

Meta-analysis of single-case treatment effects on self-injurious behavior for individuals with autism and intellectual disabilities

Stephanie Morano and Salvador Ruiz

Pennsylvania State University, USA

Jiwon Hwang

California State University, Bakersfield

Jennifer L. Wertalik, Jeremy Moeller and Muhammed Akif Karal

Pennsylvania State University, USA

Austin Mulloy

University of California, USA

Autism & Developmental Language Impairments

Volume 2: 1–26

© The Author(s) 2017

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/2396941516688399

journals.sagepub.com/home/dli



Abstract

Background and aim: Research has demonstrated that a variety of treatments can reduce or eliminate self-injurious behavior (SIB) in individuals with autism and/or intellectual disabilities but evidence suggests that not all treatments are equally effective.

Methods: We used multi-level meta-analysis to synthesize the results of 137 single-case design studies on SIB treatment for 245 individuals with autism and/or intellectual disabilities. Analyses compare the effects of various behavioral and medical treatments for SIB and assess associations between treatment effects and participant- and study-level variables.

Results: Findings suggest differential reinforcement, punishment, and treatment packages with reinforcement and punishment components resulted in the largest SIB reductions.

Conclusions: Results indicate that overall, treatment for SIB is highly effective and that participant and study characteristics do not moderate treatment effects.

Implications: Based on results and in line with current practice recommendations, we encourage use of reinforcement-based procedures in all cases of SIB. In the event that reinforcement-only treatments have failed or if SIB poses a serious, immediate threat to the health and well-being of an individual, our results suggest that overcorrection paired with reinforcement may be the most effective as well as less invasive alternative.

Keywords

Self-injurious behavior, autism, intellectual disabilities, multi-level meta-analysis, single-case research designs

Introduction

Self-injurious behavior (SIB) is a broad term encompassing behaviors that cause unintentional, self-inflicted, socially unacceptable physical injury to the individual's own body (Yates, 2004). Examples of SIB topographies include head hitting, hand mouthing, hair

pulling, eye gouging, and hitting self with objects (e.g., Matson & LoVullo, 2008). In addition to causing physical injury, SIB can have serious social repercussions. Distress among caregivers and severely limited community acceptance and participation are commonly associated with SIB (Fisher, Piazza, Bowman, Hanley, & Adelinis, 1997; Schalock, 2004; Tate & Baroff, 1966).

Corresponding author:

Stephanie Morano, Pennsylvania State University, 220 Cedar Building, University Park, State College, PA 16802-1503, USA.

Email: s.m.morano@gmail.com



Creative Commons Attribution-NonCommercial 3.0 License (<http://www.creativecommons.org/licenses/by-nc/3.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

The prevalence of SIB is estimated at 50% for people diagnosed with an autism spectrum disorder (ASD) and in the range of 4%–24% for people diagnosed with intellectual disability (ID; Baghdadli, Pascal, Grisi, & Aussilloux, 2003; Prangnell, 2009; Richards, Oliver, Nelson, & Moss, 2012). Furthermore, nearly half (47%) of all individuals with SIB exhibit these behaviors for 10 years or more (Totsika, Toogood, Hastings, & Lewis, 2008). The persistence of SIB over time, high prevalence rates, and the potential for severe physical and social consequences necessitate the identification of efficient, effective, and lasting treatments for SIB.

Theoretical frameworks for emergence and treatment of SIB

The fields of medicine and behavior analysis offer differing theoretical frameworks for the emergence of SIB. The medical perspective emphasizes biochemical processes internal to individuals, while behavior analytic theories focus on relations between environmental conditions and individuals' behaviors.

Authors from the field of medicine have hypothesized that engaging in SIB precipitates the release of endorphins, hormones, or neurotransmitters, which act to reinforce and maintain the behavior (Garcia & Smith, 1999). Treatments for SIB based on these hypotheses involve administering psychotropic drugs that interrupt key biochemical processes. For example, the drug naltrexone blocks opioid receptors and is used in an effort to prevent experience of a potentially reinforcing "endorphin rush" during SIB episodes (e.g., Sandman, Barron, & Colman, 1990). Anti-depressant medications (e.g., Sertraline and Paroxetine; Hellings, Kelley, Gabrielli, Kilgore, & Shah, 1996; Davanzo, Belin, Widawski, & Bryan, 1997) and serotonin-rich foods (e.g., bananas, Gedy, 1990) have been administered to alter individuals' baseline state and may render SIB and its biochemical effects of less interest or reinforcing value. Further, antipsychotic drugs have been used to suppress motor activity and thereby limit incidence of SIB (e.g., Risperidone and Clozapine; Cohen, Ihrig, Lott & Kerrick, 1998; Hammock, Schroeder, & Levine, 1995).

Alternatively, authors from the field of behavior analysis have argued that, while the physical sensation a person experiences during SIB could reinforce the behavior, SIB could additionally function as a form of communication and/or means to solicit behavior in others (Sigafos, Arthur, & O'Reilly, 2003). For example, a person may engage in SIB to obtain attention from a caregiver, gain access to tangible items (e.g., food and toys), or avoid participation in a non-preferred activity. In this framework, people with ASD and/or ID, who often have limited proficiency

in communication and/or mobility with which to manipulate their environment, understandably resort to challenging behaviors such as SIB to express their preferences and needs. Conversely, the ability to effectively communicate requests and refusals with speech or alternative communication methods may serve as a protective factor against SIB (Schroeder, Schroeder, Smith, & Dalldorf, 1978).

Landmark studies conducted by Iwata, Pace, Cowdery, and Miltenberger (1982/1994) established functional analysis (FA) as a reliable, experimental method for identifying the function(s) of individuals' SIB. In the years since, behavior analysts have embraced FA as the gold standard for determining why an individual exhibits a behavior and consider the method an essential step in selecting appropriate and maximally effective interventions (Cooper, Heron, & Heward, 2007). Behavior analytic treatments for SIB generally involve manipulation of environmental conditions and teaching of skills. For example, in cases in which FA identifies the function of SIB to obtain a particular toy or form of attention, interventions may include reinforcement of alternative communicative behaviors with provision of the particular food or form of attention (e.g., Hanley, Piazza, Fisher, & Adelinis, 1997). For SIB with an avoidance function, interventionists may teach individuals new ways to request or initiate release from non-preferred stimuli, and then provide structured opportunities to reinforce the new behaviors with release (e.g., Kahng, Iwata, DeLeon, & Worsdell, 1997). Punishments for SIB are also common methods of intervention that are intended to break associations between SIB and its reinforcing consequences. Following instances of SIB, aversive stimuli may be applied (e.g., ammonia tablet broken under nose; e.g., Singh, Dawson, & Gregory, 1980) or reinforcers may be withheld (e.g., attention or toy withdrawn, e.g., Lucero, Frieman, Spoering, & Fehrenbacher, 1976; further movement restrained, e.g., Fisher et al., 1997). Typically, interventions involve multiple components, such as a punishment contingency, instruction in leisure or communication skills, or new routines, plus several reinforcement procedures.

Effects of medical treatments for SIB

Reviews of the research have reported conflicting results regarding the effectiveness of anti-depressants, opioid antagonists, and antipsychotics in the treatment of SIB. Symons, Thompson, and Rodriguez (2004) conducted a multi-level meta-analysis and narrative review of studies on naltrexone use in individuals with ID. Results indicated that 80% of subjects showed reductions in SIB during short-term treatment and that males

required lower doses and responded more favorably than females. Based on their narrative review, Mahatmya, Zobel, and Valdovinos (2008) also found that anti-depressants, opioid antagonists, and anti-psychotics are generally effective for treating SIB in individuals with ASD, but that several variables appear to moderate the effects of medication (i.e., dosage, treatment duration, sex, severity of SIB, and variables maintaining SIB). In contrast, results from Gormez, Rana, and Varghese's (2014) narrative review of pharmacological treatments for SIB in adults with ID suggested that no active drug was more effective than a placebo. Authors of these reviews and others on pharmacological treatments for challenging behaviors (e.g., Matson et al., 2000) have noted a variety of difficulties in drawing conclusions confidently regarding the effects of specific drugs, for example, individual differences in reactivity, interactions with other medications or extra-experimental changes in other medications, and, at the study level, a lack of consistency in measurement and reporting on side effects (e.g., suppression of learning or communication).

Effects of behavioral treatments for SIB

Syntheses of behavior analytic research similarly report varying results. Findings of a quantitative review of studies on behavioral treatments for SIB published between 1965 and 2002 suggest that increased use of FA over time coincides with a marked increase in the use of reinforcement-based treatments (e.g., differential reinforcement of alternative behavior) and a gradual decrease in the use of punishment procedures (Kahng, Iwata, & Lewin, 2002). Kahng et al.'s (2002) analysis of treatment effects depicted all methods as capable of substantial reductions of SIB, favored punishment and preventative methods (e.g., manipulation of antecedents, mechanical or manual blocking of SIB), and suggested combinations of treatments were slightly more effective than single component interventions.

Prangnell (2009) reviewed studies on behavioral treatments for SIB published from 1998 to 2008 and, in contrast to Kahng et al. (2002), noted a substantial proportion of studies reported the use of aversive treatments. Prangnell's (2009) synthesis findings suggested treatment effects were highly variable within particular categories/types of interventions, punishment procedures typically resulted in rapid and substantial reductions in SIB, and combinations of behavioral treatments were more effective than single treatments. Prangnell (2009) cautioned that conclusions about the long-term effects of treatments could not be drawn due to a lack of follow-up or maintenance data reported across studies.

Additionally, Denis, Van den Noortgate, and Maes (2011) conducted a multi-level meta-analysis of non-aversive and non-intrusive, reinforcement-based SIB treatment effects for participants with profound ID. Results showed that non-aversive and non-intrusive behavioral treatments resulted in large reductions in SIB. Participant sensory impairment was identified as a moderator of treatment effects, while medication, motor impairment, setting, age, gender, and matching of treatment with behavioral function were not found to account for variations in treatment effects.

Limitations of previous research syntheses on SIB treatment effects

While a number of reviews have synthesized research on effects of behavioral treatment on SIB, each limited their sample of studies to focus on a single disability category: ASD or ID. Behavior analytic tradition suggests, with regard to SIB and other forms of challenging behavior, disability condition is negligible and greatly surpassed in relevance by an individual's history of contingencies (Watson, 1913). Thus, previous syntheses' omission of large portions of the research literature on behavioral treatments may have skewed their findings.

As described above, the majority of reviews on SIB treatment have focused on either medical or behavioral treatments. Only one narrative review (Mahatmya et al., 2008) included studies on both medical and behavioral treatments for SIB in individuals with ASD. However, authors limited their sample to function-based behavioral treatments and thereby excluded several decades of research from before the establishment of functional assessment methods.

Additionally, the scope and internal validity of previous reviews' findings are limited by their synthesis methods. With the exception of the quantitative analyses conducted by Symons et al. (2004), Kahng et al. (2002), and Denis et al. (2011), each which focused on narrow selections of treatments and/or disability conditions, all other syntheses employed narrative review methods. While narrative review methods have a variety of strengths (White, 1987), they do not permit precision in assessment or comparison of treatment effects nor analysis of associations between treatment effects and potential moderators. Also, in recent years, methodological researchers have greatly advanced understandings of and techniques for meta-analysis of single-case research data (i.e., the primary research design employed in research on treatment of SIB; "Methodological Dilemmas," April 2012; "Methodological Issues," April 2013; "Handling Methodological Issues," April 2014; "Single-Case Experimental Designs," April 2015). In empirical research and theoretical

commentary, authors have identified systematic biases and inaccuracies in summary statistics used in previous syntheses (e.g., Mean Baseline Reduction, standardized difference between means), which are associated with common features of single-case research data (e.g., variability across datasets in numbers of repeated measures and/or outcome metrics, variability across observations in behavior levels; Beretvas & Chung, 2008; Pustejovsky, 2015).

Limitations in previous reviews' samples and methods warrant further meta-analytic work in the area of SIB treatments. The present review was designed to address these limitations. We sampled the entire body of literature on SIB treatment and synthesized research on all types of SIB treatment, involving individuals with ID and/or ASD, using the most current and methodologically sound quantitative procedures.

Meta-analysis of single-case research

Meta-analysis allows estimation and aggregation of treatment effects, analysis of associations between treatment effects, i.e., effect sizes (ESs), and independent variables, and calculation of confidence intervals and significance testing for treatment effects and associations (Cooper & Hedges, 1994). By synthesizing bodies of literature, meta-analysis can yield new knowledge about educational and behavioral practices (Jenson, Clark, Kircher, & Kristjansson, 2007). The practices are well established in fields of psychology and education that employ group-design research methods, although development and uptake of quantitative techniques for synthesizing results of single-case design research is still in process. Despite general agreement that meta-analysis of single-case experimental data is useful and should be undertaken, there has been considerable debate regarding the validity of various methods (e.g., Allison & Gorman, 1993; Beretvas & Chung, 2008; Ferron, 2002; Kratochwill & Levin, 2014; Salzberg, Strain, & Baer, 1987; Shadish, Rindskopf, & Hedges, 2008; Scruggs, Mastropieri, & Casto, 1987; Scruggs & Mastropieri, 2013). Criticisms primarily focus on the use of non-parametric summary statistics and methods developed for group-design research, but also pertain to methodological challenges that result from variations in standards and rigor in primary single-case research (Kratochwill & Levin, 2010; Kratochwill et al., 2012).

Non-parametric statistics and standardized mean differences. Non-parametric summary statistics (e.g., percentage of non-overlapping data (PND; Scruggs et al., 1987), percentage of zero data (PZD; Scotti, Evans, & Meyer, 1991), and mean baseline reduction (Kahng et al., 2002)) and application of the

Standardized Mean Difference (SMD; Busk & Serlin, 1992) to single-case data carry the risks of systematic bias and misrepresentation of data phenomena. For example, PND has an inverse relationship with the number of baseline data points (i.e., higher PND values are associated with fewer baseline data points; Allison & Gorman, 1994). PZD has an inverse relationship with the number of observation sessions conducted during treatment phases beyond the first observation of a zero level of behavior (i.e., longer treatment phases are associated with lower PZD scores; Beretvas & Chung, 2008). Comparison and aggregation of mean baseline reduction (MBR) statistics across studies are confounded by variations in the levels of baseline data, such that different levels of baseline data produce statistics on different metrics (i.e., 50% reduction in a rate of 10 behaviors per minute does not equate to 50% reduction in a rate of 100 behaviors per minute, in terms of magnitude of reduction or practical value of the treatment outcome). Also, low PND, PZD, and MBR values can also result from slow acquisition rates, even when treatments are ultimately effective in changing or eliminating behaviors over time (Allison & Gorman, 1994; Scotti et al., 1991; White, 1987).

