

DISCUSSION

Caregiver involvement in applied behavior-analytic research: A scoping review and discussion

Jessica L. Becraft^{1,2}  | Samantha L. Hardesty^{1,2}  | Kissel J. Goldman¹  |
Lesley A. Shawler³  | Matthew L. Edelstein^{1,2}  | Phillip Orchowicz¹

¹Department of Behavioral Psychology,
Kennedy Krieger Institute, Baltimore,
MD, USA

²Department of Psychiatry and Behavioral
Sciences, Johns Hopkins University School of
Medicine, Baltimore, MD, USA

³School of Psychological and Behavioral
Sciences, Southern Illinois University,
Carbondale, IL, USA

Correspondence

Jessica L. Becraft, Department of Behavioral
Psychology, Kennedy Krieger Institute, 9810
Patuxent Woods Drive Suite B, Columbia,
MD 21046.

Email: becraft@kennedykrieger.org

Editor-in-Chief: John Borrero

Handling Editor: Bridget Taylor

Abstract

We conducted a scoping review to characterize the role of caregiver involvement in behavior-analytic research. We reviewed eight behavioral-learning journals from 2011–2022 for works that included children or caregivers as participants and characterized caregiver involvement as passive (implications for caregivers, input, social validity) and active (implementation, caregiver behavior, training, caregiver-collected data). The review identified 228 studies, and almost all (96.1%; $n = 219$) involved caregivers in some capacity; 94.3% ($n = 215$) had passive involvement (26.8% had only passive involvement; $n = 61$), 69.3% ($n = 158$) had active involvement (1.8% had only active involvement; $n = 4$), and 3.9% ($n = 9$) had neither passive nor active involvement. Involvement generally increased over publication years. The most common types of involvement were implications for caregivers, implementation, and input; caregiver-collected data were rare. We propose considerations when engaging caregivers in research and suggest new avenues of inquiry related to caregivers' treatment objectives and social validity, treatment implementers, and caregiver-collected data.

KEYWORDS

applied behavior analysis, caregivers, clinical endpoints, parents, scoping review

Children are frequent recipients of services based in applied behavior analysis and are commonly included in behavior-analytic research. Nearly 90% of Board Certified Behavior Analysts practice in the areas that presumably serve children¹ (Behavior Analyst Certification Board, n.d.). Children also encompassed most participants of the first 25 years of the *Journal of Applied Behavior Analysis (JABA)* publications (Northup et al., 1993). In a search of *JABA* from 2011 to 2021 using the PsycINFO database, we found similar results, with 72% (655/906) of studies reporting child participants.²

In addition to the child, caregivers³ play a crucial role in their child's treatment. Caregivers often initiate and consent to clinical services or research participation

and contribute to determining clinical necessity for services (e.g., frequency of services). Per the Behavior Analyst Certification Board Ethics Code, behavior analysts have an ethical responsibility to involve caregivers in treatment planning and implementation (Behavior Analyst Certification Board, 2020, Guideline 2.09). Therefore, caregivers may be asked to implement treatment protocols and behavior plans recommended by a behavior analyst, teach other caregivers to implement treatment plans, and collect data to confirm whether treatment effects maintain and generalize. Because caregivers are important in practice, it is important to study caregivers and their involvement in research. Ecological validity is the extent to which the phenomenon under study relates to the "real world" (Fahmie et al., 2023). Fahmie et al. (2023) encourage behavior analysts to systematically program for ecological validity by designing

¹Categories used to calculate this value include autism spectrum disorder (72.1%), education (6.9%), clinical behavior analysis (4.4%), intellectual and development disabilities (2.7%), behavioral pediatrics (1.0%), parent and caregiver training (0.4%), child welfare (0.3%), and corrections and delinquency (0.1%).

²We conducted a rough search using the PsycINFO filters to identify experimental studies (i.e., not discussions or reviews) and studies that were indexed as including child participants.

³For the remainder of this paper, we use the term "caregiver" to refer to the parent or someone who fulfills the role of a parent (e.g., responsible for making treatment decisions for the child).

studies around the participants' natural environments and incorporating input from participants and their caregivers. In other words, caregiver involvement enhances ecological validity.

Caregiver involvement could include providing a priori input on aspects of the study, providing post hoc feedback on the study (social validity), implementing procedures, being taught to implement procedures, having data collected on their behavior, and collecting data on events or behavior during the study or services. However, research discussing and evaluating areas of caregiver involvement is scant. Some research may mention the inclusion of caregivers but does not focus on studying their involvement. For instance, study authors may mention that caregivers were interviewed to identify potential reinforcers, but they may not discuss how frequently caregiver input on reinforcers was solicited. On a larger scale, there are no reviews summarizing how often or what type of input caregivers provide in research.

Understanding of caregiver involvement is incomplete even among well-researched topics. For instance, reviews on caregiver training and implementation indicate that caregivers can be taught to implement behavioral interventions with corresponding improvements in their child's behavior (e.g., Kemmerer et al., 2023; Sun, 2022; Unholz-Bowden et al., 2020). Reviews on caregiver training tend to be narrowly focused on a certain population, treatment procedure, or setting. Kemmerer et al. (2023) focused on caregiver training for individuals with autism spectrum disorder, Sun (2022) on individuals with intellectual or development disabilities (IDD) during behavioral skills training, and Unholz-Bowden et al. (2020) on caregiver training via telehealth. Each of these reviews is important to establish that parent training occurs and can be effective for these populations, interventions, or settings. However, related topics such as the percentage of studies involving caregivers that provide training, the type of training, whether data are collected during training or implementation, and whether caregivers collect data have all received less attention.

Given the interrelatedness of areas of caregiver involvement, research is needed to summarize various areas of caregiver involvement to identify interacting and complementary trends across categories and gaps in the literature with respect to caregiver involvement. Systematic reviews are often viewed as the highest level of evidence because they summarize a large amount of research (Oxford Centre for Evidence-Based Medicine, 2011). In turn, they produce implications for many stakeholders including researchers, practitioners, and policy makers. Reviews allow researchers to identify trends on a given topic, evaluate the effectiveness of interventions, and identify areas for further study. In practice, following the scientist-practitioner model (Dorsey & Harper, 2018), clinicians may turn to research to inform service delivery. For policy makers, reviews can help guide decisions about what types of health care should be provided, which in

turn influence insurance companies and other third-party payers who may reimburse for behavioral services. Scoping reviews, like systematic reviews, provide a synthesis of evidence in a particular area. Systematic reviews typically have a narrow focus or question designed to confirm or challenge a current practice or determine the conditions under which said practice is suitable (e.g., the effect of an intervention across different cultural groups). In contrast, scoping reviews ask broader questions on a topic such as determining the size, variety, and nature of evidence on a topic, summarizing findings when methods or approaches are heterogeneous, and identifying gaps in the literature (Arksey & O'Malley, 2005; Munn et al., 2018; Tricco et al., 2018).

Because caregivers can be valuable participants in research and a considerable amount of behavior-analytic research incorporates children, it would be useful to characterize how caregivers are incorporated in research. In doing so, it would also be possible to identify gaps in the literature that research could address to better inform clinical practice. To date, there are no scoping reviews of the various types of caregiver involvement in behavior-analytic research with children. The purpose of the current study was to (a) determine the nature of caregiver involvement in existing behavior-analytic research with children and (b) identify gaps in the literature where further research on caregiver involvement is warranted. Because we anticipated that there would be a large and heterogeneous sample of evidence related to caregiver involvement, we used a scoping review approach.

METHOD

This scoping review followed the guidelines established by the *Preferred Reporting Items for Systematic Reviews and Meta-Analysis-Scoping Review* (PRISMA-ScR; Tricco et al., 2018; see Supporting Information A for the PRISMA-ScR checklist). The method for identifying, screening, and including articles and planned analyses were specified in advance in a published protocol on the Open Science Framework (<https://osf.io/gS5qx>).

