a child with autism spectrum diagnosis through synthesized analysis and treatment: A replication in an EIBI home program. *Behavioral Interventions, 33*(1), 102–111. https://doi.org/10.1002/bin.1505

Tarbox, C., Tarbox, J., Bermudez, T. L., Silverman, E., & Servellon, L. (2023). Kind extinction: A procedural variation on traditional

extinction. *Behavior Analysis in Practice*. https://doi.org/10.1007/s40617-023-00833-w

Tarlow, K. R. (2017). An improved rank correlation effect size statistic for single-case designs: Baseline corrected tau. *Behavior Modification, 41*(4), 427–467. https://doi.org/10.1177/0145445516676750

Taylor, S. A., Phillips, K. J., & Gertzog, M. G. (2018). Use of synthesized analysis and informed treatment to promote school reintegration. *Behavioral Interventions, 33*(4), 364–379. https://doi.org/10.1002/bin.1640

Tiger, J. H., Hanley, G. P., & Bruzek, J. (2008). Functional communication training: A review and practical guide. *Behavior Analysis in Practice, 1*(1), 16–23. https://doi.org/10.1007/bf03391716

Tomlinson, S. R. L., Gore, N., & Mcgill, P. (2018). Training individuals to implement applied behavior analytic procedures via telehealth: A systematic review of the literature. *Journal of Behavioral Education, 27*(2), 172–222. https://doi.org/10.1007/s10864-018-9292-0

Vannest, K. J., Peltier, C., & Haas, A. (2018). Results reporting in single case experiments and single case meta-analysis. *Research in Developmental Disabilities, 79*, 10–18. https://doi.org/10.1016/j.ridd.2018.04.029

Vollmer, T. R., & Pendergrass, J. A. (2025). Social validity and contemporary applied behavior analysis. *Perspectives on Behavior Science, 48*(3), 519–527. https://doi.org/10.1007/s40614-025-00452-6

Ward, S. N., Hanley, G. P., Warner, C. A., & Gage, E. E. (2021). Does teaching an omnibus mand preclude the development of specifying mands? *Journal of Applied Behavior Analysis, 54*(1), 248–269. https://doi.org/10.1002/jaba.784

Warner, C. A., Hanley, G. P., Landa, R. K., Ruppel, K. W., Rajaraman, A., Ghaemmaghami, M., Slaton, J. D., & Gover, H. C. (2020). Toward accurate inferences of response class membership. *Journal of Applied Behavior Analysis, 53*(1), 331–354. https://doi.org/10.1002/jaba.598

Weber, J., Fahmie, T., Walker, S., Lambert, J., Copeland, B., Freetly, T., & Zangrillo, A. (2024). Exploring factors that influence the efficacy of functional communication training. *Journal of Applied Behavior Analysis, 57*(3), 709–724. https://doi.org/10.1002/jaba.1078

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*(2), 203–214. https://doi.org/10.1901/jaba.1978.11-203

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.