Similarly, trends in data confound SMD values by introducing error in variance estimates and skewing means (Marquis, Horner, & Carr, 2000). Further, standardization of treatment effects with the variance of phase data does not logically correspond to the import of variability in behavior levels (Pustejovsky, 2015). In the tradition of behavior analysis, variability in behavior levels purely represents unreliability of measurement procedures, created by uncontrolled variables, whose moderating effects are unknown and immeasurable. Presumably, effects of uncontrolled variables vary in composition and intensity across observation sessions. In group-design research, variability in measures of individuals represents the distribution of a trait, plus some amount of error due to uncontrolled variables. While standardization of mean differences with variances scales effects from group research in terms of meaningful distributions, approximately, such standardization in single-case research scales effects in terms of error and the presence of confounds (i.e., effects are compressed or inflated according to the degree of unreliability of the measurement system).

Additional bias and misrepresentation of phenomena can result from autocorrelation (i.e., serial dependence in repeated measures) and variations in outcome metrics (e.g., frequency counts, percent of partial intervals). Autocorrelation has the potential to interfere with accurate estimation of variances (Baek & Ferron, 2013). When used as a standardization factor, inaccurate variances confound ESs. Variations in outcome

metrics across studies typically necessitate standardization of data prior to aggregation (Van den Noortgate & Onghena, 2003). Given no methods of standardization exist for non-parametric summary statistics, aggregation of data from disparate metrics yields results with very low internal validity.

Recognition of these and other methodological problems has led authors to caution against use of non-parametric summary statistics and methods developed for group-design research. While their flaws differ, each statistical method is compromised in terms of accuracy and bias in estimation and aggregation of treatment effects (e.g., Allison & Gorman, 1994) and moderator analyses (e.g., Haevaert, Saenen, Maes, & Onghena, 2015).

Multi-level modeling. Multi-level modeling (MLM) constitutes a viable alternative to non-parametric summary statistics for synthesizing single-case data. MLM estimation (a) is not biased by differences in the number of data points collected in each phase or levels of behavior across individuals, (b) can model trends in time-series data, and (c) is robust to the presence of auto-correlation (Baek & Ferron, 2013; Jenson et al., 2007; Raudenbush & Bryk, 2002; Singer & Willet, 2003; Van den Noortgate & Onghena, 2003, 2008). MLM also can identify differential effects when distributions of statistics vary or are not known, and allows for accurate moderator analyses. When used with datasets comparable in study design specifics or in conjunction with a rigorous ES summary statistic, mounting evidence suggests MLM can be a sound and useful tool for identifying evidence-based practices and generating new knowledge from research literatures (Baek & Ferron, 2013; Baek et al., 2014; Moeyaert, Ferron, Beretvas, & Van den Noortgate, 2014; Moeyaert, Ugille, Ferron, Beretvas, & Van den Noortgate, 2013; Moeyaert, Ugille, Ferron, Beretvas, & Van den Noortgate, 2014; Ugille, Moeyaert, Beretvas, Ferron, & Van den Noortgate, 2012).

Only one previous synthesis used MLM to investigate SIB treatments (i.e., the effects of non-aversive and non-intrusive reinforcement-based treatments in participants with profound ID; Denis et al., 2011). However, Denis et al.'s (2011) results may have been confounded by use of a SMD ES. The current study contributes to SIB research literature by (a) analyzing data on all types of treatment for SIB, from studies with both participants with ASD and ID; (b) making use of a more robust ES; and (c) assessing the relations between treatment effects and a variety of subject characteristics (e.g., disability diagnosis), study characteristics (e.g., duration of treatment), and intervention methods (e.g., matching treatments to SIB function, form of punishment).

The following research questions guided preliminary analyses (1), the formulation and estimation of multi-level models (2, 3, 4), and post-hoc analyses (5, 6):

1. Across categories of each participant- and study-level independent variable (e.g., disability condition, metric for measurement of SIB), are there apparent, systematic differences in ESs' distributions and means? For which variables do data justify their inclusion as explanatory variables in meta-analytic models?
2. In an unconditional model, involving aggregation of individual ESs within studies and then combination of study aggregates:
 - a. Overall, what is the expected effect of treatment on SIB level?
 - b. Across studies, do aggregate ESs differ?
 - c. Within studies, do individual ESs differ?
 - d. What proportions of the total variance occur between studies and within studies (i.e., between individual ESs in studies)?
3. In a first conditional model, including treatment type as a explanatory variable for study aggregate ESs:
 - a. For each treatment type, what is the expected effect of treatment on SIB level?
 - b. Do data suggest any treatments are more effective than others?
 - c. After controlling for treatment type, do study aggregate ESs differ? In other words, is there any variation that remains to be explained?
4. In a second conditional model, including treatment type, as well as whether interventions were matched to SIB function as explanatory variables for study aggregate ESs:
 - a. For each treatment type, what is the difference between expected effects for interventions matched to function and those not matched to function?
 - b. Do data suggest the practice of matching interventions to function is more effective than not matching interventions to function?
 - c. Does modeling whether treatments were matched to SIB function result in greater explanatory power in the model (i.e., greater precision in effect estimates, less unexplained variance in study aggregate ESs)?
5. In the subset of studies that investigated punishment procedures, do data suggest effects differ across negative punishment (i.e., withdrawal of preferred stimuli), positive punishment (i.e., application of aversive stimuli) that relates to SIB function, and positive punishment that is not known to relate SIB function?
6. In the subset of studies that investigated positive punishments, do data suggest effects differ across

irritants, movement suppression, overcorrection, and interventions that couple these components with reinforcement procedures?

Method

Literature search

We searched the PsychInfo, ProQuest, and Web of Science databases using combinations of the following keywords: (a) treatment, therapy, training, or intervention, (b) self-injur*, SIB, self-destruct*, or self-harm, and NOT suicide, and (c) autis*, ASD, autism spectrum, intellectual disabilit*, retard*, or mental deficien*. We limited results to peer-reviewed journal articles published in English, but did not constrain by date of publication. Our database search yielded 679 articles for potential inclusion.

Inclusion criteria

In examination of the 679 articles, we used the following criteria to select studies or datasets for inclusion: (a) the experimental study used a single-case research design, beginning with a baseline phase that was followed by a treatment phase; (b) the dependent variable was a quantitative measure of SIB (e.g., frequency of head-hitting); (c) the independent variable was a treatment which targeted reduction in SIB; (d) participants had ASD and/or ID; (e) outcome data were presented graphically or numerically in a table for individual measurements/observation sessions; and (f) authors reported at least two data points per phase. After screening all identified articles, 137 studies (reported in 131 articles) that included a total of 245 unique participants were selected for inclusion in our analysis (see Appendix 1 for reference list of included studies).

Independent variables

Based on information available in articles, we selected independent variables, at two levels, that we hypothesized had potential to influence ESs (i.e., the individual participant-level and the study-level). For each variable, we operationally defined coding categories for each variable to capture the diversity across studies. Coding category definitions were expanded and refined at regular research team meetings, as needed to unambiguously sort all studies and participants. After initial training on coding conventions, two team members coded each article independently and then compared all coding decisions for agreement. All coding inconsistencies were resolved during research team meetings by group consensus.

Participant characteristics. At the participant level, we coded information on three characteristics (with variable categories in parentheses): (a) diagnosis (ASD, ID, and dual diagnosis); (b) communication limitations (no form of communication, some form of communication, verbal communication, and not specified); and (c) physical limitations (non-ambulatory, uses wheelchair, ambulatory, restraints, and not specified). We did not include participant age as a dependent variable because behavior analytic treatments have proven effective with all age groups (The National Autism Center, 2009), and we did not include IQ because reporting on participant IQ was missing from the majority of included original studies.

Study characteristics. At the study level, we coded information on four characteristics pertinent to treatment of SIB (with variable categories in parentheses): (a) assessment of SIB function (none/informal observation, FA, indirect assessment scale (e.g., Functional Analysis Screening Tool and/or Motivation Assessment Scale), and mixed assessment (e.g., FA plus indirect assessment scale)); (b) treatment matched to function (yes and no (e.g., providing verbal praise statements for behaviors that are maintained by attention)); (c) hypothesized function of SIB, as identified by authors or inferred during coding process from information provided in article (attention, escape/demand, tangible, automatic, multiply maintained, and not specified/other); and (d) treatment type (medication, functional communication training (FCT), differential reinforcement, extinction + reinforcement treatment packages, extinction, punishment, punishment + reinforcement treatment packages, and punishment + punishment treatment packages). Within categories of treatments, there were various forms of heterogeneity in intervention specifics, most notably among medications. Decisions to group treatments into the eight categories were driven by interventions' commonalities in procedures and principles, and the analytical benefits of maximizing sample sizes in categories, as well as the meta-analysis norm of tolerating minor forms of heterogeneity across studies (Cooper & Hedges, 1994) and the statistical capacity of MLM to model and test heterogeneity.

We also coded three study-level characteristics that we hypothesized could introduce bias or confound analysis of ESs, but are irrelevant to practical outcomes of treatment. These included (a) dependent variable metric (rate, frequency, percent of intervals, percent of trials, and duration); (b) length of data collection sessions (0–5 min, 6–10 min, 11–20 min, 21–45 min, 46–90 min, 91+ min, and not specified), and (c) baseline condition/treatment phase comparison (no treatment/no contingency in place, pretreatment routine, FA

condition, participant-perceived reinforcement contingency, and participant-perceived punishment contingency).

Analysis procedures

Data extraction. After completing coding for the independent variables, we extracted dependent measure data from included studies. We used a web application, WebPlotDigitizer (Rohatgi, 2016), to define the coordinate space of graphs, pinpoint coordinates of data points, and tabulate coordinate values. Investigation of the use of WebPlotDigitizer to extract data from single-case research graphs has found the program to yield highly reliable coordinates for data points (Moeyaert, Maggin, & Verkuilen, 2016). After extracted data were tabulated, we visually checked coordinate values against graphical displays for accuracy and then transferred the data into a spreadsheet.

ES calculation. We used Pustejovsky's measurement-comparable log-ratio ES measure to quantify treatment

effects (Pustejovsky, 2015). In contrast to other available summary statistics, Pustejovsky's ES is not subject to systematic bias related to sample sizes or baseline levels of behavior, and aggregates are not confounded by differences in outcome metrics across datasets. Also, because behavior changes, in theory, are multiplicative, as opposed to additive, linear aggregation of ESs for behavior is inappropriate and risks misrepresentation of phenomena in composite. The logarithmic form of Pustejovsky's ES bolsters the internal validity of ES aggregates and meta-analytic results.

Figure 1 illustrates graphically the steps in calculation of a log-ratio ES and the corresponding percent reduction. ESs were estimated using the following equation:

$$\omega = \ln\left(\frac{M_T}{M_B}\right) \quad (1)$$

where M_B and M_T are means for the baseline and treatment phases, respectively. In order to achieve unbiased estimates, we adjusted individual ESs using the correction factor J (Hedges & Olkin, 2014).

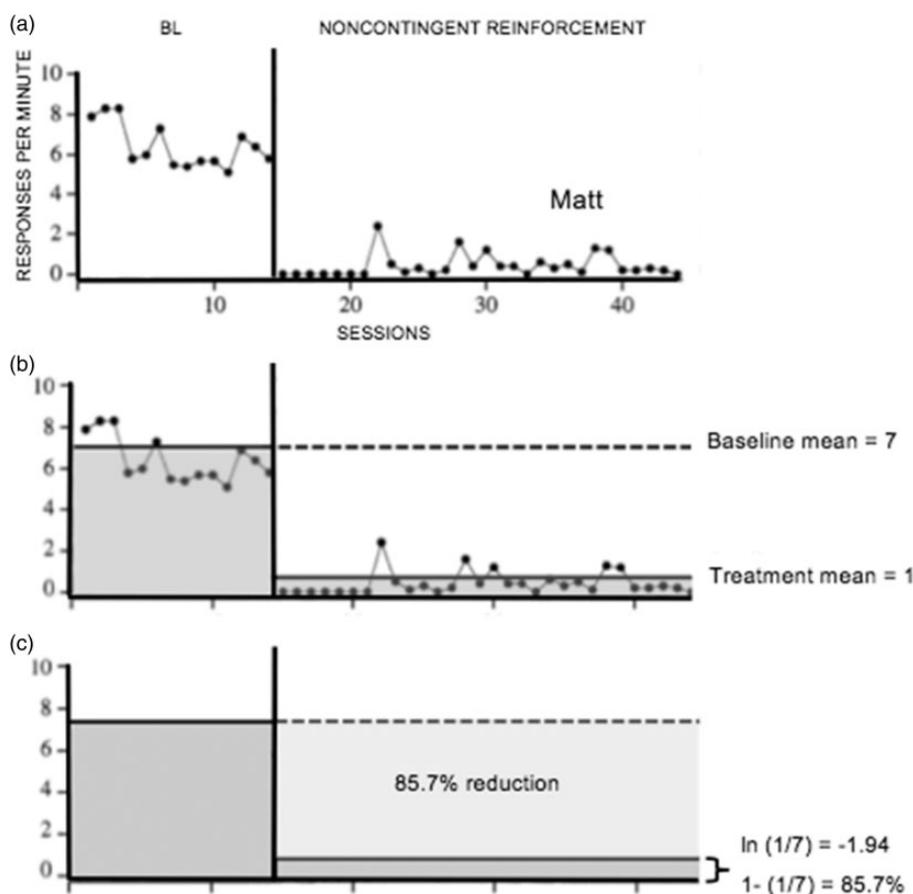


Figure 1. Illustration of log ratio and percent reduction calculations. The top panel (a) shows an original data display (Kahng, Iwata, Thompson, & Hanley, 2000); the middle panel (b) shows mean lines and values for baseline and treatment phases; and the bottom panel (c) shows the formulae for the log ratio and percent reduction and illustrates the percent reduction area on the original graph.

In aggregate analyses, we weighted individual ESs by their precision. We calculated ES precision as the inverse of the conditional variance of the log-ratio:

$$\frac{1}{\left(\frac{S_B^2}{n_B(M_B^2)}\right) + \left(\frac{S_T^2}{n_T(M_T^2)}\right)} \quad (2)$$

where S_B and S_T are standard deviations, and n_B and n_T were sample sizes for baseline and treatment phases.

As recommended by Pustejovsky (2015) to facilitate interpretation and increase accessibility of results, we transformed ESs from analysis output to percent reduction figures by converting log-ratios into exponentiation values and multiplying by 100. In this study, percent reduction figures represented the proportion of baseline SIB levels eliminated during treatment phases.