Inclusion criteria

To be included in the current review, articles had to involve either (a) at least one child (i.e., a neurotypical individual under the age of 18 or an individual with a reported IDD under the age of 21) or (b) at least one caregiver who was selected for participation in the study because they were a caregiver (i.e., a parent or someone that was fulfilling a parental role for the child). The child did not need to be the primary participant of the study; that is, studies were eligible for inclusion even if they focused on only the behavior of a caregiver of a child. In addition, studies had to have been published between

2011 and 2022. This range was selected to evaluate the most recent trends in caregiver involvement. To better capture research on practices based in applied behavior analysis, we included articles in which the authors targeted a socially significant behavior in the study (i.e., arbitrary responses in basic or translational investigations were excluded), directly observed at least one child or caregiver response, manipulated an independent variable using a group or single-case design, and published in a journal that commonly publishes behavior-analytic research. Articles were excluded if they did not meet the inclusion criteria, used only nonhuman subjects, or were not primary experimental studies (i.e., reviews, discussions, or secondary data analyses).

For the purposes of this review, journals included: *Behavior Analysis in Practice (BAP)*, *Behavior Analysis: Research and Practice (BARP)*, *Behavioral Interventions (BI)*, *Education and Treatment of Children (ETC)*, *Journal of Applied Behavior Analysis (JABA)*, *Journal of Behavioral Education (JOBE)*, *Journal of Positive Behavior Interventions (JPBI)*, and *The Analysis of Verbal Behavior (TAVB)*. Additional information about how these journals were selected is described in Search and Article Identification below.

Search and article identification

We conducted four electronic searches in PsycINFO and ERIC databases to identify articles. The following Boolean phrases were searched in both databases: (“parent” OR “parents” OR “mother” OR “father” OR caregiver*) N5 (input OR choice OR answer OR provide* OR g*ve OR “fill out” OR “filled out” OR collect* OR conduct* OR “data” OR record* OR implement* OR “run” OR “ran” OR train* OR “taught” OR learn OR “social validity” OR prefer* OR implication OR consider* OR “try”) AND (“intervention” OR “treatment” OR “therapy” OR “services”). The first five search terms were selected to capture a parent or primary caregiver; search terms following “N5” were selected to capture the type of involvement in proximity to the caregiver term (five words apart or fewer); the last four search terms were selected to capture clinical studies. Articles were filtered to include only articles published between 2011 and 2021 (the first three searches did not include 2022).

In the first search conducted in January 2022, we specified that the databases search all text fields, which yielded 921 unique articles. We did not restrict the journals in which studies could be published. We reviewed the full text of these articles, which resulted in 180 articles that met the inclusion criteria. However, upon review, we determined that the search was missing at least 30 articles known to meet the inclusion criteria. To troubleshoot concerns with the first search, we modified settings in the database searches. Specifically, in the second search (also conducted in January 2022), we did not specify where the database should search for the terms (i.e., we left the field blank), which resulted in nearly 90,000 articles. This search

included the missing articles but required further refinement due to the high number of irrelevant articles.

To both obtain the most relevant articles (as in the first search) and a large breadth of articles (as in the second search), the third search (conducted October 2022) included search terms for the journal source from the list of journals described in the inclusion criteria. Because this was a scoping review focused on caregiver involvement in behavior-analytic research, we limited the search to journals that most frequently published behavior-analytic work. We identified journals by analyzing the articles from the first search. For each journal, we counted the total number of articles identified in the search and the number of articles that met the inclusion criteria and calculated the percentage of articles included. Journals with inclusion percentages above 30% were selected, which included *TAVB* (100%), *JABA* (85.4%), *BI* (84.9%), *JOBE* (61.5%), *ETC* (54.6%), *JPBI* (54.6%), and *BAP* (37.2%). After reviewing this list, we also elected to include *BARP*. No articles from *BARP* were screened in the first search, so we did not have an inclusion percentage; however, *BARP* is a behavior-analytic journal focusing on the intersection of research and practice in behavior analysis and thus is likely to publish studies involving caregivers. In April 2023, we conducted the fourth and final search to include articles from 2022 using the same procedures described in the third search, resulting in 383 unique articles for screening. Articles published in 2022 or earlier as advanced online prints but not yet assigned to an issue at the time of the search were included in the review.

Screening and eligibility

The full screening and eligibility process for the fourth search is shown in the PRISMA diagram (Figure 1). Screening and eligibility were conducted using an online review management system (www.covidence.org). The study authors, which included five doctoral-level and one masters-level Board Certified Behavior Analyst, screened the title and abstracts of all articles identified in the initial search after duplicates were removed ($n = 383$) to determine articles that could be excluded without full-text review. Each article was initially screened by two people. Sixty-nine articles were excluded during the screening as either nonexperimental (e.g., review or discussion paper) or not in the journal list (all articles in this category were chapters from a book with “Behavioral Interventions” in the title). Agreement between screeners was 87.3%, and disagreements were resolved by a third independent screener (also a study author). Two independent screeners then both read the remaining 317 articles and applied the inclusion and exclusion criteria. Agreement about whether to include or exclude was 89.5%, and disagreements were resolved by a third independent coder. An additional 89 articles were excluded (see Figure 1 for exclusion reasons), leaving 228 articles in the review.

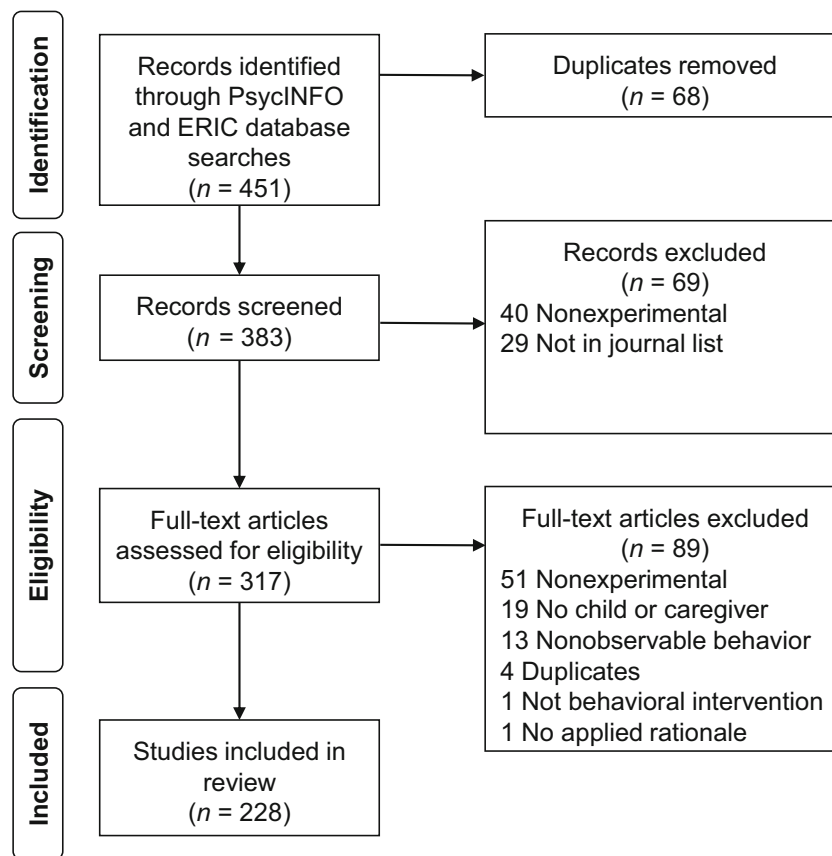


FIGURE 1 PRISMA diagram. PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Article coding

Coding procedure and agreement

Article codes (described in detail below) were entered into Google Sheets. Each of the 228 articles were coded by two independent coders. All coders were this study's authors. Interobserver agreement was calculated on a point-by-point basis. If both coders agreed on the code, a value of 1 was ascribed. If they disagreed, a value of 0 was ascribed. These values were averaged across articles for each coding category and subcategories. Across all codes and articles, mean interobserver agreement between coders was 98.1% (range across codes: 95.2%–100%; range across articles: 80.0%–100.0%). All but six articles had mean agreement at or above 90%; the other six had agreement between 80% and 89%. The nature of disagreements varied across studies. All disagreements were resolved by group consensus of the study authors, resulting in 100% agreement for all articles.