Multi-level analysis. Multi-level models in this meta-analysis were composed of linear regression equations that represented individual ESs as nested within studies. In the most basic model, at level-1, individual ESs were modeled as randomly varying around study average ESs. At level-2, study average ESs were modeled as randomly varying around an overall average ES. In subsequent models, explanatory variables (i.e., fixed effects) were added to level-2 equations to analyze associations between study-level variables and ESs. All models were estimated using SAS software (SAS Institute Inc., 2014). An alpha of .05 was selected for all statistical tests.

Our analysis began with preliminary inspections of means and distributions of ESs for each independent variable category (i.e., research question 1). Based on lack of apparent differences, we ruled out inclusion of participant and study characteristic variables.

To address research questions 2a–2d and establish a comparison for subsequent models, we estimated an unconditional model that included random effects at level-1 and level-2, to capture variation within studies (i.e., across individual ESs) and between study averages. The unconditional model included no fixed effects.

To investigate effects of different treatment types (i.e., research questions 3a–3c), we next estimated a conditional model that included fixed and random effects at level-2 for each treatment type, and a random effect at level-1. In the level-2 regression equation, we used dummy variables to represent each category of treatment. To test for differences in ESs between treatment types, we recoded dummy variables to rotate the reference category and re-ran the model to obtain contrasts between all pairs of treatments.

In a second conditional model, we evaluated the impact on effect magnitudes of matching treatments to SIB function (i.e., research questions 4a–4c).

The model included a series of dummy variables that represented treatment types when not matched to function and differences in effect when matched to function (i.e., the additive or subtractive effect of matching to function). To obtain estimates of ESs for each treatment type when matched to function, we reverse coded dummy variables and re-ran the model.

After completing our planned investigations, we undertook two post-hoc analyses to explore curiosities regarding differential effects of various forms of punishment interventions (i.e., research questions 5 and 6). We formulated hypotheses regarding variables which may account for variations in effects, operationally defined further categories of punishment interventions, and then recorded additional codes for all studies that involved punishment. As before, we estimated conditional models that included dummy variables for each category of punishment intervention. For these final two models, we forwent tests of differences between each form of punishment intervention and instead drew conclusions from the overlap of confidence intervals.

Results

Table 1 presents descriptive statistics for independent variables pertaining to participant characteristics (i.e., diagnosis, communication, physical limitation, and SIB function) and study characteristics (i.e., assessment, DV metric, session length, and comparison condition), which we opted to not include in meta-analytic models. We inspected descriptive statistics and ES plots, by independent variable category, for apparent patterns suggestive of systematic differences across categories or violations of assumptions of MLM. No evidence of homoscedasticity was found (nor other violations of assumptions). ES plots showed no patterns in density or clustering in ES magnitudes or distributions within or across categories, with the exception of treatment type. Means and standard deviations of each independent variable category similarly suggested homogeneity across categories. Although three variable categories had outlying average ES values (verbal communication, mixed assessment, and duration DV metric), all three values were unreliable due to very small category sample sizes. As a result, the outlying ESs could not be interpreted as indications of true category differences and we decided not to include these particular participant and study characteristic variables as explanatory variables in the models.

Overall average treatment effect

Table 2 summarizes results of the unconditional model, which aggregated ESs within and then across studies, and yielded a weighted average study-level ES.

Table 1. Descriptive statistics for independent variables not included in meta-analytic models.

	n	Weighted average ES	Standard deviation	Percent reduction
Diagnosis				
Autism spectrum disorder	11	-1.19	0.99	70%
Intellectual disability	212	-1.68	1.65	81%
Dual diagnosis	22	-2.12	1.77	88%
Communication				
No method	52	-1.58	1.45	79%
Some method	118	-1.59	1.53	80%
Verbal	3	-0.37	0.76	31%
Not specified	72	-2.14	2.04	88%
Physical limitation				
Non-ambulatory	32	-1.40	1.10	75%
Wheelchair	8	-1.42	0.78	76%
Ambulatory	63	-1.43	1.29	76%
Restraints	30	-1.88	1.94	85%
Not specified	112	-1.97	1.93	86%
Assessment of SIB function				
None or informal observation	119	-1.88	1.68	85%
Functional analysis	113	-1.61	1.62	80%
Indirect assessment	6	-3.50	2.46	97%
Mixed assessment	7	-0.78	0.58	54%
SIB function				
Attention	34	-1.23	0.72	71%
Escape/demand	33	-1.73	1.57	82%
Tangible	19	-1.45	2.03	77%
Automatic	35	-1.25	0.82	71%
Multiply maintained	114	-2.10	1.96	88%
Not specified	10	-2.13	2.16	88%
DV metric				
Rate/frequency	175	-1.79	1.81	83%
Percent of intervals	59	-1.38	1.04	75%
Percent of trials	10	-3.62	2.11	97%
Duration	1	0.03	0	-3%
Session length				
0-5 minutes	20	-1.63	1.03	80%
5-10 minutes	54	-2.04	2.11	87%
10-20 minutes	118	-1.71	1.52	82%
20-45 minutes	17	-1.05	1.44	65%
45-90 minutes	17	-1.40	1.46	75%
91+ minutes	13	-1.99	1.67	85%
Not specified	6	-2.68	1.94	93%
Comparison condition				
No treatment	119	-1.63	1.67	80%
Pretreatment	27	-1.43	1.34	76%
Functional analysis condition	57	-1.64	1.51	81%
Perceived reinforcement	36	-2.4	2.26	91%
Perceived punishment	6	-3.45	2.04	97%

Note. *n* = number of effect sizes for variable categories; SIB = self-injurious behavior

Table 2. Parameter estimates and percent reduction for the multilevel analysis.

Variable	N	Parameter estimate (SE)		Percent reduction [CI]	
<i>Unconditional model</i>					
Overall mean ES, γ .		−2.20***	(0.16)	89%	[85,92]
Between study variance, $\sigma_{u_k}^2$		2.06***	(0.39)		
Within study variance, σ_e^2		20.19***	(2.39)		
<i>Conditional model 1</i>					
Treatment					
Medication, γ_0	8	−0.32	(0.40)	27%	[−60,67]
FCT, γ_1	18	−1.62*	(0.62)	80%	[33,94]
Differential reinforcement, γ_2	66	−2.45***	(0.29)	91%	[85,95]
Extinction + reinforcement, γ_3	11	−0.87	(0.31)	58%	[24,77]
Extinction, γ_4	24	−1.63*	(0.50)	80%	[47,93]
Punishment, γ_5	76	−2.26***	(0.29)	90%	[82,94]
Punishment + reinforcement, γ_6	27	−2.79***	(0.56)	94%	[82,98]
Punishment + punishment, γ_7	15	−2.90**	(0.56)	94%	[84,97]
Variance between studies					
FCT, $\sigma_{u_{1k}}^2$		2.05	(1.57)		
Differential reinforcement, $\sigma_{u_{2k}}^2$		1.97**	(0.68)		
Extinction, $\sigma_{u_{4k}}^2$		2.17	(1.38)		
Punishment, $\sigma_{u_{5k}}^2$		1.74**	(0.60)		
Punishment + reinforcement, $\sigma_{u_{6k}}^2$		2.97*	(1.60)		
Punishment + punishment, $\sigma_{u_{7k}}^2$		1.38	(1.16)		
Within study variance, σ_e^2		20.24***	(2.40)		
<i>Conditional Model 2</i>					
Medication					
Unmatched, γ_{00}	8	−0.32	(0.40)	27%	[−61,67]
FCT					
Matched, γ_{10+11}	18	−1.62*	(0.62)	80%	[33,94]
Differential reinforcement					
Unmatched, γ_{20}	26	−2.63	(0.43)	93%	[89, 97]
Matched, γ_{20+21}	40	−2.27***	(0.41)	90%	[77, 95]
Matched-unmatched, γ_{21}		−0.36	(0.59)		
Extinction + reinforcement					
Matched, γ_{30+31}	11	−0.87	(0.31)	58%	[23, 77]
Extinction					
Unmatched, γ_{40}	5	−1.91	(1.25)	85%	[−70, 99]
Matched, γ_{40+41}	19	−1.57*	(0.56)	79%	[38, 93]
Matched-unmatched, γ_{41}		−0.34	(1.34)		
Punishment					
Unmatched, γ_{50}	62	−2.42	(0.32)	91%	[83, 95]
Matched, γ_{50+51}	14	−1.55*	(0.65)	79%	[24, 94]
Matched-unmatched, γ_{51}		−0.87	(0.73)		
Punishment + reinforcement					
Unmatched, γ_{60}	21	−2.92	(0.60)	95%	[82, 98]
Matched, γ_{60+61}	6	−1.87	(1.66)	85%	[−297, 99]
Matched-unmatched, γ_{61}		−1.06	(1.76)		

(continued)

Table 2. (continued)

Variable	N	Parameter estimate (SE)		Percent reduction [CI]	
Punishment + punishment					
Unmatched, γ_{70}	14	-2.96	(0.58)	95%	[84, 98]
Matched, γ_{70+71}	1	-0.96	(0.63)	62%	[-32, 89]
Matched-unmatched, γ_{71}		-1.00	(3.14)		
Variance between studies					
FCT, σ_{u1k}^2		2.04	(1.57)		
Differential reinforcement, σ_{u2k}^2		2.03**	(0.70)		
Extinction + reinforcement, σ_{u3k}^2		0.11	(0.30)		
Extinction, σ_{u4k}^2		2.38	(1.51)		
Punishment, σ_{u5k}^2		1.68**	(0.60)		
Punishment + reinforcement, σ_{u6k}^2		3.07*	(1.74)		
Punishment + punishment, σ_{u7k}^2		1.42	(1.20)		
Within study variance, σ_e^2		20.40***	(2.42)		

Note. FCT and Reinforcement + Ext. treatment packages were always matched to function; SE = standard error; CI = confidence interval. Matched-Unmatched values are the difference between effect size values for treatments matched to function of SIB and treatments not matched to SIB function.

*** $p < .001$.

** $p < .01$.

* $p < .05$.

The unconditional model included two random effects, one that represented variation in study average ESs, and one that represented variation in individual ESs within studies. In addition to addressing research questions 2a–2d, the purpose of the unconditional model was to serve as a baseline for interpretation of subsequent conditional models.

The average log-ratio ES for overall treatment effect was -2.2 ($SE = 0.16$, $df = 117$, $p < .001$), which corresponds to an 89% reduction in SIB with a 95% CI [85, 92]. The random effect variance component representing variation in study average ESs was moderate and statistically significant, indicating that variation in study level ESs was greater than would be expected due to sampling bias alone ($\hat{\sigma}_{u1k}^2 = 2.06$, $se = 0.39$, $df = 117$, $p > .001$). The random effect variance component representing variation in individual ESs within studies was large and statistically significant, similarly indicating that there was substantial variation within studies ($\hat{\sigma}_e^2 = 20.19$, $se = 2.39$, $df = 117$, $p > .001$). Of the total ES variance in the unconditional model, 91% occurred within studies. This large proportion of within study variance suggested that participant-level variables accounted for much more variation in ESs than study-level variables.

Treatment effect by treatment type

The first conditional model addressed research questions 3a–3c, which regarded effects of different treatment types. We also compared the first conditional model against the unconditional model to determine whether including treatment type increased the explanatory value

of the model. We included eight fixed effects for the eight treatment types: medication, FCT, differential reinforcement, extinction + reinforcement packages, extinction, punishment, punishment + reinforcement packages, and punishment + punishment packages (more than one punishment procedure used at the same time). Results for the first conditional model appear in Table 2. The fixed effects, γ_0 through γ_7 , represent ES estimates for the eight treatment types; and the random effects, $\hat{\sigma}_{u1k}^2$ through $\hat{\sigma}_{u7k}^2$, capture the amount of variance between study average ESs within each treatment type. Random effects for medication ($\hat{\sigma}_{u0k}^2$) and extinction + reinforcement packages ($\hat{\sigma}_{u3k}^2$) were excluded to enable model estimation. The small sample sizes for categories of medication and extinction + reinforcement package treatments ($n = 8$ and $n = 11$, respectively) prevented convergence of effect estimates in SAS software.

Treatments with small average effects. Medication and extinction + reinforcement packages had small weighted ESs relative to the other six treatment types. The ES estimate for medication was -0.32 ($SE = 0.4$, $df = 6$, $p = .43$), which corresponds to a 27% reduction in SIB with a 95% CI [-60, 67]. The ES estimate for extinction + reinforcement packages was -0.87 ($SE = 0.31$, $df = 20$, $p = .09$), which corresponds to a 58% reduction in SIB with a 95% CI [24, 77]. The difference between effects of medication and extinction + reinforcement packages was not significant ($p = .29$).

Treatments with moderate to large effects. Extinction and FCT were associated with moderate to large weighted

ESs, which were statistically significant. The ES estimate for extinction was -1.63 ($SE=0.5$, $df=17$, $p=.01$), which corresponds to an 80% reduction in SIB with a 95% CI [47, 93]. The ES estimate for FCT was -1.62 ($SE=0.62$, $df=14$, $p=.04$), which corresponds to an 80% reduction in SIB with a 95% CI [33, 94]. The difference between ES estimates for extinction and FCT was not statistically significant. Contrasts were not significant between ES estimates for medication and ES estimates for extinction and FCT ($p=.055$ and $p=.10$, respectively) and extinction + reinforcement and extinction and FCT ($p=.23$ and $p=.31$, respectively).

Treatments with very large effects. Punishment, differential reinforcement, punishment + reinforcement packages, and punishment + punishment packages were associated with very large, statistically significant ESs. The ES estimate for punishment was -2.3 ($SE=0.29$, $df=63$, $p<.0001$), which corresponds to a 90% reduction in SIB with a 95% CI [81.5, 94]. The ES estimate for differential reinforcement was -2.5 ($SE=0.29$, $df=60$, $p<.0001$), which corresponds to a 91% reduction in SIB with a 95% CI [85, 95]. The ES estimate for punishment + reinforcement packages was -2.8 ($SE=0.56$, $df=32$, $p=.0003$), which corresponds to a 94% reduction in SIB with a 95% CI [82, 98]; and the ES estimate for punishment + punishment packages was -2.89 ($SE=0.56$, $df=14$, $p=.0001$), which corresponds to a 94.5% reduction in SIB with a 95% CI [83, 98]. Differences between the average ESs for these four treatments were not statistically significant. The ES estimate for extinction was substantially smaller than the ES estimates for the four treatments, but the differences were not statistically significant ($.11 \leq p \leq .30$). Contrasts were not significant between the ES estimates for the four treatments and FCT ($.15 \leq p \leq .38$). Contrasts were significant between ES estimates for the four treatments and medication ($p \leq .001$ for all tests). Contrasts were significant between ES estimates for the four treatments and extinction + reinforcement ($p \leq .01$ for all tests).

Variation between studies in treatment effects. The between-study variance estimates for extinction ($\hat{\sigma}_{u4k}^2 = 2.17$, $se = 1.38$, $df = 17$, $p = .06$), FCT ($\hat{\sigma}_{u1k}^2 = 2.05$, $se = 1.57$, $df = 3$, $p = .09$), and punishment + punishment packages ($\hat{\sigma}_{u7k}^2 = 1.38$, $se = 1.16$, $df = 14$, $p = .12$) were not statistically significant, suggesting there was no longer substantial variation in study average ESs after controlling for the effect of treatment type. However, small intra-class correlation coefficient (ICC) values for the groups of studies using extinction (ICC=0.10), FCT (ICC=0.09), and punishment + punishment packages (ICC=0.06) indicate that only 6% to 10% of the

variance in treatment effects was at the study-level, while 90% to 94% of the variance was at the participant-level.

The between-study variance estimates for punishment ($\hat{\sigma}_{u5k}^2 = 1.74$, $SE = .605$, $df = 63$, $p = .002$), differential reinforcement ($\hat{\sigma}_{u2k}^2 = 1.97$, $SE = .68$, $df = 60$, $p = .002$), and punishment + reinforcement packages ($\hat{\sigma}_{u6k}^2 = 2.97$, $SE = 1.6$, $df = 32$, $p = .03$) were all statistically significant, revealing there was still substantial variation in study average ESs after controlling for treatment type. Similar to above, small values for the residual ICC for the groups of studies using punishment (ICC=0.08), differential reinforcement (ICC=0.09), and reinforcement + punishment packages (ICC=0.13) indicate only 8% to 13% of the variance in treatment effects was at the study level, while 87% to 92% was at the participant level. As noted previously, random effects for medication and extinction + reinforcement packages were excluded to allow data to meet convergence criteria.

Variation within studies in treatment effects. The variance component $\hat{\sigma}_e^2$ increased slightly from 20.19 ($SE=2.39$) in the unconditional model to 20.24, $se=2.42$, $df=244$, $p<.001$, in the first conditional model. In other words, the variance for individual ESs within studies remained similarly large after estimating separate treatment effects in the first conditional model. The lack of change suggests participant-level variables accounted for most variation in treatment effects. Intra-class correlation coefficients reported above were calculated using the variance component for the first conditional model.

Treatment effect by treatment type and matched to SIB function

Table 2 presents the results for the second conditional multilevel model, which addressed research questions 4a–4c, regarding whether matching treatment to function had an impact of SIB treatment effects. We also compared the second conditional model against the first model to determine whether including matched to function increased the overall explanatory value of the model. We ran the second conditional multilevel model twice using reverse coding of dummy variables to achieve standard error estimates and calculate confidence intervals for both treatments matched to function and not matched to function. Figure 2 provides a visual representation of the percent reduction figures and confidence intervals for each treatment as estimated by the unconditional model, the first conditional model, and the second conditional model.

The fixed effects, γ_{20} , γ_{40} , γ_{50} , γ_{60} , and γ_{70} , represent ES estimates for five treatment types when not matched

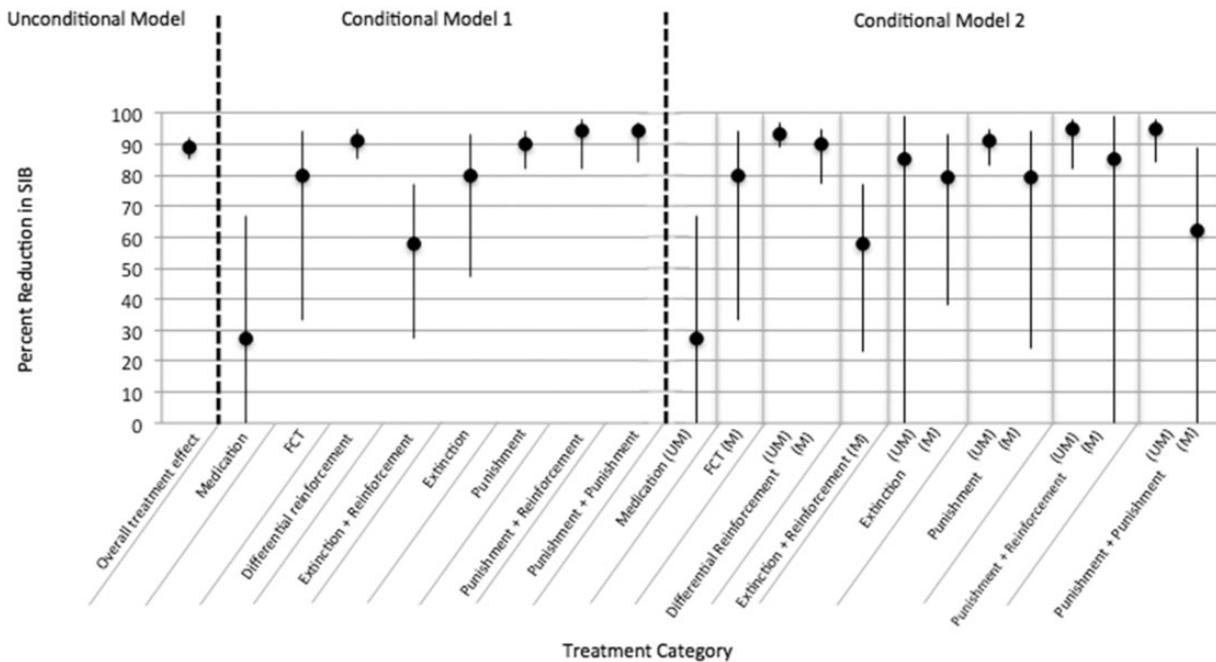


Figure 2. Mean percent reduction in SIB for the three multilevel models with 95% confidence intervals. The unconditional model includes the overall average effect of treatment. Conditional model 1 includes average percent in SIB by treatment category. Conditional model 2 includes average percent reduction in SIB by treatment category matched to function (M) and unmatched (UM) to function. Cut-off for lower bound of confidence intervals is sent to 0 for medication in conditional model 1, and extinction (UM), punishment + reinforcement (M), and punishment + punishment (M) in conditional model 2.

to the function of SIB. ES estimates for medication, FCT, and extinction + reinforcement packages (γ_{00} , γ_{10} and γ_{30}) are excluded from this list because these treatments were always matched to function or, in the case of medication, not matched to function. The fixed effects γ_{10+11} , γ_{20+21} , γ_{30+31} , γ_{40+41} , γ_{50+51} , γ_{60+61} , and γ_{70+71} represent ES estimates for treatments when matched to function, and the fixed effects γ_{21} , γ_{41} , γ_{51} , γ_{61} , and γ_{71} represent the difference between matched and unmatched treatment effects.

ES differences between matched and unmatched treatments.

For each treatment type, ESs, ES differences, and percent reduction figures for treatments unmatched and matched to function are summarized in Table 2. ESs for unmatched treatments (γ_{20} , γ_{40} , γ_{50} , γ_{60} , and γ_{70}) had greater magnitude than ESs for treatments matched to function ($\gamma_{10} + \gamma_{11}$ through $\gamma_{70} + \gamma_{71}$) across all treatment types for which contrasts were possible; however, none of the differences were statistically significant. Regarding effect magnitudes (i.e., differences from zero), the ESs for punishment and differential reinforcement were statistically significant when treatments were both matched and unmatched to function (matched punishment, $\omega = -1.55$, $SE = .65$, $df = 13$, $p = .02$; unmatched punishment, $\omega = -0.87$, $SE = 0.73$, $df = 61$, $p = 0.24$; matched differential reinforcement, $\omega = -2.27$, $SE = 0.41$, $df = 39$, $p < .001$;

unmatched differential reinforcement, $\omega = -0.36$, $SE = 0.59$, $df = 25$, $p = 0.55$). ESs for punishment + reinforcement packages and punishment + punishment packages were statistically significant only when treatment was unmatched to function (matched punishment + reinforcement packages, $\omega = -1.87$, $SE = 1.66$, $df = 5$, $p = 0.29$; unmatched punishment + reinforcement packages, $\omega = -1.06$, $SE = 1.76$, $df = 20$, $p = 0.56$; matched punishment + punishment packages, $\omega = -0.96$, $SE = 3.10$, $p = 0.76$; unmatched punishment + punishment packages, $\omega = -2.01$, $SE = 3.14$, $df = 13$, $p = 0.53$). Finally, the ES for extinction was statistically significant only when treatment was matched to function (matched extinction, $\omega = -1.57$, $SE = 0.56$, $df = 18$, $p = 0.02$; unmatched extinction, $\omega = -0.34$, $SE = 1.34$, $df = 4$, $p = 0.80$).

Variation between studies in treatment effects. Similar to results of the first conditional model, random effect variance components remained significant for punishment ($\hat{\sigma}_{u5k}^2 = 1.68$, $SE = .6$, $p = .003$), differential reinforcement ($\hat{\sigma}_{u2k}^2 = 2.03$, $SE = .7$, $p = .002$), and punishment + reinforcement packages ($\hat{\sigma}_{u6k}^2 = 3.07$, $SE = 1.74$, $p = .04$), indicating that even after conditioning on “matched to SIB function,” there was additional explainable variation across study average ESs for these three treatment types. Random effect variance components were also used to calculate the proportion of

Table 3. Post-hoc analysis of positive and negative punishment treatments.

Punishment type	n	Weighted average ES	Standard error	Percent reduction [95% CI]	p value
Negative punishment					
Not matched to SIB function	4	-1.34	1.32	74% [-249, 98]	.317
Positive punishment					
Not matched to SIB function	58	-2.48	0.34	92% [84, 96]	<.001
Matched to SIB function	14	-1.52	0.66	78% [20, 94]	.031
Random Effects					
Between study variance, σ_{uk}^2		1.61**	0.64		
Within study variance, σ_e^2		26.65***	5.25		

Note. n = number of effect sizes included in weighted average; ES = effect size; CI = confidence interval. P-values in table are for t-tests of effect sizes' differences from zero generated by SAS software. Differences between effect sizes were not significant.

*** $p < .001$.

** $p < .01$.

* $p < .05$.

variance explained in treatment categories from the first conditional model to the second conditional model (i.e., the percentage of the variation in ESs explained by whether treatments were matched to function). The proportion of variance explained by matching to function was small and positive for punishment (3.33%), but was small and negative for differential reinforcement (-2.76%), extinction (-9.75%), punishment + reinforcement packages (-3.47%), and punishment + punishment packages (-2.3%). Negative percentages indicate that the second condition model failed to explain additional variation in ESs and actually decreased in explanatory power. All values were expectedly small because they involve between study variance and, as described above, most variance was within studies.

Variation within studies in treatment effects. The lack of reduction in $\hat{\sigma}_e^2$ from the first conditional model to the second conditional model ($\hat{\sigma}_e^2 = 20.40$, $SE = 2.42$, $df = 244$, $p < .001$) affirmed that participant-level variables rather than study-level variables account for a greater proportion of variance in individual ESs.

Differences in the effects of punishment treatments

We were interested to examine whether the remaining variation in punishment effects could be accounted for by the possibility that certain types of punishment were more effective than others. In a first post-hoc analysis, we investigated the possibility that positive punishment treatments were more effective than negative punishment treatments. Positive punishment is defined as the presentation of a stimulus that decreases behavior (e.g., administration of irritants) and negative punishment is defined as the removal of a stimulus that decreases behavior (e.g., time out; Cooper et al., 2007).

Negative punishment may not be as effective as positive punishment due to the removal of a stimulus. While positive punishments are highly effective, they are also generally more aversive and more ethically objectionable than negative punishments. In a second post-hoc analysis, we grouped positive punishment interventions by punishing stimuli (i.e., irritant, movement suppression, overcorrection/exercise) and intervention components (i.e., punishment alone or packaged with reinforcement procedures) to investigate the possibility that use of particular stimuli was associated with differences in treatment effects. As meta-analysts free from ethical concerns of applied experimentation with punishments, we viewed the post-hoc analyses as a unique opportunity to investigate the relative efficacy of different forms of punishments in treatment for SIB.

Table 3 displays the results of the first analysis which contrasted the effects of positive and negative punishment. None of the negative punishment treatments were matched to SIB function, and negative punishment treatments had an average ES of -1.34 ($SE = 1.32$, $df = 3$, $p = .317$), corresponding to a 74% reduction in SIB with a 95% CI [-249, 98]. The large standard error and wide confidence interval are likely due to the very small sample size ($n = 4$). Positive punishment treatments were split between those matched to function ($n = 14$) and those not matched to function ($n = 58$). Positive punishment treatments matched to function had an ES of -1.52 ($SE = .66$, $df = 13$, $p = .03$), corresponding to 78% reduction in SIB with a 95% CI [20, 94]. Positive punishment treatments not matched to function had an ES of -2.48 ($SE = .43$, $df = 57$, $p < .001$), corresponding to 95% reduction in SIB with a 95% CI [84, 96]. All three confidence intervals for ES estimates overlapped substantially, indicating the difference in ES magnitudes were not statistically significant. Significance of the study-level

Table 4. Post-hoc analysis of punishing stimuli.

Punishment type	n	Weighted average ES	Standard error	Percent reduction [95% CI]	p value
Single component intervention					
Irritant	37	-2.45	0.47	91% [78, 97]	<.001
Movement suppression	29	-2.46	0.46	91% [79, 97]	<.001
Overcorrection or exercise	6	-1.36	0.78	74% [-19, 94]	.089
Package with reinforcement					
Irritant	1	-1.91	0.32	85% [72, 92]	<.001
Movement suppression	25	-2.41	0.51	91% [76, 97]	.001
Overcorrection	1	-6.91	0.24	99.9% [99.8, 99.9]	<.001
Random effects					
Between study variance		1.91**	0.70		
Single component studies, σ_{1k}^2					
Package with reinforcement, σ_{2k}^2		1.77	1.24		
Within study variance, σ_e^2		22.17***	4.01		

Note. n = number of effect sizes included in weighted average; ES = effect size; CI = confidence interval. Where $n > 0$, p -values are for t-tests of weighted average effect sizes' differences from zero generated by SAS software. Differences between groupings' effect sizes are insignificant. Where $n = 1$, p -values are for z-tests of individual effect sizes' difference from zero.