Topic and treatment

We characterized the topic area of each article based on the abstract, keywords, and primary target response under study. The topic categories were created based on

the authors' prior experience with assigning categories for behavior-analysis studies in another unpublished review; in that study, we initially had a free-text field for the topic and categorized studies retrospectively based on themes. Topic areas included behavior reduction, skill acquisition, feeding disorder, health behavior, infant behavior, and other. Table 1 includes definitions of topics. Topic areas were mutually exclusive, and topics were assigned based on the primary purpose and dependent variable. For example, a study on functional communication training to treat aggression may report both aggression (behavior reduction) and communication (skill acquisition). However, if the primary purpose was to reduce aggression, it would be assigned as behavior reduction. We further coded, regardless of the target behavior, whether a treatment was developed and evaluated for the child. Thus, interventions for caregiver, teacher, or therapist behavior only (e.g., a study focused on caregiver implementation of discrete trial teaching that did not report on a specific child target behavior for intervention) were not coded as treatment.

Child characteristics

We also grouped the studies based on the modal age of the child (infant: 0–12 months; young child: 1–5 years;

TABLE 1 Operational definitions of topics and settings

Category	Definition
Topic	Topics were mutually exclusive
Behavior reduction	Child challenging behavior assessment or treatment to reduce or eliminate unwanted behavior (e.g., self-injury, aggression, stereotypy); exclude behavior excesses related to feeding disorder
Skill acquisition	Teaching skills to children (e.g., verbal behavior, academic skills, social skills); exclude skills related to feeding disorder
Feeding disorder	Target behavior for assessment or treatment is related to a feeding disorder
Health behavior	Target behavior is relevant to child's overall health and does not fall into other categories (e.g., medication adherence, physical activity)
Infant behavior	Focus on children under 12 months and on targets specific to infants (e.g., reflexes, crying)
Other	All other topics
Setting	More than one setting per study could be coded
Home	Child and/or caregivers primary permanent place of residence; included group homes
Outpatient	Clinic for medical or psychological assessment and/or treatment; included university-based research clinics if providing clinical services to treat a medical or psychological problem
School	Daycare, preschool, primary, and secondary schools
Laboratory	Space, typically in a university, designated for research studies; did not include lab space that served as an outpatient clinic to treat a medical, psychological, or educational problem
Telehealth	Researchers/clinician connected to child and/or caregiver via videoconferencing; location of the child/caregiver also recorded
Community	Any location outside of the home, school, laboratory, or clinic
Inpatient	Temporary residence in a hospital for medical or psychological assessment and/or treatment
Other	Any location other than specified above

elementary child: 6–12 years; adolescent: 13–17 years; adult⁴: 18–21 years) and whether most children in the study were diagnosed with an IDD. The child did not

⁴Only applicable for individuals with IDD.

need to be the participant whose behavior was targeted in the study. For instance, in a study about a caregiver's implementation of functional communication training with their 8-year-old child, the child's age would be coded as "elementary child" even though the caregiver was the primary participant. If there was no modal age group, we used the mean (e.g., if participants in each study were 5 and 8 years of age, we coded the child age as "elementary child").

Caregiver characteristics

We coded characteristics of caregivers that were included in each study. For the following characteristics, we recorded the number of caregivers that were reported in each category: gender (male, female, nonbinary, transgender), race or ethnicity (Asian or Pacific Island, Black or African American, Native American or Alaska Native, White, Multiracial, Hispanic or Latinx, Other), highest education (high school, trade school, some college, bachelor's degree, master's degree, doctorate), and employment status (full-time, part-time, unemployed). For gender, we coded any caregiver labeled mother or mom as female and father or dad as male. In cases in which the authors reported that both parents were involved but did not use gendered nouns or pronouns, we did not assume the caregivers' gender and coded gender as "not reported." For race or ethnicity, categories except "Hispanic/Latinx" were mutually exclusive. We did not make assumptions about race or ethnicity based on information provided about the child participant. For example, we did not assume a caregiver was Black because their child was reported as Black. We also did not assume race or ethnicity based on the language the caregiver spoke or the country in which they lived. For highest education, we did not make assumptions about education based on the caregiver's occupation (e.g., if parent was reported as a carpenter, we did not assume they completed trade school). For employment status, we coded employment as "full-time" if the author reported them as currently employed and did not specify that they were part-time employees. For those reported to be stay-at-home caregivers, we coded them as "unemployed" unless it was explicitly stated that they held a job. We also recorded the mean and range of the caregivers' ages. In addition, we listed all languages mentioned as the languages fluently spoken by the caregiver or the family. If it was documented in the study, we reported socioeconomic status. We expected variability in how socioeconomic status was reported (e.g., labels like "low" or "middle" vs. family income values like \$50,000–\$74,999). Rather than creating categories of socioeconomic status, we simply listed all socioeconomic status information provided in each article. We did assume that the child's reported socioeconomic status was the same as the caregivers unless it was explicitly stated that the child did not live in the same household.

TABLE 2 Operational definitions of caregiver involvement categories

Category	Definition	Example
Passive caregiver involvement		
Implications for caregivers	Authors discussed the implications/impact for how a caregiver interacts with his or her child, his or her parenting style, or the caregiver's relation to a therapist/behavior analyst based on the results of the study.	Recommendation for caregiver to participate in functional analysis sessions.
Input	Caregivers were described by authors as assisting in identifying stimuli (e.g., reinforcers), situations (e.g., setting for the study, tasks assigned), and/or procedures (e.g., picking the intervention). Did not include caregivers referring their child for treatment or volunteering for a study.	Caregiver concern about extinction procedures resulted in adding noncontingent reinforcement to the treatment.
Social validity	Caregivers completed a formal social validity or treatment acceptability measure.	Caregiver completed the Treatment Acceptability Rating Form at the conclusion of the study.
Active caregiver involvement		
Implementation	Caregivers served as therapist or change-agent in some (e.g., follow-up) or all sessions.	Intervention for activity engagement was generalized to the caregiver for the final phase of the study.
Training	Caregivers were formally trained to implement procedures in the study. Training procedures had to be described by the authors.	Researcher used behavioral skills training to teach caregiver to implement discrete trial training.
Caregiver behavior	Data were collected by trained observers on caregiver behavior. Could have included procedural fidelity data on caregiver implementation or other caregiver behavior.	Researcher recorded the number of positive comments made by caregiver when interacting with their child.
Caregiver-collected data	Caregivers collected data on any behavior or event for study; excluded indirect assessments.	Caregiver recorded the number of bites consumed during each meal.

Setting

Table 1 includes operational definitions for the settings. We categorized the setting(s) in which the study was conducted as community, home, inpatient, laboratory, outpatient, school, telehealth, and other. Because many studies occurred in multiple settings, more than one setting per study was coded.

Caregiver involvement

Caregivers' involvement within the study was examined across two broad categories: passive and active involvement. *Passive involvement* was defined as procedures that did not include active participation in the study procedures but may have involved caregivers giving feedback or input during the study. We coded the following passive involvement types: implications for caregivers, input, and social validity. *Active involvement* was defined as procedures that required caregivers to directly implement at least a portion of procedures. Active involvement types included: implementation, training, caregiver behavior, and caregiver-collected data. Table 2 includes operational definitions for each type of passive and active involvement.

For each involvement type, we coded whether the article contained that type of caregiver involvement. In addition, we subcoded information for input, social validity, implementation, caregiver behavior, and caregiver-collected data. Subcodes for caregiver *input* included feedback on stimuli (e.g., potential reinforcers), task or response (e.g., the highest priority target for treatment), setting (i.e., where the study took place), assessment (e.g., conditions to include in a functional analysis), or treatment (e.g., specific treatment components). For *social validity*, we coded what tool was used (either the name of the tool or experimenter designed). For *implementation*, we coded whether caregivers interacted with their child only during a descriptive assessment (i.e., were not instructed on how to respond or interact with the child), only implemented procedures at the end (e.g., generalization), intermittently implemented procedures for some sessions throughout the study or implemented all procedures. For *caregiver behavior*, we coded what caregiver behavior was recorded. This was a free text box but could include behaviors such as procedural fidelity, treatment choice, or correct responses. If procedural fidelity was coded, we noted whether it was the primary target response in the study (i.e., a study focused on caregiver procedural fidelity). Subcategories for *caregiver-collected data* were the type of data recorded by caregivers and included frequency (e.g., count or rate), interval

(e.g., partial-interval recording), duration (e.g., how long an event lasted), event recording (i.e., whether a specific event or response occurred), rating (e.g., Likert-type scale), or permanent product (e.g., evidence that a behavior occurred without direct observation). We further noted the target behavior and whether interobserver agreement was assessed between the caregiver and another observer, and, if so, the mean interobserver agreement was coded.