*** $p < .001$.

** $p < .01$.

* $p < .05$.

random effect variance components suggests that there is still variance in study average ESs remaining to be explained after modeling the type of punishment.

Table 4 presents results of the second post-hoc analysis. No meaningful differences were apparent across average ESs for the single component and reinforcement package punishment treatments. Overcorrection/exercise was associated with 74% reduction in SIB when used alone (95% CI [-19, 94], $p = .09$), and 99.9% reduction in SIB when used in combination with reinforcement (95% CI [99.8, 99.9], $p < .001$). Movement suppression resulted in 91% in SIB when used alone (95% CI [79, 97], $p < .001$), and 91% reduction in SIB when used in combination with reinforcement (95% CI [76, 97], $p = .001$). Irritants were associated with 91% reduction in SIB when used alone (95% CI [78, 97], $p < .001$), and 85% reduction in SIB when combined with reinforcement (95% CI [72, 92], $p < .001$). The very small sample sizes for overcorrection/exercise alone, irritant + reinforcement, and overcorrection + reinforcement packages render these ESs and percent reduction figures unreliable and unfit for comparison. As in the first post-hoc analysis, significance of the study-level random effect variance component for studies using positive punishments alone indicates there is variance in study average ESs for which the model could not account ($\hat{\sigma}_{1k}^2 = 1.91$, $SE = 0.70$, $df = 244$, $p = .003$). However, the lack of significance of the study-level random effect variance component for studies that packaged reinforcement with punishment suggests study average ES does not

vary, after controlling for punishing stimuli, beyond what is expected due to sampling bias ($\hat{\sigma}_{2k}^2 = 1.77$, $SE = 1.24$, $df = 244$, $p > .08$).

Discussion

Relations between participant-level variables and treatment effects

Descriptive statistic results corroborate the behavior analytic perspective that environmental manipulations are more influential than participant characteristics (Cooper et al., 2007). For example, the percent reduction figures corresponding to weighted mean ESs were nearly identical for participants with differing physical abilities, including those in wheelchairs (75%), those who are ambulatory (75%), and those who are non-ambulatory (76%). Percent reduction figures were also similar across the communication ability categories with the exception of the verbal communication category.

Only three individuals in the sample had verbal communication abilities. The small number of verbal participants in the sample is consistent with the suggestion that an ability to manipulate the environment with speech may serve as a protective factor against SIB (Schroeder et al., 1978). The percent reduction figures associated with weighted mean ESs for participants with verbal skills (31% reduction in SIB) was much smaller than the percent reduction figures across the limited or no communication categories (79%–88%). Based on visual inspection of the three original graphs

for verbal participants, it appears that SIB levels for one subject (Clauser & Gould, 1988) were very similar during the baseline and treatment (extinction) phases. The similarity of means across phases would have yielded a very small ES, and given the small sample size ($n = 3$), this one effect likely skewed the group average giving the false impression that individuals with verbal skills have SIB that is somewhat more resistant to treatment.

An alternative explanation for insignificant differences between weighted mean ESs across diagnosis categories is that changes in ASD and ID diagnostic criteria over time resulted in low internal validity in our coding categories (i.e., ASD, ID, dual diagnosis) and obscured potential differences. Participant diagnoses reported in original studies reflect the fields' understanding and characterization of ASD and ID at the time of publication. Our sample of studies spans five decades and included 38 participants in studies published before 1987, when autistic disorder first appeared in the DSM III. ASD may have been under-identified or conflated with ID for participants in studies published before 1987. The validity of diagnosis coding for studies published after 1987 is also questionable given that criteria for ASD diagnosis changed again in subsequent versions of the DSM.

Additionally, we found that treatment of SIB had similar effects across all functions of behavior (i.e., attention, escape/demand, tangible, automatic, and multiply maintained). In other words, sorting by function (without regard for treatment type) did not reveal differences in the magnitude of SIB treatment effects.

Given we did not detect category differences in variables we coded, we assume we did not have access to information on participant-level characteristics that influenced treatment effects. For example, it is possible that interventionist expertise, interventionist-participant rapport, or variables extraneous to intervention procedures (e.g., uncontrolled setting events or discriminative stimuli) were responsible for the observed variability in treatment effects.

Effectiveness of different treatments for SIB

Our results indicate that one group of treatments yielded large reductions in SIB (differential reinforcement, punishment, punishment + reinforcement, and punishment + punishment), and a second group of treatments yielded relatively smaller reductions in SIB (medication, extinction, extinction + reinforcement, and FCT). Within the group of treatments associated with smaller effects, our data suggest that medication in particular is not an effective treatment for SIB. The distribution of medication ESs skewed towards no effect, and visual analysis of individual graphs reveals that some participants in the sample experienced slight

increases in SIB while receiving medical treatment. It is important to note, however, that we aggregated the effects of different medications due to very small sample sizes (i.e., Naltrexone ($n = 5$), Clozapine ($n = 1$), and Clozapine + Depakote ($n = 1$), and Carbamazepine ($n = 1$)). Given the potential for individual differences in reactions to medications and issues related to dosages, doing so may have resulted in low internal validity in this category. Relatedly, it is important to remember that a much greater proportion of the variance in treatment effects was found at the participant level than at the study level in the meta-analysis.

It is also important to note that our choice of ES statistic may have privileged faster-acting treatments (i.e., punishment) over treatments that affect behavior change more slowly. The percent reduction figure is driven by phase mean differences without consideration of time. A quick reduction in SIB (i.e., level change) during treatment yields a large difference between baseline and treatment means and a correspondingly large ES. The more gradual slope associated with a slower decrease in SIB yields a smaller mean difference between phases and a smaller ES. Visual analysis of the set of graphical displays for FCT treatment in our dataset indicates that FCT was sometimes associated with a more gradual reduction in SIB. As a result, the magnitude of the ES associated with FCT may be smaller than ESs associated with other treatments not because FCT results in smaller reductions in SIB, but because FCT treatment reduces SIB more gradually.

Extinction and extinction + reinforcement may be associated with smaller SIB treatment effects because participants may have had extinction bursts or increases in behavior in response to the withdrawal of expected consequences. In the event of an extinction burst, the increase in behavior following the start of treatment would confound the log-ratio ES. Further, punishment + reinforcement has the potential to suppress behavior and create new behavior. Extinction does not teach behavior and reinforcement of a new behavior may take time to show results.

Differential reinforcement and punishment-based treatments were associated with the largest treatment effects. Although there was no measurable difference in the magnitude of differential reinforcement and punishment-based effects, research does suggest there are differences in how quickly and sustainably these treatments work to reduce SIB. Punishment may act quickly but may not affect long-term behavior change (e.g., Favell, McGimsey, & Schell, 1982). Differential reinforcement may result in a slower reduction in SIB, but reinforcement procedures can teach new behaviors and result in long-term behavior change (e.g., Favell, McGimsey, & Schell, 1982). Our results support use of punishment + reinforcement treatment

packages in efforts to reduce SIB quickly. While we did not synthesize data on maintenance of treatment effects, our review of studies did expose us to many studies which suggested that teaching new, adaptive behaviors sustained reductions in SIB (e.g., Bass & Speak, 2005; Iwata et al., 1994).

Current practice recommendations from a variety of government organizations advocate for use of reinforcement-based treatments (e.g., Department of Health, 2014; Epstein, Atkins, Cullinan, Kutash, & Weaver 2008; Individuals with Disabilities Education Act, 2004; Royal College of Psychiatrists, British Psychological Society, & Royal College of Speech and Language Therapists, 2007). Similarly, applied behavior analysts assert aversive punishment procedures should be reserved for situations in which other treatments have failed (e.g., Iwata, 1988; Matson & Taras, 1989; Worlery, Bailey, & Sugai, 1998). Yet, it can also be argued that allowing SIB to persist is unethical when using a fast-acting treatment could prevent serious harm (Houten et al., 1988). Consideration of punishment-based treatments requires consideration of ethical objections and potential adverse effects. Research suggests that an individual may respond to punishment by engaging in aggressive behaviors if aggression has been successful in escaping punishment in the past (e.g., Azrin & Holz, 1966). Additionally, punishment alone fails to teach a replacement behavior (e.g., Alberto & Troutman, 2012). In the absence of a replacement behavior, it is likely that SIB or a functionally equivalent behavior will reemerge, even if punishment initially eliminated the behavior (Cooper et al., 2007).

It is important to note that punishment is technically defined only as a consequence that reduces behavior. Ethical objections to punishment may stem from a belief that punishment is cruel or unjust, and may rest on the assumption that punishment is invasive and aversive. Of great relevance to debates on which and when punishments should be employed, results of our post-hoc analyses suggest that less aversive and invasive punishment procedures (e.g., overcorrection, time out) yield similar treatment effects as highly aversive and invasive punishments (e.g., electric shock, lemon juice sprays).

Effectiveness of matching treatment to function

Our analysis did not reveal differences in ESs for treatments matched and not matched to function. These results run counter to the common expectation that treatments matched to the function of behavior are more effective in behavior modification (e.g., as codified in requirements to conduct functional behavior assessments in US schools; Individuals with Disabilities Education Act, 2004). However, our results align with

similar findings reported by Denis et al. (2011) and Machalicek, O'Reilly, Beretvas, Sigafoos, and Lancioni (2007).

The principle of matching interventions to needs identified by assessments is widely supported by research in other areas of education and development (e.g., curriculum-based assessment; Hosp, Hosp, & Howell, 2016). While the practice of matching treatment to SIB function is consistent with this time-honored and research-supported principle, our data suggest further variables at the individual-level, yet unrecognized or targeted in established assessment methods, greatly influence the outcomes of intervention.

Limitations

Several limitations associated with our dataset should be noted. While we sampled the entire single-case SIB treatment literature, small sample sizes in several variable categories resulted in a number of unreliable estimates, namely those pertaining to medications, a few punishment categories in post-hoc analyses, and several matched and unmatched treatment categories. MLM typically provides robust estimates, but precision and accuracy of estimates positive correlates with sample size (Raudenbush & Liu, 2000).

Many primary studies did not include detailed participant-level information or report maintenance data. The large proportion of within-study variance associated with our models led us to hypothesize that one or more unaccounted-for participant-level variable(s) may have influenced treatment effects. Access to more participant-level information might have allowed us the opportunity to code and analyze additional and potentially meaningful participant-level variables in our analysis. The inclusion of maintenance data might have decreased the magnitude of effects for treatments that work quickly to reduce SIB but do not result in long-term behavior change (i.e., punishment only treatments).

We must also address study limitations related to our analysis. Over the five decades during which primary studies in our sample were published, behavior analysis and single-case research evolved (e.g., Odom, Brantlinger, Gesten, Horner, Thompson, & Harris, 2005). At different time points, study participants with ASD and ID would have had differing levels of access to care and educational opportunities; interventionists would have had different training experiences; and researchers would have been bound by different standards for ethics and rigor. Research in education and medicine has demonstrated that baseline and treatment values can 'drift' over time with changes in societal practices and environmental conditions (e.g., Lemons, Fuchs, Gilbert, & Fuchs, 2014). Though the potential effect of time on our dataset raises interesting

questions, we did not model time in our analyses for the sake of parsimony and diminishing benefits of greater complexity in statistical models.

Also, single-case research is an inductive method, while statistical modeling is inferential. In single-case designs, conditions are manipulated across time, and manipulations are replicated, in efforts to induce reliable patterns of behavior change. In group design research, conditions are manipulated across people, and from differences in groups' behavior patterns, inferences are drawn regarding variable associations. As described in the introduction, much empirical research and wide-ranging consensus supports the practice of meta-analysis of single-case data. However, this fundamental difference has an unknown impact on the internal validity of use of inferential statistics with single-case data.

Implications for practice

Results suggest that treatments involving differential reinforcement, punishment, and reinforcement plus punishment treatments result in similar reductions in SIB. In line with current practice recommendations, we encourage use of reinforcement-based procedures in all cases of SIB. In the event that reinforcement-only treatments have failed or if SIB poses a serious, immediate threat to the health and well-being of an individual, our results suggest that overcorrection paired with reinforcement may be the most effective as well as less invasive alternative. Overcorrection can quickly reduce SIB, and by including a reinforcement component to teach more adaptive behavior, treatment can result in sustainable behavior change.

Results of the meta-analysis could be interpreted to support use of punishment procedures as sole components of interventions. We caution against this interpretation and discourage the practice of punishment in isolation. Research suggests that punishment overrides the effect of reinforcement from problem behavior, but the effects of punishments depend on having the punishing stimulus presented upon the occurrence of the challenging behavior. The common presence of obstacles to consistent delivery of punishing stimuli, along with legal and ethical issues of punishment make use of reinforcement-based practices alone or in conjunction with less aversive and invasive punishments more reasonable, practical, and efficient.

As a final point, despite the lack of support for treatments that were matched to function over those that were not matched to function, it is advisable to follow current practice recommendations for assessing SIB and basing treatment decisions on functional assessment findings. In our study, we did not investigate pro-social behaviors, however, functional assessments can guide and support the selection of reinforcement-based

procedures that target increases in socially acceptable behaviors, enhancements in an individual's community involvement, and improvements in the quality of their relationships (Mace, 1994).