Data analysis

We analyzed data using summary and descriptive statistics. We counted the number of studies meeting each coding criteria (e.g., number of studies in which the age of participant was grouped as “young”) and converted them to a percentage of studies. We also separated the percentage with each caregiver involvement category by year of publication.

RESULTS

The detailed codes assigned to each article are available in Supporting Information B and C. The full reference list for included articles is in Supporting Information D. Below, we summarize the main results for each coding category.

Journal and topic area

Approximately one third of articles were published in *JABA* (34.6%; $n = 79$). The next most frequent publication outlets included *BI* (20.6%; $n = 47$) and *BAP* (16.7%; $n = 38$). The remaining journals accounted for fewer than 20 articles each.

Skill acquisition was the most common topic area with 38.6% of studies ($n = 88$) addressing academic, daily living, communication, or social skills. Behavior reduction (i.e., assessing or treating behavioral excesses such as aggression, elopement, or disruptive behavior) was the next most common topic area (37.7%; $n = 86$). Fewer studies focused on feeding disorders, health behavior, infant behavior, and other topics. Across all topic areas, 89.0% of studies ($n = 203$) described and reported data on intervention(s) to improve child behavior (e.g., reduce challenging behavior, increase social skills). The remaining 11.0% of studies either only included an assessment of child behavior or did not report on child behavior at all (i.e., it focused on caregiver behavior).

Participants and setting

The target child population age was most commonly between 1 and 13 years old (young and elementary

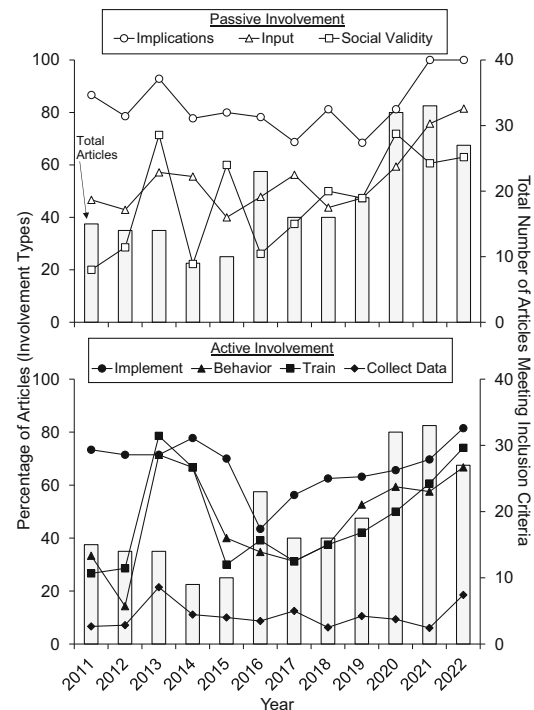


FIGURE 2 Percentage of articles by year that included caregiver involvement. The total number of articles (bars) is duplicated in top and bottom panels.

children; 87.3%; $n = 199$). Most studies targeted children with IDD (78.5%; $n = 179$). Most caregiver characteristics were infrequently reported. Full details on caregiver demographics by study are provided in Supporting Information B, and a table summarizing caregiver demographics across studies is provided in Supporting Information E. Gender was the only consistently reported caregiver demographic (65.4%; $n = 149$). Researchers reported other caregiver demographics including caregiver age in 51 studies (22.4%), highest level of education in 40 studies (17.5%), employment status in 33 studies (14.5%), race or ethnicity in 31 studies (13.6%), language in 27 studies (11.8%), and socioeconomic status in 12 studies (5.3%).

The most common settings were home (50.9%; $n = 116$) and outpatient clinics (35.1%; $n = 80$). A smaller proportion took place in school (15.6%; $n = 35$), laboratory (13.6%; $n = 31$), or telehealth settings (16.2%; $n = 37$). Very few ($n < 15$ each) took place in community, inpatient, or other settings. Additionally, most took place in one setting category (62.3%; $n = 142$).

Caregiver involvement

Overall, nearly all studies included at least one form of active or passive caregiver involvement ($n = 219$, 96.1%), and most studies included both passive and active involvement components ($n = 154$, 67.5%). Sixty-one studies (26.8%) included only passive involvement, and

four studies (1.8%) had only active components. Nine studies (3.9%) failed to mention any caregiver involvement. The amount and type of caregiver involvement varied widely across studies. Across the seven caregiver involvement types, studies included a mean of 3.7 different categories ($SD = 1.9$; range: 0–7). Figure 2 shows the percentage of articles with each involvement type by year and the total number of articles included in the review each year. The top panel (passive categories) and bottom panel (active categories) of Figure 2 show an increase in the number of articles that met the inclusion criteria for the review over the past three years. Moreover, caregiver involvement also increased across many of the types. In the sections that follow, we summarize the number of articles with each caregiver involvement category, from most to least included in our sample.

Passive involvement

The top panel of Figure 2 shows the percentage of articles that met criteria for each passive involvement category: implications for caregivers, input, and social validity. Most studies included at least one passive involvement category (94.3%; $n = 215$). If excluding implications for caregivers, which does not actually include the caregiver in the study, 75.9% ($n = 173$) included caregivers passively.

Implications for caregivers

Implications for caregivers were mentioned in most studies (84.6%; $n = 193$). Figure 2 (white circles) shows a steady percentage of articles including implications over time, with an increase in 2021 and 2022. Implications often focused on ways in which the study results inform or apply to treatment generalization with caregivers or limitations of the study as they relate to caregivers (e.g., Riviere et al., 2011). Implications for caregivers was the only caregiver involvement category in a small proportion of studies ($n = 11$; e.g., Roane & DeRosa, 2014).

Caregiver input

Caregiver input was described explicitly in 57.9% of studies ($n = 132$). Caregiver input was on an increasing trend from 2018 to 2022 as seen in Figure 2 (white triangles). Caregiver input was most used to identify potential reinforcers or other stimuli ($n = 77$; e.g., Clark et al., 2015) or to select target responses ($n = 63$; e.g., Coon et al., 2022). Caregiver input was also used to individualize assessment procedures or conditions ($n = 30$; e.g., Call et al., 2016; Slaton et al., 2017). Caregivers had explicit input on selecting treatment procedures or components in 38 studies (e.g., Gabor et al., 2016; Mendres-Smith et al., 2020). Less commonly, caregiver input was used to identify the setting where the study should take place ($n = 15$; e.g., Donaldson et al., 2013).

Social validity

Caregivers completed a formal social-validity assessment, typically in the form of a Likert-type questionnaire, regarding the study procedures in 50.0% of studies ($n = 114$). There was a slight, but variable, increasing trend over the years in reporting social validity (Figure 2, white squares). Most studies ($n = 79$) used a social-validity questionnaire designed by the experimenter. The remainder used empirically validated questionnaires such as the Treatment Acceptability Rating Form–Revised (TARF-R; Reimers et al., 1992) or Treatment Evaluation Inventory–Short Form (TEI-SF; Kelley et al., 1989); however, some studies that used these questionnaires noted making modifications (e.g., Cho & Sonoyama, 2020, modified the TARF-R). Caregivers implemented procedures with their child in 82 of the 114 studies with social validity.

Active involvement

Figure 2 (bottom panel) shows the percentage of articles that met criteria for each active involvement category: implementation, caregiver behavior, training, and caregiver-collected data. Approximately two thirds of studies included caregivers in at least one active way (69.3%; $n = 158$).

Caregiver implementation

Caregivers implemented procedures in approximately 66.7% of studies ($n = 152$). Caregiver implementation has been fairly stable (except for a decrease in 2016) over the past 10 years with an increase in 2022 (see Figure 2, black circles). Most often, caregivers were responsible for implementing all procedures ($n = 86$; e.g., Tsami & Lerman, 2020). In other studies, caregivers intermittently implemented procedures throughout the study ($n = 47$; e.g., Najdowski et al., 2012) or only implemented generalization or follow-up procedures at the end of the study ($n = 18$; e.g., Wiskow et al., 2017). Only one study included caregiver implementation during only a descriptive assessment (Blackman et al., 2020).