Conclusion

SIB is common, persistent, and can have serious physical and social consequences. The present review used a rigorous multi-level meta-analysis to synthesize the large body of single-case design SIB treatment literature in order to investigate the relative effectiveness of various treatments for SIB. Results indicate that overall, treatment for SIB is highly effective and that participant and study characteristics do not moderate treatment effects. Treatment effects for punishment, differential reinforcement, and treatment packages including punishment and differential reinforcement were largest; effects for extinction and FCT were slightly smaller; and effects for medication and extinction + reinforcement treatment packages were smallest. Differences between treatments not matched to function and treatments matched to function were observed, although evaluated to be negligible.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Alberto, P. A., & Troutman, A. C. (2012). *Applied behavior analysis for teachers* (9th ed.). Upper Saddle River, NJ: Pearson Education.
- Allison, D. B., & Gorman, B. S. (1994). "Make things as simple as possible, but no simpler." A rejoinder to Scruggs and Mastropieri. *Behaviour Research and Therapy*, 32, 885–890.
- Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behaviour Research and Therapy*, 31, 621–631.
- Azrin, N. H., & Holz, W. C. (1966). Punishment. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application* (pp. 380–447). New York, NY: Appleton-Century Crofts.
- Baek, E. K., & Ferron, J. M. (2013). Multilevel models for multiple-baseline data: Modeling across-participant variation in autocorrelation and residual variance. *Behavior Research Methods*, 45, 65–74.
- Baek, E. K., Moeyaert, M., Petit-Bois, M., Beretvas, S. N., Van den Noortgate, W., & Ferron, J. M. (2014). The use of multilevel analysis for integrating single-case

- experimental design results within a study and across studies. *Neuropsychological Rehabilitation*, 24, 590–606.
- Baghdadli, A., Pascal, C., Grisi, S., & Aussilloux, C. (2003). Risk factors for self-injurious behaviours among 222 young children with autistic disorders. *Journal of Intellectual Disability Research*, 47, 622–627.
- Bass, M. N., & Speak, B. L. (2005). A behavioural approach to the assessment and treatment of severe self-injury in a woman with Smith-Magenis syndrome: A single case study. *Behavioural and Cognitive Psychotherapy*, 33, 361–368.
- Beretvas, S. N., & Chung, H. (2008). A review of meta-analyses of single-subject experimental designs: Methodological issues and practice. *Evidence-Based Communication Assessment and Intervention*, 2, 129–141.
- Busk, P. L., & Serlin, R. (1992). Meta-analysis for single case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Clauser, B., & Gould, K. (1988). Visual screening as a reductive procedure: An examination of generalization and duration. *Behavioral Interventions*, 3, 51–61.
- Cohen, S. A., Ihrig, K., Lott, R. S., & Kerrick, J. M. (1998). Risperidone for aggression and self-injurious behavior in adults with mental retardation. *Journal of Autism and Developmental Disorders*, 28, 229–233.
- Cooper, H., & Hedges, L. (1994). *The handbook of research synthesis*. New York, NY: Russell Sage Foundation.
- Cooper, J. O., Heron, E. T., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Davanzo, P. A., Belin, T. R., Widawski, M. H., & King, B. H. (1997). Paroxetine treatment of aggression and self-injury in persons with mental retardation. *American Journal on Mental Retardation*, 102, 427–437.
- Denis, J., Van den Noortgate, W., & Maes, B. (2011). Self-injurious behavior in people with profound intellectual disabilities: A meta-analysis of single-case studies. *Research in Developmental Disabilities*, 32, 911–923.
- Department of Health. (2014). Positive and proactive care: Reducing the need for restrictive intervention. Retrieved from <https://www.gov.uk/government/publications/positive-and-proactive-care-reducing-restrictive-interventions>
- Epstein, M., Atkins, M., Cullinan, D., Kutash, K., & Weaver, R. (2008). *Reducing behavior problems in the elementary school classroom: A practice guide*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://ies.ed.gov/ncee/wwc/publications/practiceguides>
- Favell, J. E., McGimsey, J. F., & Schell, R. M. (1982). Treatment of self-injury by providing alternate sensory activities. *Analysis and Intervention in Developmental Disabilities*, 2, 83–104.
- Ferron, J. (2002). Reconsidering the use of general linear model with single-case data. *Behavior Research Methods*, 34, 324–331.
- Fisher, W. W., Piazza, C. C., Bowman, L. G., Hanley, G. P., & Adelinis, J. D. (1997). Direct and collateral effects of restraint and restraint fading. *Journal of Applied Behavior Analysis*, 30, 105–120.
- Garcia, D., & Smith, R. G. (1999). Using analog baselines to assess the effects of naltrexone on self-injurious behavior. *Research in Developmental Disabilities*, 20, 1–21.
- Gedye, A. (1990). Dietary increase in serotonin reduces self-injurious behaviour in a Down's syndrome adult. *Journal of Intellectual Disability Research*, 34, 195–203.
- Gormez, A., Rana, F., & Varghese, S. (2014). Pharmacological interventions for self injurious behaviour in adults with intellectual disabilities: Abridged republication of a Cochrane systematic review. *Journal of Psychopharmacology*, 28, 624–632.
- Haevaert, M., Saenen, L., Maes, B., & Onghena, P. (2015). Comparing the percentage of non-overlapping data approach and the hierarchical linear modeling approach for synthesizing single-case studies in autism research. *Research in Autism Spectrum Disorders*, 11, 112–125.
- Hammock, R. G., Schroeder, S. R., & Levine, W. R. (1995). The effect of clozapine on self-injurious behavior. *Journal of Autism and Developmental Disorders*, 25, 611–626.
- Handling methodological issues in the analysis of single-subject experimental design data*. (2014, April). Structured poster session at the annual meeting of the American Educational Research Association, Philadelphia, PA.
- Hanley, G. P., Piazza, C. C., Fisher, W. W., & Adelinis, J. D. (1997). Stimulus control and resistance to extinction in attention-maintained SIB. *Research in Developmental Disabilities*, 18, 251–260.
- Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Orlando, FL: Academic press.
- Hellings, J. A., Kelley, L. A., Gabrielli, W. F., Kilgore, E., & Shah, P. (1996). Sertraline response in adults with mental retardation and autistic disorder. *Journal of Clinical Psychiatry*, 53, 333–336.
- Heron, T. E. (1978). Punishment: A review of the literature with implications for the teacher of mainstreamed children. *The Journal of Special Education*, 12, 243–252.
- Hosp, M. K., Hosp, J. L., & Howell, K. W. (2016). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York: Guilford Press.
- Houten, R., Axelrod, S., Bailey, J. S., Favell, J. E., Foux, R. M., Iwata, B. A., & Lovaas, O. I. (1988). The right to effective behavioral treatment. *Journal of Applied Behavior Analysis*, 21, 381–384.
- Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004).
- Iwata, B. A. (1988). The development and adoption of controversial default technologies. *The Behavior Analyst*, 11, 149–157.
- 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, 197–209. (Reprinted from *Analysis and Intervention in Developmental Disabilities*, 2, 3–20, 1982).
- Jenson, W. R., Clark, E., Kircher, J. C., & Kristjansson, S. D. (2007). Statistical reform: Evidence-based practice, meta-

- analyses, and single subject designs. *Psychology in the Schools*, 44, 483–493. doi:10.1002/pits.20240
- Kahng, S., Iwata, B. A., DeLeon, I. G., & Worsdell, A. S. (1997). Evaluation of the “control over reinforcement” component in functional communication training. *Journal of Applied Behavior Analysis*, 30, 267–277. doi:10.1901/jaba.1997.30-267
- Kahng, S. W., Iwata, B. A., & Lewin, A. B. (2002). Behavioral treatment of self-injury, 1964 to 2000. *American Journal on Mental Retardation*, 107, 212–221.
- Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., . . . Shadish, W. R. (2012). Single-case intervention research design standards. *Remedial and Special Education*, 34, 26–38.
- Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods*, 15, 124–144.
- Kratochwill, T. R., & Levin, J. R. (2014). *Single-case intervention research: Methodological and statistical advances*. Washington, DC: American Psychological Association.
- Lemons, C. J., Fuchs, D., Gilbert, J. K., & Fuchs, L. S. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher*, 43, 242–252.
- Lucero, W. J., Frieman, J., Spoering, K., & Fehrenbacher, J. (1976). Comparison of three procedures in reducing self-injurious behavior. *American Journal of Mental Deficiency*, 80, 548–554.
- Mace, C. F. (1994). The significance and future of functional analysis methodologies. *Journal of Applied Behavior Analysis*, 27, 385–392.
- Machalicek, W., O'Reilly, M. F., Beretvas, N., Sigafos, J., & Lancioni, G. E. (2007). A review of interventions to reduce challenging behavior in school settings for students with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 1, 229–246.
- Mahatmya, D., Zobel, A., & Valdovinos, M. G. (2008). Treatment approaches for self-injurious behavior in individuals with autism: Behavioral and pharmacological methods. *Journal of Early and Intensive Behavior Intervention*, 5, 106–118.
- Marquis J. G., Horner R. H., Carr E. G., Turnbull A. P., Thompson M., & Behrens G. A., (2000). A meta-analysis of positive behavior support. In Gerston R. M., & Schiller E. P. (Eds.), *Contemporary special education research: Syntheses of the knowledge base on critical instructional issues* (pp. 137–178). Mahwah, NJ: Erlbaum.
- Matson, J. L., Bambrugge, E. A., Mayville, J. P., Bielecki, J., Kuhn, D., Smalls, Y., . . . Logan, J. R. (2000). Psychopharmacology and mental retardation: A 10 year review (1990–1999). *Research in Developmental Disabilities*, 21, 263–296.
- Matson, J. L., & LoVullo, S. V. (2008). A review of behavioral treatments for self-injurious behaviors of persons with autism spectrum disorders. *Behavior Modification*, 32, 61–76.
- Matson, J. L., & Taras, M. E. (1989). A 20 year review of punishment and alternative methods to treat problem behaviors in developmentally delayed persons. *Research in Developmental Disabilities*, 10, 85–104.
- Methodological dilemmas encountered in analyses and meta-analyses of single-case research*. (2012, April). Structured poster session at the annual meeting of the American Educational Research Association, Vancouver, BC.
- Methodological issues in the analysis and meta-analysis of single-subject experimental design data*. (2013, April). Structured poster session at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Moeyaert M., Ferron J., Beretvas S., Van den Noortgate W. (2014). From a single-level analysis to a multilevel analysis of single-case experimental designs. *Journal of School Psychology*, 52, 191–211. doi:10.1016/j.jsp.2013.11.003
- Moeyaert, M., Maggin, D., & Verkuilen, J. (2016). Reliability, validity, and usability of data extraction programs for single-case research designs. *Behavior Modification*, 40, 874–900. doi:10.1177/0145445516645763
- Moeyaert, M., Ugille, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2013). The three-level synthesis of standardized single-subject experimental data: A Monte Carlo simulation study. *Multivariate Behavioral Research*, 48, 719–748.
- Moeyaert, M., Ugille, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2014). Three-level analysis of single-case experimental data: Empirical validation. *The Journal of Experimental Education*, 82, 1–21.
- The National Autism Center. (2009). *The national standards project—Addressing the need for evidenced-based practice guidelines for autism spectrum disorders*. Randolph, MA: National Autism Center.
- Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, 71, 137–148.
- Prangnell, S. J. (2009). Behavioural interventions for self-injurious behavior: A review of recent evidence (1998–2008). *British Journal of Learning Disabilities*, 38, 259–270.
- Pustejovsky, J. E. (2015). Measurement-comparable effect sizes for single-case studies of free-operant behavior. *Psychological Methods*, 20, 342–359. doi:10.1037/met0000019
- Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, 5, 199–213.
- Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Richards, C., Oliver, C., Nelson, L., & Moss, J. (2012). Self-injurious behaviour in individuals with autism spectrum disorder and intellectual disability. *Journal of Intellectual Disability Research*, 56, 476–489.
- Rohatgi, A. (2016). WebPlotDigitizer (Version 3.1) [Software]. Available from <http://arohatgi.info/WebPlotDigitizer/app/>
- Royal College of Psychiatrists, British Psychological Society and Royal College of Speech and Language Therapists.

- (2007). *Challenging behaviour: Clinical and service guidelines for supporting people with learning disabilities who are at risk of receiving abusive or restrictive practices*. Retrieved from <http://www.rcpsych.ac.uk/files/pdfversion/cr144.pdf>
- Salzberg, C. S., Strain, P. S., & Baer, D. M. (1987). Meta-analysis for single-subject research: When does it clarify, when does it obscure? *Remedial and Special Education, 8*, 43–48.
- Sandman, C. A., Barron, J. L., & Colman, H. (1990). An orally administered opiate blocker, naltrexone, attenuates self-injurious behavior. *American Journal on Mental Retardation, 95*, 93–102.
- Schalock, R. L. (2004). The concept of quality of life: What we know and do not know. *Journal of Intellectual Disability Research, 48*, 203–216.
- Schroeder, S. R., Schroeder, C. S., Smith, B., & Dalldorf, J. (1978). Prevalence of self-injurious behaviors in a large state facility for the retarded: A three-year follow up study. *Journal of Autism and Schizophrenia, 8*, 261–269.
- Scotti, J. R., Evans, I. M., & Meyer, L. H. (1991). A meta-analysis of intervention research with problem behavior: Treatment, validity, and standards of practice. *American Journal on Mental Retardation, 96*, 233–256.
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research. *Remedial and Special Education, 8*, 24–33.
- Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25 past, present, and future trends in summarizing single-subject research. *Remedial and Special Education, 34*, 9–19.
- Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention, 2*, 188–196.
- Sigafoos, J., Arthur, M., & O'Reilly, M. (2003). *Challenging behaviour and developmental disability*. Hoboken, NJ: Wiley.
- Singer, J. D., & Willet, J. B. (2003). *A framework for investigating change over time*. In Singer and Willet (Eds.). *Applied longitudinal data analysis: Modeling change and event occurrence* (pp. 3–15). New York: Oxford University Press.
- Singh, N. N., Dawson, M. J., & Gregory, P. R. (1980). Self-Injury in the profoundly retarded: clinically significant versus therapeutic control. *Journal of Intellectual Disability Research, 24*, 87–97. doi:10.1111/j.1365-2788.1980.tb00061.x
- Single-case experimental designs: Developments in statistical analysis, effect size metrics, and meta-analysis methods*. (2015, April). Structured poster session at the annual meeting of the American Educational Research Association, Chicago, IL.
- Sohanpal, S. K., Deb, S., Thomas, C., Soni, R., Lenôtre, L., & Unwin, G. (2007). The effectiveness of antidepressant medication in the management of behaviour problems in adults with intellectual disabilities: A systematic review. *Journal of Intellectual Disability Research, 51*, 750–765.
- Symons, F. J., Thompson, A., & Rodriguez, M. C. (2004). Self-injurious behavior and the efficacy of naltrexone treatment: A quantitative synthesis. *Mental Retardation and Developmental Disabilities Research Reviews, 10*, 193–200.
- Tate, B. G., & Baroff, G. S. (1966). Aversive control of self-injurious behavior in a psychotic. *Behaviour Research and Therapy, 4*, 281–287.
- Totsika, V., Toogood, S., Hastings, R. P., & Lewis, S. (2008). Persistence of challenging behaviour in adults with intellectual disabilities over a period of 11 years. *Journal of Intellectual Disability Research, 52*, 446–457.
- Ugille, M., Moeyaert, M., Beretvas, S. N., Ferron, J., & Van den Noortgate, W. (2012). Multilevel meta-analysis of single-subject experimental designs: A simulation study. *Behavior Research Methods, 44*, 1244–1254.
- Van den Noortgate, W., & Onghena, P. (2003). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments & Computers, 35*(1), 1–10.
- Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence-Based Communication Assessment and Intervention, 2*, 142–151.
- Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review, 20*, 158–177.
- White, O. R. (1987). Some comments concerning “The quantitative synthesis of single-subject research”. *Remedial and Special Education, 8*, 34–39.
- Worlery, M., Baliley, D., & Sugai, G. (1998). *Effective teaching principles and procedures for applied behavior analysis with exceptional students*. Boston, MA: Allyn & Bacon.
- Yates, T. M. (2004). The developmental psychopathology of self-injurious behavior: Compensatory regulation in post-traumatic adaptation. *Clinical Psychology Review, 24*, 35–74.