Caregiver training

Although caregivers implemented procedures in two thirds of studies, explicit training procedures for caregivers were only described in 49.1% of all studies ($n = 112$). Framed differently, out of the 152 studies with caregiver implementers, 71.1% ($n = 108$; four reported training caregivers, but they did not implement procedures) of the articles described caregiver training. It is unclear what training, if any, caregivers received in the remaining 44 studies in which they implemented some or all the procedures. Formal training procedures commonly involved behavioral skills training, which includes instruction, modeling, role play, and in vivo feedback (e.g., Dogan et al., 2017). In most studies with explicit

caregiver training, procedural-fidelity data were collected on caregiver responding ($n = 94$). Caregiver training increased in 2013 and 2014, followed by a sharp decrease in 2015 and then a steady increasing trend through 2022 (Figure 2, black squares).

Caregiver behavior

Data on caregiver behavior included any caregiver behavior that was directly observed and recorded by trained observers, excluding responses to indirect assessments, interviews, or surveys. Data on caregiver behavior were collected in 48.7% of studies ($n = 111$) with a variable but increasing trend over the years, similar to caregiver training (Figure 2, black triangles). Most studies that reported on caregiver behavior reported on the caregiver's procedural fidelity ($n = 104$). For 93 of those 104 articles, procedural fidelity was the only caregiver behavior recorded. Fidelity was the primary dependent variable in 33 studies. That is, the focus of the study was on teaching the caregiver to implement procedures, and the study was designed to demonstrate experimental control over caregiver fidelity (e.g., Campos et al., 2020).

Eighteen studies reported on caregiver behavior other than fidelity, including treatment selection (e.g., Gabor et al., 2016), responses to hypothetical discounting scenarios (Chadwell et al., 2019), interactions with their child (that were not part of fidelity monitoring, e.g., Blackman et al., 2020), affect (e.g., Jull & Mirenda, 2011), and correct responses on knowledge assessments or other tasks (e.g., Sivaraman & Fahmie, 2020).

Caregiver-collected data

Caregivers were involved with collecting data on their child's behavior in 10.1% of studies ($n = 24$). Over the past decade, caregiver-collected data has been consistently low (Figure 2, black diamonds). In 11 of the 24 studies, caregivers reported event data (e.g., Tanner & Andreone, 2015). Caregivers collected data on the frequency of some behavior(s) in six studies, such as the number of bites consumed (Bloomfield et al., 2021) and instruction following (Cavell et al., 2018). Caregivers collected duration data in two studies (e.g., duration of naps, McLay et al., 2019) and interval data in two studies (e.g., partial-interval recording of sibling playtime, Walton & Ingersoll, 2012). Caregivers reported permanent product data in one study (steps from a pedometer; Ek et al., 2016) and a rating in one study (level of appropriate behavior and challenging behavior; Coon et al., 2022).

DISCUSSION

The purpose of this scoping review was to summarize the amount and nature of caregiver involvement in peer-reviewed behavior-analytic articles focused on children. Our findings demonstrate that researchers recognize the importance of caregiver involvement; approximately 96%

of all the studies reviewed incorporated caregivers in some capacity. Passive involvement was common, with nearly all articles including at least one passive involvement category. In contrast, active involvement was reported in about two thirds of studies. The reviewed studies highlight various ways in which caregivers may be involved, from providing input on study materials, procedures, or settings, serving as change agents, or collecting data.

Despite a relatively high frequency of researchers reporting caregiver involvement, the description of caregiver involvement within the existing literature is limited and, at times, not technologically precise (e.g., not specifying how frequently caregivers collected data). Technologically precise descriptions of caregiver involvement can provide details necessary for replication. Furthermore, it allows for secondary research (e.g., systematic reviews, meta-analyses) directed at identifying when and how to include caregivers. As with any other aspect of treatment, clinicians should follow evidence-based practices when including caregivers. However, the extent to which or the conditions under which research translates to practice is not always clear. For example, high caseloads may limit the frequency of behavioral services provided directly by a clinician; instead, the parent may be expected to implement most of the treatment in the absence of direct clinical oversight. Toward that end, more research is needed to understand the best ways to include caregivers in practice and bridge the gap between caregiver involvement in research and the realities of practice.

Table 3 provides recommendations for caregiver involvement. These are discussed in the sections that follow.

Caregiver input and social validity

Over a half of studies described how caregivers had input on the study components such as identifying reinforcers or target responses. Future research should determine the optimal amount, type, or method of soliciting input from caregivers, which would better inform how and when to collect caregiver input in practice. For example, caregivers may provide input on stimuli, responses, and assessment conditions prior to conducting a functional analysis, which can help the researchers design relevant test and control conditions for the assessment (e.g., Gerow et al., 2021). Although collecting input from caregivers allows researchers to program for ecological validity (Fahmie et al., 2023), researchers do not always describe how input influences the assessment. For instance, studies did not usually report exact interview questions, follow-up questions, or detailed caregiver responses, so it is unclear exactly how input influences the actual assessment. Moreover, collecting caregiver input can be costly in terms of time and resources (e.g., time to administer and score indirect measures; cost of measures). Future research should include a cost-benefit analysis to evaluate relative gains for potentially costly or time-consuming measures.

TABLE 3 Recommendations for researchers for caregiver involvement

Category	Research recommendation
Caregiver input and social validity	Assess optimal ways to solicit caregiver input
	Determine types of caregiver input that are most helpful
	Evaluate whether and how caregiver involvement influences social validity
	Evaluate the interrelation of caregiver involvement, social validity, treatment adherence, and clinical outcomes
	Assess correspondence between caregiver and clinician goals
	Evaluate barriers to use psychometrically validated social validity measures
	Review the frequency with which culturally responsive care is reported in research
	Develop valid and reliable social validity measures
Caregiver implementation, training, and behavior	Provide more comprehensive descriptions of training received by caregivers prior to or during study
	Evaluate level of caregiver implementation necessary to produce improved outcomes
	Assess whether having caregivers conduct functional analysis improves maintenance or generality
	Assess barriers to including caregivers as change agents
	Evaluate conditions that influence collecting treatment-fidelity data
Caregiver-collected data	Explore relation of fidelity, adherence, and treatment outcomes
	Examine the influence of competing responsibilities of data collection reliability and validity
	Develop and evaluate data-collection tools for caregivers to collect data on macro-level changes in behavior
	Evaluate reliability and validity of caregiver data in more natural contexts, for longer observation intervals, and with a variety of target behaviors
	Evaluate the effect of caregiver data collection on procedural fidelity
Other considerations	Develop tools that automate home data collection
	Evaluate variables that contribute to caregiver adherence and attendance
	Identify other predictive behavioral markers that may be indicative of caregiver involvements
	Compare caregiver involvement based on child on child or caregiver characteristics such as diagnosis, age, race, or income
	Partner with clinicians to identify research questions that address challenges in clinical practice

Social-validity assessments are another way caregivers can provide input. Social validity is generally defined in terms of socially significant goals, procedures, and outcomes (Wolf, 1978). In the current review, nearly half of studies included some type of social-validity assessment with caregivers, which is higher than has been previously reported (10%–17%; Carr et al., 1999; Ferguson et al., 2019). Taken together, these data suggest that social-validity measures are more likely to be included when caregivers are involved. There are several lines of research that could be initiated to better understand how involvement is related to social validity. First, researchers could evaluate whether caregiver involvement influences each domain of social validity (i.e., goals, procedures, and outcomes). That is, does involving parents as change agents, data collector, and so on differentially influence their evaluation of the study goals, procedures, and outcomes? Second, researchers could examine the interrelation of caregiver involvement, social validity, treatment adherence, and clinical outcomes. For instance, there may be a mediated pathway from caregiver involvement to clinical outcomes (e.g., more involvement → increased acceptability → more likely to follow treatment recommendations → improved outcomes). Further research

could investigate what type and amount of caregiver involvement influences these pathways. Third, researchers could investigate the alignment of caregiver goals and expectations for treatment with that of the researcher or clinician. Social-validity assessments are currently the primary evidence available that the behavior analyst and caregiver goals may align. However, social-validity assessments typically ask about the treatment overall, rather than specific goals or treatment components. Moreover, these questions may omit other goals that are important to the family (e.g., reduction of medications, making friends). Last, social-validity assessments tend to be conducted at the conclusion of the study rather than at the beginning of treatment when goals are created. To that end, researchers could compare caregiver goals prior to beginning clinical services to the actual treatment goals created by the clinician in their treatment plan.⁵

Another notable finding related to social validity is that most social-validity instruments were experimenter-developed or a modified version of instruments, which

⁵It is important to note that the child's goals and objectives for treatment are essential. We focus here on caregivers as they are the focus of this review, but the same statement could and should apply to the recipient of treatment.

raises several concerns. First, the psychometric properties (i.e., reliability and validity) of these surveys are unknown (Anderson et al., 2022). Second, a lack of standardization limits the ability to draw inferences from social-validity data. If every study uses a slightly different social-validity survey, it may be impossible to synthesize and compare social-validity findings across studies. As a field, it is important to determine why validated measures are not being used (i.e., what need are they not fulfilling) and to produce psychometrically sound social-validity measures that address that reason. Such efforts are underway; for example, Anderson et al. (2022) analyzed open-ended interviews with caregivers to develop social-validity questionnaires and reported interrater reliability and content validity metrics. Researchers should continue to focus on developing valid and reliable social validity measures.

Active involvement

Caregiver implementation, training, and behavior

An important goal in the provision of behavioral services is to work toward discontinuation of services, which is accomplished, in part, by building the skills and capacities of the caregivers. The results of our review indicated that two thirds of studies involved caregivers as change agents and the majority had caregivers implement all sessions. These results are promising because caregivers would likely be expected to implement treatment procedures outside of treatment sessions or upon completion of the study. The procedures used to train caregivers were reported less frequently (49% of all studies). It is unclear what, if any, training caregivers received in the 44 studies in which they implemented procedures but training was not described. It is possible caregivers were trained but it was not reported (e.g., due to purported relevance or page limits).

Although caregivers implemented all procedures in most studies, it is unclear under which conditions using caregivers as the only change agent is warranted and what is sufficient to achieve good outcomes. It may not always be appropriate for caregivers to implement all procedures. Caregivers and their children have a long history of delivering reinforcement to one another, which may be difficult or dangerous to manage or extinguish. Starting treatment with a novel therapist may be more efficient and potentially safer. However, treatment may not transfer without explicit training (Stokes & Baer, 1977). In fact, it may be beneficial to involve the caregiver as a change agent early on to identify potential variables that differentially affect the parent-child interaction and allow for better programming for generalization. For example, Kurtz et al. (2013) found that a function was more likely to be identified in a caregiver-conducted than a staff-conducted functional analysis.

Including caregivers in the assessment period may also provide an opportunity to give caregivers hands-on, concrete experience with functional control of their child's behavior. However, it is unclear whether incorporating caregivers in assessment sessions is appropriate in all settings or situations, and further research should continue to parse out when using caregivers as change agents is indicated and whether serving as a change agent influences other outcomes such as treatment acceptability or adherence.

Furthermore, when caregivers are responsible for implementing treatment procedures, it is important to carefully monitor fidelity. Of the 152 articles in which caregivers served as change agents, 104 of them (68.4%) reported caregiver procedural fidelity. Several reviews of procedural fidelity reported in behavior-analytic journals suggest that the rate of reporting is infrequent (McIntyre et al., 2007; Peterson et al., 1982), but is improving to around 50% (Falakfarsa et al., 2022; Han et al., 2023). Therefore, it appears that researchers may be more likely to report fidelity for caregivers than for other change agents. Fidelity may be reported more frequently for caregivers because researchers may have concerns about the caregivers' ability to implement the procedures or because they anticipate others may question the fidelity of caregivers' implementation. In total, these data suggest researchers are tracking and reporting caregiver procedural fidelity, particularly when caregivers implement all sessions. Some potential areas for further research are to compare fidelity reporting and scores across different change agents and to further investigate the conditions under which researchers collect and report fidelity data. For instance, research could explore whether certain target behaviors, settings, or participant characteristics influence their decision to collect fidelity data. Reporting fidelity may be more likely in less controlled situations (e.g., home setting vs. clinical setting), which may suggest it is collected when there is less certainty that the procedures will be implemented as planned.

Another avenue of research is to explore the relation of procedural fidelity, treatment adherence, and treatment outcomes. Treatment adherence is an extension of procedural fidelity in which behaviors consistent with treatment recommendations are practiced in the absence of oversight (Moore & Amado, 2021). It seems plausible that higher fidelity (i.e., observed correct implementation of recommendations) would predict higher adherence. However, there are other variables that may influence adherence such as weak establishing operations (e.g., the change in their child's behavior is not immediately large enough to be reinforcing), inadequate stimulus generalization (e.g., the caregiver was not trained to conduct the treatment in a sufficient number of settings), and competing contingencies (e.g., other responsibilities that require the caregivers' attention; Allen & Warzak, 2000). The relation between fidelity, adherence, and outcomes could be evaluated in a longitudinal study that tracks these variables over time.

Caregiver-collected data

The articles we reviewed very rarely made use of caregiver-collected data. Similarly, a recent survey of Board Certified Behavior Analysts and their data-collection practices reported that only 2% used caregivers as the primary data collector (Morris et al., 2022). However, in practice, caregivers may be able to provide quantifiable evidence as to the efficacy of a behavior change program outside of the clinical setting during the entire course of treatment, allowing for better collection of maintenance and generalization data. Historically, behavior analysts have questioned the accuracy of caregiver report (Baer et al., 1968, 1987). Based on the small number of studies that included caregivers as data collectors and reported interobserver agreement, caregivers' interobserver agreement was generally acceptable (>80%), suggesting that caregivers could be trustworthy data reporters, at least by the standards with which we evaluate trained observers. Notably, these studies often involved brief periods of data collection in a highly controlled setting. Researchers should evaluate the influence of competing responsibilities (e.g., work, caring for other children) on data collection on frequency and accuracy. In particular, researchers should develop and test data-reporting methods that are feasible for caregivers (i.e., simple and low effort), cost effective, and sensitive to behavior change. It may not be necessary for caregivers to collect precise frequency counts of behavior. Rather, caregiver data could be used to detect macro changes—that is, treatment effects or outcomes—rather than micro session-by-session discriminations (see Morris et al., 2022, for a discussion of micro- and macro-data collection). There is some evidence that caregiver report of macro-level changes may correspond to trained observer data, but results are preliminary and need to be extended for longer observation durations, in more natural contexts, and with other types of target behaviors (Becraft et al., 2023).

Furthermore, researchers should evaluate changes in procedural fidelity and the accuracy of data collection. Data collection might improve fidelity because it may facilitate caregiver's discrimination of their child's behavior, which is presumably related to the behavioral contingencies that the caregiver is expected to implement. Conversely, accuracy may be lower if the caregiver is focused on accurately implementing the treatment. Another alternative is to develop and test methods to automatically collect data at home such as wearables and computerized responses. Such technology may not be capable of recording all behavior that may be of interest to behavior analysts, but there are several ways in which automated data collection is already in use and could be tested further (Bak et al., 2021). For example, if the goal of an intervention is to increase vigorous physical activity, the child could wear a watch with an activity tracker and heart rate monitor (e.g., Van Camp & Berth, 2018). Researchers might investigate whether displays or other permanent products of data-collection efforts (e.g., graphs

of progress) affect caregivers' procedural fidelity and the frequency and accuracy of nonautomated data collection.

Other considerations for caregiver involvement

Our review also highlighted several other considerations for caregiver involvement in behavior-analytic research. Although we generally believe caregiver involvement is good, there may be times when caregiver involvement is not indicated or is counterproductive. We do not suggest that researchers should always attempt to include caregivers in all ways. For example, when testing a new intervention, it may be important for highly skilled therapists to serve as change agents to ensure procedural fidelity. Similarly, when the target behavior is complex or difficult to observe, it may not be helpful for caregivers to serve as data collectors. Some types of caregiver involvement may also be less relevant in contexts such as school or work. Also, caregiver involvement may not be warranted for certain populations (e.g., adults).

Studying caregiver involvement can have practical challenges and barriers for researchers. Caregivers may have limited time to participate in research sessions, or they may not be motivated to participate. Researchers can provide incentives for their participation, but this can be costly, and funds to support it may not be available. Broadly, researchers can address these barriers by using a functional approach. That is, researchers should identify the behavioral target and its controlling variables. For example, the Performance Diagnostic Checklist–Parent (Hodges et al., 2020) may be helpful. Such an approach may highlight potential interventions to improve involvement. For instance, to reduce time and access burdens, caregivers could complete surveys or interviews online or in a telehealth meeting.