Appendix

- Adelinis, J. D., Piazza, C. C., Fisher, W. W., & Hanley, G. P. (1997). The establishing effects of client location on self-injurious behavior. *Research in Developmental Disabilities, 18*, 383–391.
- Altman, K., Haavik, S., & Cook, J. W. (1978). Punishment of self-injurious behavior in natural settings using contingent aromatic ammonia. *Behaviour Research and Therapy, 16*, 85–96.
- Altmeyer, B. K., Williams, D. E., & Sams, V. (1985). Treatment of severe self-injurious and aggressive biting. *Journal of Behavior Therapy and Experimental Psychiatry, 16*, 169–172.
- Anderson, S., Hanson, R., Malecha, M., Oftelie, A., Erickson, C., & Clark, J. M. (1997). The effectiveness of Naltrexone in treating task attending, aggression, self-injury, and stereotypic mannerisms of six young males with autism or Pervasive developmental disorders. *Journal of Developmental and Physical Disabilities, 9*, 211–221.

- Augustine, A., & Cipani, E. (1983). Treating self-injurious behavior. *Child & Family Behavior Therapy, 4*, 53–69.
- Ball, T. S., Campbell, R., & Barkemeyer, R. (1980). Air splints applied to control self-injurious finger sucking in profoundly retarded individuals. *Journal of Behavior Therapy and Experimental Psychiatry, 11*, 267–271.
- Ball, T. S., Datta, P. C., Rios, M., & Constantine, C. (1985). Flexible arm splints in the control of a Lesch-Nyhan victim's finger biting and a profoundly retarded client's finger sucking. *Journal of Autism and Developmental Disorders, 15*, 177–184.
- Banda, D. R., McAfee, J. K., & Hart, S. L. (2009). Decreasing self-injurious behavior in a student with autism and Tourette Syndrome through positive attention and extinction. *Child & Family Behavior Therapy, 31*, 144–156.
- Banda, D. R., McAfee, J. K., & Hart, S. L. (2012). Decreasing self-injurious behavior and fading self restraint in a student with autism and Tourette Syndrome. *Behavioral Interventions, 27*, 164–174.
- Barkley, R. A., & Zupnick, S. (1976). Reduction of stereotypic body contortions using physical restraint and dro. *Journal of Behavior Therapy and Experimental Psychiatry, 7*, 167–170.
- Barrett, R. P., Payton, J. B., & Burkhart, J. E. (1988). Treatment of self-injury and disruptive behavior with carbamazepine (Tegretol) and behavior therapy. *Journal of the Multihandicapped Person, 1*, 79–91.
- Baumeister, A. A., & Baumeister, A. A. (1978). Suppression of repetitive self-injurious behavior by contingent inhalation of aromatic ammonia. *Journal of Autism and Childhood Schizophrenia, 8*, 71–77.
- Blankenship, M. D., & Lamberts, F. (1989). Helmet restraint and visual screening as treatment for self-injurious behavior in persons who have profound mental retardation. *Behavioral Interventions, 4*, 253–265.
- Borreson, P. M. (1980). The elimination of a self-injurious avoidance response through a forced running consequence. *Mental Retardation, 18*, 73–77.
- Braithwaite, K. L., & Richdale, A. L. (2000). Functional communication training to replace challenging behaviors across two behavioral outcomes. *Behavioral Interventions, 15*, 21–36.
- Carter, S. L. (2009). Use of treatment analysis following ambiguous functional analysis results. *Behavioral Interventions, 24*, 205–213.
- Cavalier, A. R., & Ferretti, R. P. (1980). Stereotyped behaviour, alternative behaviour and collateral effects: A comparison of four intervention procedures. *Journal of Mental Deficiency Research, 24*, 219–230.
- Cipani, E., Brendlinger, J., McDowell, L., & Usher, S. (1991). Continuous vs. intermittent punishment: A case study. *Journal of Developmental and Physical Disabilities, 3*, 147–156.
- Clauser, B., & Gould, K. (1988). Visual screening as a restrictive procedure: An examination of generalization and duration. *Behavioral Interventions, 3*, 51–61.
- Conley, O. S., & Wolery, M. R. (1980). Treatment by over-correction of self-injurious eye gouging in preschool blind children. *Journal of Behavior Therapy and Experimental Psychiatry, 11*, 121–125.
- Corte, H. E., Wolf, M. M., & Locke, B. J. (1971). A comparison of procedures for eliminating self-injurious behavior of retarded adolescents. *Journal of Applied Behavior Analysis, 4*, 201–213.
- Davis, T. N., Dacus, S., Strickland, E., Copeland, D., Chan, J. M., Blenden, K., . . . Christian, K. (2013). The effects of a weighted vest on aggressive and self-injurious behavior in a child with autism. *Developmental Neurorehabilitation, 16*, 210–215.
- Davis, T. N., Dacus, S., Strickland, E., Machalick, W., & Coviello, L. (2013). Reduction of automatically maintained self-injurious behavior utilizing noncontingent matched stimuli. *Developmental Neurorehabilitation, 16*, 166–171.
- Day, H. M., Horner, R. H., & O'Neill, R. E. (1994). Multiple functions of problem behaviors: Assessment and intervention. *Journal of Applied Behavior Analysis, 27*, 279–289.
- Day, R. M., Rea, J. A., Schussler, N. G., Larsen, S. E., & Johnson, W. L. (1988). A functionally based approach to the treatment of self-injurious behavior. *Behavior Modification, 12*, 565–589.
- DeLeon, I. G., Anders, B. M., Rodriguez-Catter, V., & Neidert, P. L. (2000). The effects of noncontingent access to single versus multiple-stimulus sets on self-injurious behavior. *Journal of Applied Behavior Analysis, 33*, 623–626.
- DeLeon, I. G., Fisher, W. W., & Marhefka, J.-M. (2004). Decreasing self-injurious behavior associated with awakening in a child with autism and developmental delays. *Behavioral Interventions, 19*, 111–119.
- Demanche, J., & Chok, J. T. (2013). The use of wrist weights and vibratory stimulation to treat self-injurious behavior. *Journal of Developmental and Physical Disabilities, 25*, 79–90.
- Denys, A. D., & Graham, B. (1978). Effective treatment of self-injurious behavior through a forced arm exercise. *American Journal of Mental Deficiency, 82*, 433–439.
- Derby, K. M., Fisher, W. W., & Piazza, C. C. (1996). The effects of contingent and noncontingent attention on self-injury and self-restraint. *Journal of Applied Behavior Analysis, 29*, 107–110.
- Dorsey, M. F., Iwata, B. A., Ong, P., & McSween, T. E. (1980). Treatment of self-injurious behavior using a water mist: Initial response suppression and generalization. *Journal of Applied Behavior Analysis, 13*, 343–353.
- Dorsey, M. F., Iwata, B. A., Reid, D. H., & Davis, P. A. (1982). Protective equipment: Continuous and contingent application in the treatment of self-injurious behavior. *Journal of Applied Behavior Analysis, 15*, 217–230.
- Drabman, R. S., Ross, J. M., Lynd, R. S., & Cordua, G. D. (1978). Retarded children as observers, mediators, and generalization programmers using an icing procedure. *Behavior Modification, 2*, 371–385.
- Dura, J. R., Mulick, J. A., & Hammer, D. (1988). Rapid clinical evaluation of sensory integrative therapy for self-injurious behavior. *Mental Retardation, 26*, 83–87.
- Favell, J. E., McGimsey, J. F., & Jones, M. L. (1978). The use of physical restraint in the treatment of self-injury and as

- positive reinforcement. *Journal of Applied Behavior Analysis*, 11, 225–241.
- Fisher, W. W., Bowman, L. G., Thompson, R. H., Contrucci, S. A., Burd, L., & Alon, G. (1998). Reductions in self-injury produced by transcutaneous electrical nerve stimulation. *Journal of Applied Behavior Analysis*, 31, 493–496.
- Fisher, W. W., Grace, N. C., & Murphy, C. (1996). Further analysis of the relationship between self-injury and self-restraint. *Journal of Applied Behavior Analysis*, 29, 103–106.
- Fisher, W. W., Piazza, C. C., Bowman, L. G., Hanley, G. P., & Adelinis, J. D. (1997). Direct and collateral effects of restraints and restraint fading. *Journal of Applied Behavior Analysis*, 30, 105.
- Garcia, D., & Smith, R. G. (1999). Using analog baselines to assess the effects of naltrexone on self-injurious behavior. *Research in Developmental Disabilities*, 20, 1–21.
- Gaylord-Ross, R. J., Weeks, M., & Lipner, C. (1980). An analysis of antecedent, response, and consequence events in the treatment of self-injurious behavior. *Education and Training of the Mentally Retarded*, 15, 35–42.
- Goh, H.-L., & Iwata, B. A. (1994). Behavioral persistence and variability during extinction of self-injury maintained by escape. *Journal of Applied Behavior Analysis*, 27, 173–174.
- Goh, H.-L., Iwata, B. A., & DeLeon, I. G. (2000). Competition between noncontingent and contingent reinforcement schedules during response acquisition. *Journal of Applied Behavior Analysis*, 33, 195–205.
- Hammock, R. G., Schroeder, S. R., & Levine, W. R. (1995). The effect of clozapine on self-injurious behavior. *Journal of Autism and Developmental Disorders*, 25, 611–626.
- Hanley, G. P., Piazza, C. C., Fisher, W. W., & Adelinis, J. D. (1997). Stimulus control and resistance to extinction in attention-maintained SIB. *Research in Developmental Disabilities*, 18, 251–260.
- Hanley, G. P., Piazza, C. C., Keeney, K. M., Blakeley-Smith, A. B., & Worsdell, A. S. (1998). Effects of wrist weights on self-injurious and adaptive behaviors. *Journal of Applied Behavior Analysis*, 31, 307–310.
- Harris, S. L., & Romanczyk, R. G. (1976). Treating self-injurious behavior of a retarded child by overcorrection. *Behavior Therapy*, 7, 235–239.
- Healey, J. J., Ahearn, W. H., Graff, R. B., & Libby, M. E. (2001). Extended analysis and treatment of self-injurious behavior. *Behavioral Interventions*, 16, 181–195.
- Herman, B. H., Kathryn Hammock, M., Arthur-Smith, A., Egan, J., Chatoor, I., Werner, A., & Zelnik, N. (1987). Naltrexone decreases self-injurious behavior. *Annals of Neurology*, 22, 550–552.
- Horner, R. H., Day, H. M., & Day, J. R. (1997). Using neutralizing routines to reduce problem behaviors. *Journal of Applied Behavior Analysis*, 30, 601–614.
- Horner, R. H., Day, H. M., Sprague, J. R., O'Brien, M., & Heathfield, L. T. (1991). Interspersed requests: A nonaversive procedure for reducing aggression and self-injury during instruction. *Journal of Applied Behavior Analysis*, 24, 265–278.
- Hosp, M. K., Hosp, J. L., & Howell, K. W. (2016). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York, NY: Guilford Publications.
- Irvin, D. S., Realon, R. E., Hartley, J. R., Phillips, J. F., Bradley, F., & Daly, M. (1996). The treatment of self-injurious hand mouthing by using a multi-component intervention with individuals positioned in a small group. *Journal of Developmental and Physical Disabilities*, 8, 43–59.
- 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, 197–209.
- Jenson, W. R., Rovner, L., Cameron, S., Petersen, B. P., & Kesler, J. (1985). Reduction of self-injurious behavior in an autistic girl using a multifaceted treatment program. *Journal of Behavior Therapy and Experimental Psychiatry*, 16, 77–80.
- John, R. L. (1978). Reducing self-injurious behavior by facial screening. *American Journal of Mental Deficiency*, 82, 510–513.
- Jones, L. J., Singhm, N. N., & Kendall, K. A. (1991). Comparative effects of gentle teaching and visual screening on self-injurious behaviour. *Journal of Intellectual Disability Research*, 35, 37–47.
- Kahng, S., Iwata, B. A., DeLeon, I. G., & Wallace, M. D. (2000). A comparison of procedures for programming noncontingent reinforcement schedules. *Journal of Applied Behavior Analysis*, 33, 223–231.
- Kahng, S., Iwata, B. A., DeLeon, I. G., & Worsdell, A. S. (1997). Evaluation of the “control over reinforcement” component in functional communication training. *Journal of Applied Behavior Analysis*, 30, 267–277.
- Kennedy, C. H., & Souza, G. (1995). Functional analysis and treatment of eye poking. *Journal of Applied Behavior Analysis*, 28, 27–37.
- Kerth, D. M., Progar, P. R., & Morales, S. (2009). The effects of non-contingent self-restraint on self-injury. *Journal of Applied Research in Intellectual Disabilities*, 22, 187–193.
- Konarski, E. A., & Johnson, M. R. (1989). The use of brief restraint plus reinforcement to treat self-injurious behavior. *Behavioral Interventions*, 4, 45–52.
- Kuhn, D. E., DeLeon, I. G., Fisher, W. W., & Wilke, A. E. (1999). Clarifying an ambiguous functional analysis with matched and mismatched extinction procedures. *Journal of Applied Behavior Analysis*, 32, 99–102.
- Ladd, M. V., Luiselli, J. K., & Baker, L. (2009). Continuous access to competing stimulation as intervention for self-injurious skin picking in a child with autism. *Child & Family Behavior Therapy*, 31, 54–60.
- Lalli, J. S., Mace, F. C., Livezey, K., & Kates, K. (1998). Assessment of stimulus generalization gradients in the treatment of self-injurious behavior. *Journal of Applied Behavior Analysis*, 31, 479–483.
- Lerman, D. C., Iwata, B. A., Shore, B. A., & DeLeon, I. G. (1997). Effects of intermittent punishment on self-injurious behavior: An evaluation of schedule thinning. *Journal of Applied Behavior Analysis*, 30, 187–201.
- Lerman, D. C., Iwata, B. A., Shore, B. A., & Kahng, S. W. (1996). Responding maintained by intermittent reinforcement: Implications for the use of extinction with problem