Another important consideration is the relation between caregiver involvement and treatment outcomes. The “dose” of involvement could predict outcomes and may help behavior analysts better design interventions to serve their clients. Reporting outcomes was beyond the scope of the current review, but future research could test whether involvement is related to outcomes. Borrowing from precision medicine, Hagopian et al. (2018) provide a potential model for identifying predictive behavior markers of treatment success. Using a similar approach, the nature and amount of caregiver involvement could be conceptualized and tested as a predictor of treatment success. For example, as noted previously, caregiver implementation during the functional analysis appears to be associated with a greater likelihood of identifying the function of challenging behavior. This presumably leads to a greater likelihood of selecting an effective treatment, which could result in a greater or quicker reduction in challenging behavior.

Moreover, caregiver demographics were reported infrequently, which is similar to the lack of reporting on participant demographics in behavior-analytic research

(Jones et al., 2020). When demographics were reported, our review suggested that the caregivers most likely to be involved were White females with at least some college education. Several cultural and demographic variables may influence the amount and nature of caregiver involvement. For instance, if a caregiver's native language is different than that of the researchers, then any involvement would likely involve obtaining an interpreter and translating materials. If such resources are not available, then caregiver involvement may be limited. Similarly, if a caregiver is a low-income single parent who must work long hours, it may not be reasonable to expect the same level of involvement as a middle-class stay-at-home parent. Because demographics are reported rarely, it is difficult to draw conclusions about caregiver involvement based on such variables. Therefore, researchers should report demographics for participants and their caregiver whenever possible, which would allow for the evaluation of ideal ways to arrange caregiver involvement across different cultural groups in systematic reviews and meta-analyses.

Limitations of review

There are a few limitations of this review. Reporting or editorial practices may limit how or if caregiver involvement was reported. The decisions to include or not include this information could be based on relevance to the study purpose. Alternatively, authors may choose not to include an element of caregiver involvement because it did not meet standards for research (e.g., interobserver agreement was not assessed for caregiver data). Therefore, it is possible that we underestimated caregiver involvement. However, we were most interested in the nature of involvement. Moreover, we believe the absolute amount of involvement is less interesting than the relative amount (i.e., the comparison of the amount of involvement across different types of involvement). Future research may investigate relations between caregiver involvement and other variables that we measured (e.g., study topic, child age, parent demographics). To that end, we have made our database available in Supporting Information B and C.

Another limitation of this review is that the search terms were designed to identify articles that involved caregivers. Therefore, the articles we included were biased to include caregivers and do not represent caregiver involvement in studies with children more generally. As a clear example, social validity was reported at least three times more frequently in our review than in previous reviews (Carr et al., 1999; Ferguson et al., 2019). However, the purpose of this scoping review was not to determine the prevalence of caregiver involvement but to understand the nature of caregiver involvement that has been reported in behavior-analytic literature and identify potential areas for further research.

In addition, there are some limitations in how we coded articles. The coding categories were somewhat

arbitrary. For example, the article topics were based on our experience with a prior unpublished review, and there may have been a better way to categorize them. Behavior-reduction and skill-acquisition studies each accounted for about one third of articles and could have been broken down further based on the function or topography of challenging behavior or the type of skill (e.g., social, academic). Additional reviews could focus on one particular topic area and further investigate the role of caregiver involvement.

Finally, we restricted our review to eight journals that publish behavior-analytic research. This decision was the result of a cost-benefit analysis when the second search returned nearly 90,000 articles. The review's purpose was to understand the nature of caregiver involvement in behavior-analytic research, but the number of articles that would have to be screened and coded was not reasonable. We used a data-based process to select seven of the eight journals; these were journals that had the highest likelihood of including relevant articles for the review in the initial search. Restricting journals in this way did exclude articles that would have been relevant. For example, our search did not include a study on the collateral effects of function-based treatment for challenging behavior on caregiver stress (Kurtz et al., 2021) that otherwise met review criteria. However, our review includes a large sample of research that we believe is sufficient to highlight trends and identify potential under-investigated areas in the research.

Conclusion

Overall, this scoping review highlights the diverse ways in which caregivers are involved in behavior-analytic research. Importantly, outcomes of the current review were intended to both describe the nature of caregiver involvement and highlight potential gaps in the literature that could stimulate future research. Caregiver involvement continues to warrant further investigation, particularly with respect to active involvement methods. Additionally, more research is needed on commonly used methods of passive involvement such as how to best use and collect caregiver input before, during, and after treatment. The relation between research and practice should be bidirectional. We encourage researchers to solicit feedback from and collaborate with clinicians to help identify research questions that would address challenges in their clinical practice. This could be a formal (e.g., surveying a broad group of clinicians) or informal (e.g., collaborating with local clinicians) process. Such research has the potential to provide us with a better understanding of caregiver involvement, bridge the gap between clinical research and clinical practice, and improve clinical outcomes.

ACKNOWLEDGMENTS

We would like to thank Michael Cataldo for feedback on a previous version of this paper.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

The method to identify, screen, and include articles and planned analyses were specified in advance in a published protocol on the Open Science Framework (<https://osf.io/g55qx>).

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses—Scoping Review Checklist is provided in Supporting Information A.

Individual codes for each article are provided in Supporting Information B (Study Characteristics Codes) and C (Caregiver Involvement Codes).

Full citations for all articles included in the review are provided in Supporting Information D.

A summary table of caregiver demographics is provided in Supporting Information E.

ETHICS APPROVAL

No human or animal subjects were used to produce this article.