- behavior in clinical settings. *Journal of Applied Behavior Analysis*, 29, 153–171.
- Lerman, D. C., Iwata, B. A., Smith, R. G., Zarcone, J. R., & Vollmer, T. R. (1994). Transfer of behavioral function as a contributing factor in treatment relapse. *Journal of Applied Behavior Analysis*, 27, 357–370.
- Lindauer, S. E., DeLeon, I. G., & Fisher, W. W. (1999). Decreasing signs of negative affect and correlated self-injury in an individual with mental retardation and mood disturbances. *Journal of Applied Behavior Analysis*, 32, 103–106.
- Lindberg, J. S., Iwata, B. A., Kahng, S., & DeLeon, I. G. (1999). DRO contingencies: An analysis of variable-momentary schedules. *Journal of Applied Behavior Analysis*, 32, 123–136.
- Lindberg, J. S., Iwata, B. A., Roscoe, E. M., Worsdell, A. S., & Hanley, G. P. (2003). Treatment efficacy of noncontingent reinforcement during brief and extended application. *Journal of Applied Behavior Analysis*, 36(1), 1–19.
- Linn, D. M., Rojahn, J., Helsel, W. J., & Dixon, J. (1988). Acute effects of transcutaneous electric nerve stimulation on self-injurious behavior. *Journal of the Multihandicapped Person*, 1, 105–119.
- Linscheid, T. R., Iwata, B. A., Ricketts, R. W., Williams, D. E., & Griffin, J. C. (1990). Clinical evaluation of the self-injurious behavior inhibiting system (SIBIS). *Journal of Applied Behavior Analysis*, 23, 53–78.
- Linscheid, T. R., Pejeau, C., Cohen, S., & Footo-Lenz, M. (1994). Positive side effects in the treatment of SIB using the self-injurious behavior inhibiting system (SIBIS): Implications for operant and biochemical explanations of SIB. *Research in Developmental Disabilities*, 15, 81–90.
- Linscheid, T. R., & Reichenbach, H. (2002). Multiple factors in the long-term effectiveness of contingent electric shock treatment for self-injurious behavior: A case example. *Research in Developmental Disabilities*, 23, 161–177.
- Lockwood, K., & Williams, D. E. (1994). Treatment and extended follow-up of chronic hand mouthing. *Journal of Behavior Therapy and Experimental Psychiatry*, 25, 161–169.
- Lucero, W. J., Frieman, J., Spoering, K., & Fehrenbacher, J. (1976). Comparison of three procedures in reducing self-injurious behavior. *American Journal of Mental Deficiency*, 80, 548–554.
- Luiselli, J. K., Beltis, J. A., & Bass, J. (1989). Clinical analysis of naltrexone in the treatment of self-injurious behavior. *Journal of the Multihandicapped Person*, 2, 43–50.
- Luiselli, J. K., Suskin, L., & McPhee, D. F. (1981). Continuous and intermittent application of overcorrection in a self-injurious autistic child: Alternating treatments design analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 12, 355–358.
- Barton, L. E., & Lagrow, S. J. (1983). Reducing self-injurious and aggressive behavior in deaf-blind persons through overcorrection. *Journal of Visual Impairment & Blindness*, 77, 421–424.
- Mason, S. A., & Iwata, B. A. (1990). Artifactual effects of sensory-integrative therapy on self-injurious behavior. *Journal of Applied Behavior Analysis*, 23, 361–370.
- Matson, J. L., & Keyes, J. (1988). Contingent reinforcement and contingent restraint to treat severe aggression and self-injury in mentally retarded and autistic adults. *Journal of the Multihandicapped Person*, 1, 141–153.
- Matson, J. L., & Keyes, J. B. (1990). A comparison of DRO to movement suppression time-out and DRO with two self-injurious and aggressive mentally retarded adults. *Research in Developmental Disabilities*, 11, 111–120.
- Mayhew, G., & Harris, F. (1979). Decreasing self-injurious behavior: Punishment with citric acid and reinforcement of alternative behavior. *Behavior Modification*, 3, 322–336.
- Mazaleski, J. L., Iwata, B. A., Rodgers, T. A., Vollmer, T. R., & Zarcone, J. R. (1994). Protective equipment as treatment for stereotypic hand mouthing: Sensory extinction or punishment effects? *Journal of Applied Behavior Analysis*, 27, 345–355.
- Mazaleski, J. L., Iwata, B. A., Vollmer, T. R., Zarcone, J. R., & Smith, R. G. (1993). Analysis of the reinforcement and extinction components in DRO contingencies with self-injury. *Journal of Applied Behavior Analysis*, 26, 143–156.
- McEntee, J. E., Parker, E. H., Brown, M. B., & Poulson, R. L. (1996). The effects of response interruption, DRO and positive reinforcement on the reduction of hand-mouth behavior. *Behavioral Interventions*, 11, 163–170.
- McFarlain, R. A., Andy, O. J., Scott, R. W., & Wheatley, M. L. (1975). Suppression of headbanging on the ward. *Psychological Reports*, 36, 315–321.
- Measel, C. J., & Alfieri, P. A. (1976). Treatment of self-injurious behavior by a combination of reinforcement for incompatible behavior and overcorrection. *American Journal of Mental Deficiency*, 81, 147–153.
- Moore, J. W., Fisher, W. W., & Pennington, A. (2004). Systematic application and removal of protective equipment in the assessment of multiple topographies of self-injury. *Journal of Applied Behavior Analysis*, 37, 73–77.
- Morrison, D. (1972). Issues in the application of reinforcement theory in the treatment of a child's self-injurious behavior. *Psychotherapy: Theory, Research & Practice*, 9, 40–45.
- Mudford, O. C., Boundy, K., & Murray, A. D. (1995). Therapeutic shock device (TSD): Clinical evaluation with self-injurious behaviors. *Research in Developmental Disabilities*, 16, 253–267.
- Myers, D. V. (1975). Case histories and shorter communications. Extinction, DRO, and response-cost procedures for eliminating self-injurious behavior: A case study. *Behavioral Research & Therapy*, 13, 189–191.
- Neufeld, A., & Fantuzzo, J. W. (1987). Treatment of severe self-injurious behavior by the mentally retarded using the bubble helmet and differential reinforcement procedures. *Journal of Behavior Therapy and Experimental Psychiatry*, 18, 127–136.
- Nunes, D. L., Murphy, R. J., & Ruprecht, M. L. (1977). Reducing self-injurious behavior of severely retarded individuals through withdrawal of reinforcement procedures. *Behavior Modification*, 1, 499–516.
- Oliver, C., Hall, S., Hales, J., Murphy, G., & Watts, D. (1998). The treatment of severe self-injurious behavior by the systematic fading of restraints: Effects on self-injury, self-restraint, adaptive behavior, and behavioral correlates

- of affect. *Research in Developmental Disabilities*, 19, 143–165.
- Pace, G. M., Iwata, B. A., Cowdery, G. E., Andree, P. J., & McIntyre, T. (1993). Stimulus (instructional) fading during extinction of self-injurious escape behavior. *Journal of Applied Behavior Analysis*, 26, 205–212.
- Parrish, J. M., Iwata, B. A., Dorsey, M. F., Bunck, T. J., & Slifer, K. J. (1985). Behavior analysis, program development, and transfer of control in the treatment of self-injury. *Journal of Behavior Therapy and Experimental Psychiatry*, 16, 159–167.
- Patel, M. R., Carr, J. E., Kim, C., Robles, A., & Eastridge, D. (2000). Functional analysis of aberrant behavior maintained by automatic reinforcement: Assessments of specific sensory reinforcers. *Research in Developmental Disabilities*, 21, 393–407.
- Peine, H. A., Liu, L., Blakelock, H., Jenson, W. R., & Osborne, J. G. (1991). The use of contingent water misting in the treatment of self-choking. *Journal of Behavior Therapy and Experimental Psychiatry*, 22, 225–231.
- Realon, R. E., Favell, J. E., & Cacace, S. (1995). An economical, humane, and effective method for short-term suppression of hand mouthing. *Behavioral Interventions*, 10, 141–147.
- Realon, R. E., & Konarski, E. A. (1993). Using decelerative contingencies to reduce the self-injurious behavior of people with multiple handicaps: The effects of response satiation? *Research in Developmental Disabilities*, 14, 341–357.
- Reese, R. M. (1997). Biobehavior analysis of self-injurious behavior in a person with profound handicaps. *Focus on Autism and Other Developmental Disabilities*, 12, 87–94.
- Repp, A. C., Felce, D., & Barton, L. E. (1988). Basing the treatment of stereotypic and self-injurious behaviors on hypotheses of their causes. *Journal of Applied Behavior Analysis*, 21, 281–289.
- Roberts, M. L., Mace, F. C., & Daggett, J. A. (1995). Preliminary comparison of two negative reinforcement schedules to reduce self-injury. *Journal of Applied Behavior Analysis*, 28, 579–580.
- Roscoe, E. M., Iwata, B. A., & Goh, H. L. (1998). A comparison of noncontingent reinforcement and sensory extinction as treatments for self-injurious behavior. *Journal of Applied Behavior Analysis*, 31, 635–646.
- Rose, V., Sloop, W., & Baker, P. (1980). Elimination of chronic self-injurious behavior by withdrawal of staff attention. *Psychological Reports*, 46, 327–330.
- Sandler, A. G., & McLain, S. C. (2007). Use of noncontingent tactile and vestibular stimulation in the treatment of self-injury: An interdisciplinary study. *Journal of Developmental and Physical Disabilities*, 19, 543–555.
- Shirley, M. J., Iwata, B. A., Kahng, S. W., Mazaleski, J. L., & Lerman, D. C. (1997). Does functional communication training compete with ongoing contingencies of reinforcement? An analysis during response acquisition and maintenance. *Journal of Applied Behavior Analysis*, 30, 93–104.
- Singh, N. N. (1980). The effects of facial screening on infant self-injury. *Journal of Behavior Therapy and Experimental Psychiatry*, 11, 131–134.
- Singh, N. N., Dawson, M. J., & Gregory, P. R. (1980). Self-Injury in the profoundly retarded: Clinically significant versus therapeutic control. *Journal of Intellectual Disability Research*, 24, 87–97.
- Singh, N. N., Dawson, M. J., & Manning, P. J. (1981). The effects of physical restraint on self-injurious behaviour. *Journal of Intellectual Disability Research*, 25, 207–216.
- Singh, N. N., Watson, J. E., & Winton, A. S. W. (1986). Treating self-injury: Water mist spray versus facial screening or forced arm exercise. *Journal of Applied Behavior Analysis*, 19, 403–410.
- Smith, R. G., Iwata, B. A., Vollmer, T. R., & Zarcone, J. R. (1993). Experimental analysis and treatment of multiply controlled self-injury. *Journal of Applied Behavior Analysis*, 26, 183–196.
- Solnick, J. V., Rincover, A., & Peterson, C. R. (1977). Some determinants of the reinforcing and punishing effects of timeout. *Journal of Applied Behavior Analysis*, 10, 415–424.
- Steege, M. W., Wacker, D. P., Cigrand, K. C., Berg, W. K., Novak, C. G., Reimers, T. M., . . . DeRaad, A. (1990). Use of negative reinforcement in the treatment of self-injurious behavior. *Journal of Applied Behavior Analysis*, 23, 459–467.
- Symons, F. J., Fox, N. D., & Thompson, T. (1998). Functional communication training and naltrexone treatment of self-injurious behaviour: An experimental case report. *Journal of Applied Research in Intellectual Disabilities*, 11, 273–292.
- Tanner, B. A., & Zeiler, M. (1975). Punishment of self-injurious behavior using aromatic ammonia as the aversive stimulus. *Journal of Applied Behavior Analysis*, 8, 53–57.
- Thompson, R. H., Iwata, B. A., Conners, J., & Roscoe, E. M. (1999). Effects of reinforcement for alternative behavior during punishment of self-injury. *Journal of Applied Behavior Analysis*, 32, 317–328.
- Turner, W. D., Realon, R. E., Irvin, D., & Robinson, E. (1996). The effects of implementing program consequences with a group of individuals who engaged in sensory maintained hand mouthing. *Research in Developmental Disabilities*, 17, 311–330.
- Van Houten, R. (1993). The use of wrist weights to reduce self-injury maintained by sensory reinforcement. *Journal of Applied Behavior Analysis*, 26, 197–203.
- Vollmer, T. R., Iwata, B. A., Zarcone, J. R., Smith, R. G., & Mazaleski, J. L. (1993). The role of attention in the treatment of attention-maintained self-injurious behavior: Noncontingent reinforcement and differential reinforcement of other behavior. *Journal of Applied Behavior Analysis*, 26, 9–21.
- Vollmer, T. R., Marcus, B. A., & Ringdahl, J. E. (1995). Noncontingent escape as treatment for self-injurious behavior maintained by negative reinforcement. *Journal of Applied Behavior Analysis*, 28, 15–26.
- Vollmer, T. R., & Vorndran, C. M. (1998). Assessment of self-injurious behavior maintained by access to self-restraint materials. *Journal of Applied Behavior Analysis*, 31, 647–650.

- Watson, T. S., Ray, K. P., Turner, H. S., & Logan, P. (1999). Teacher-implemented functional analysis and treatment: A method for linking assessment to intervention. *School Psychology Review, 28*, 292.
- Weiher, R. G., & Harman, R. E. (1975). The use of omission training to reduce self-injurious behavior in a retarded child. *Behavior Therapy, 6*, 261–268.
- Williams, D. E., Kirkpatrick-Sanchez, S., & Iwata, B. A. (1993). A comparison of shock intensity in the treatment of longstanding and severe self-injurious behavior. *Research in Developmental Disabilities, 14*, 207–219.
- Worsdell, A. S., Iwata, B. A., Hanley, G. P., Thompson, R. H., & Kahng, S. (2000). Effects of continuous and intermittent reinforcement for problem behavior during functional communication training. *Journal of Applied Behavior Analysis, 33*, 167–179.
- Young, J. A., & Wincze, J. P. (1974). The effects of the reinforcement of compatible and incompatible alternative behaviors on the self-injurious and related behaviors of a profoundly retarded female adult. *Behavior Therapy, 5*, 614–623.
- Zarcone, J. R., Iwata, B. A., Smith, R. G., Mazaleski, J. L., & Lerman, D. C. (1994). Reemergence and extinction of self-injurious escape behavior during stimulus (instructional) fading. *Journal of Applied Behavior Analysis, 27*, 307–316.
- Zarcone, J. R., Iwata, B. A., Vollmer, T. R., Jagtiani, S., Smith, R. G., & Mazaleski, J. L. (1993). Extinction of self-injurious escape behavior with and without instructional fading. *Journal of Applied Behavior Analysis, 26*, 353–360.
- Zhou, L., Goff, G. A., & Iwata, B. A. (2000). Effects of increased response effort on self-injury and object manipulation as competing responses. *Journal of Applied Behavior Analysis, 33*, 29–40.