ORCID

Jessica L. Becraft  <https://orcid.org/0000-0002-0737-5070>

Samantha L. Hardesty  <https://orcid.org/0000-0001-8778-5477>

Kissel J. Goldman  <https://orcid.org/0000-0003-0355-0000>

Lesley A. Shawler  <https://orcid.org/0000-0002-5081-6839>

Matthew L. Edelstein  <https://orcid.org/0000-0001-6860-3068>

REFERENCES

- Allen, K. D., & Warzak, W. J. (2000). The problem of parental nonadherence in clinical behavior analysis: Effective treatment is not enough. *Journal of Applied Behavior Analysis, 33*(3), 373–391. <https://doi.org/10.1901/jaba.2000.33-373>
- Anderson, R., Taylor, S., Taylor, T., & Virues-Ortega, J. (2022). Thematic and textual analysis methods for developing social validity questionnaires in applied behavior analysis. *Behavioral Interventions, 37*(3), 732–753. <https://doi.org/10.1002/bin.1832>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*(1), 91–97. <https://doi.org/10.1901/jaba.1968.1-91>
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 20*(4), 313–327. <https://doi.org/10.1901/jaba.1987.20-313>
- Bak, M. Y. S., Plavnick, J. B., Dueñas, A. D., Brodhead, M. T., Avendaño, S. M., Wawrzonek, A. J., Weber, E., Dodson, S. N., & Oteto, N. (2021). The use of automated data collection in applied behavior analytic research: A systematic review. *Behavior Analysis: Research and Practice, 21*(4), 376–405. <https://doi.org/10.1037/bar0000228>
- Becraft, J. L., Cataldo, M. F., Yu-Lefler, H. F., Schenk, Y. A., Edelstein, M. L., & Kurtz, P. F. (2023). Correspondence between data provided by parents and trained observers about challenging behavior. *Behavior Analysis: Research and Practice*. Advance online publication. <https://doi.org/10.1037/bar0000276>
- Behavior Analyst Certification Board (2020). *Ethics code for behavior analysts*. <https://bacb.com/wp-content/ethics-code-for-behavior-analysts/>
- Behavior Analyst Certification Board (n.d.). *BACB certificant data*. <https://www.bacb.com/BACB-certificant-data>
- Blackman, A. L., Jimenez-Gomez, C., & Shvarts, S. (2020). Comparison of the efficacy of online versus in-vivo behavior analytic training for parents of children with autism spectrum disorder. *Behavior Analysis: Research and Practice, 20*(1), 13–23. <https://doi.org/10.1037/bar0000163>
- Bloomfield, B. S., Fischer, A. J., Dove, M., Clark, R. R., & Fife, M. (2021). Parent teleconsultation to increase bites consumed: A demonstration across foods for a child with ARFID and ASD. *Behavior Analysis in Practice, 14*(4), 913–926. <https://doi.org/10.1007/s40617-021-00586-4>
- Call, N. A., Miller, S. J., Mintz, J. C., Mevers, J. L., Scheithauer, M. C., Eshelman, J. E., & Beavers, G. A. (2016). Use of a latency-based demand assessment to identify potential demands for functional analyses. *Journal of Applied Behavior Analysis, 49*(4), 900–914. <https://doi.org/10.1002/jaba.341>
- Campos, C., Bloom, S. E., Weyman, J. R., & Garcia, A. R. (2020). Parent-implemented multiple schedules. *Behavioral Interventions, 35*(4), 524–541. <https://doi.org/10.1002/bin.1743>
- Carr, J. E., Austin, J. L., Britton, L. N., Kellum, K. K., & Bailey, J. S. (1999). An assessment of social validity trends in applied behavior analysis. *Behavioral Interventions, 14*(4), 223–231. [https://doi.org/10.1002/\(SICI\)1099-078X\(199910/12\)14:4<223::AID-BIN37>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1099-078X(199910/12)14:4<223::AID-BIN37>3.0.CO;2-Y)
- Cavell, H. J., Radley, K. C., Dufrene, B. A., Tingstrom, D. H., Ness, E. A., & Murphy, A. N. (2018). The effects of errorless compliance training on children in home and school settings. *Behavioral Interventions, 33*(4), 391–402. <https://doi.org/10.1002/bin.1641>
- Chadwell, M. R., Sikorski, J. D., Roberts, H., & Allen, K. D. (2019). Process versus content in delivering ABA services: Does process matter when you have content that works? *Behavior Analysis: Research and Practice, 19*(1), 14–22. <https://doi.org/10.1037/bar0000143>
- Cho, S., & Sonoyama, S. (2020). Combining stimulus fading with simultaneous stimulus presentation in a school setting for food selectivity in a child with autism spectrum disorder. *Education and Treatment of Children, 43*(4), 335–345. <https://doi.org/10.1007/s43494-020-00011-5>
- Clark, D. R., Donaldson, J. M., & Kahng, S. (2015). Are video-based preference assessments without access to selected stimuli effective? *Journal of Applied Behavior Analysis, 48*(4), 895–900. <https://doi.org/10.1002/jaba.246>
- Coon, J. C., Rapp, J. T., Bush, H., & Ramey, E. S. (2022). Preliminary findings on the effects of medication advocacy and daily behavior ratings on prescriber behavior. *Behavioral Interventions, 37*(3), 887–897. <https://doi.org/10.1002/bin.1859>
- Dogan, R. K., King, M. L., Fischetti, A. T., Lake, C. M., Mathews, T. L., & Warzak, W. J. (2017). Parent-implemented behavioral skills training of social skills. *Journal of Applied Behavior Analysis, 50*(4), 805–818. <https://doi.org/10.1002/jaba.411>
- Donaldson, J. M., Vollmer, T. R., Yakich, T. M., & Van Camp, C. (2013). Effects of a reduced time-out interval on compliance with the time-out instruction. *Journal of Applied Behavior Analysis, 46*(2), 369–378. <https://doi.org/10.1002/jaba.40>
- Dorsey, M. F., & Harper, J. M. (2018). Toward the scientist practitioner model in applied behavior analysis: Examples of a science-based practice. *Education and Treatment of Children, 41*(3), 273–275. <https://doi.org/10.1353/etc.2018.0013>
- Ek, K. E., Miltenberger, R. G., & Valbuena, D. (2016). Promoting physical activity among school-age children using feedback, goal

- setting, and rewards. *Behavior Analysis: Research and Practice*, 16(1), 41–46. <https://doi.org/10.1037/bar0000029>
- Fahmie, T. A., Rodriguez, N. M., Luczynski, K. C., Rahaman, J. A., Charles, B. M., & Zangrillo, A. N. (2023). Toward an explicit technology of ecological validity. *Journal of Applied Behavior Analysis*, 56(2), 302–322. <https://doi.org/10.1002/jaba.972>
- Falakfarsa, G., Brand, D., Jones, L., Godinez, E. S., Richardson, D. C., Hanson, R. J., Velazquez, S. D., & Wills, C. (2022). Treatment integrity reporting in *Behavior Analysis in Practice* 2008–2019. *Behavior Analysis in Practice*, 15(2), 443–453. <https://doi.org/10.1007/s40617-021-00573-9>
- Ferguson, J. L., Cihon, J. H., Leaf, J. B., Van Meter, S. M., McEachin, J., & Leaf, R. (2019). Assessment of social validity trends in the *Journal of Applied Behavior Analysis*. *European Journal of Behavior Analysis*, 20(1), 146–157. <https://doi.org/10.1080/15021149.2018.1534771>
- Gabor, A. M., Fritz, J. N., Roath, C. T., Rothe, B. R., & Gourley, D. A. (2016). Caregiver preference for reinforcement-based interventions for problem behavior maintained by positive reinforcement. *Journal of Applied Behavior Analysis*, 49(2), 215–227. <https://doi.org/10.1002/jaba.286>
- Gerow, S., Radhakrishnan, S., Davis, T.N., Zambrano, J., Avery, S., Cosottile, D.W., & Exline, E. (2021). Parent-implemented brief functional analysis and treatment with coaching via telehealth. *Journal of Applied Behavior Analysis*, 54(1), 54–69. <https://doi.org/10.1002/jaba.801>
- Hagopian, L. P., Rooker, G. W., & Yenokyan, G. (2018). Identifying predictive behavioral markers: A demonstration using automatically reinforced self-injurious behavior. *Journal of Applied Behavior Analysis*, 51(3), 443–465. <https://doi.org/10.1002/jaba.477>
- Han, J. B., Bergmann, S., Brand, D., Wallace, M. D., St. Peter, C. C., Feng, J., & Long, B. P. (2023). Trends in reporting procedural integrity: A comparison. *Behavior Analysis in Practice*, 16(2), 388–398. <https://doi.org/10.1007/s40617-022-00741-5>
- Hodges, A. C., Villacorta, J., Wilder, D. A., Ertel, H., & Luong, N. (2020). *Performance Diagnostic Checklist–Parent*. APA PsycTESTS. <https://doi.org/10.1037/t83504-000>
- Jones, S. H., St. Peter, C. C., & Ruckle, M. M. (2020). Reporting of demographic variables in the *Journal of Applied Behavior Analysis*. *Journal of Applied Behavior Analysis*, 53(3), 1304–1315. <https://doi.org/10.1002/jaba.722>
- Jull, S., & Mirenda, P. (2011). Parents as play date facilitators for pre-schoolers with autism. *Journal of Positive Behavior Interventions*, 13(1), 17–30. <https://doi.org/10.1177/1098300709358111>
- Kelley, M. L., Heffer, R., Gresham, F. M., & Elliott, S. N. (1989). Development of a modified treatment evaluation inventory. *Journal of Psychopathology and Behavioral Assessment*, 11(3), 235–247. <https://doi.org/10.1007/BF00960495>
- Kemmerer, A. R., Vladescu, J. C., DeBar, R. M., Sidener, T. M., & Bell, M. C. (2023). A scoping review of the caregiver training literature for individuals with autism spectrum disorder. *Behavioral Interventions*, 38(3), 767–792. <https://doi.org/10.1002/bin.1939>
- Kurtz, P. F., Fodstad, J. C., Huete, J. M., & Hagopian, L. P. (2013). Caregiver- and staff-conducted functional analysis outcomes: A summary of 52 cases. *Journal of Applied Behavior Analysis*, 46(4), 738–749. <https://doi.org/10.1002/jaba.87>
- Kurtz, P. F., Strohmeier, C. W., Bercraft, J. L., & Chin, M. D. (2021). Collateral effects of behavioral treatment for problem behavior on caregiver stress. *Journal of Autism and Developmental Disorders*, 51(8), 2852–2865. <https://doi.org/10.1007/s10803-020-04694-y>
- McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the *Journal of Applied Behavior Analysis* 1991–2005. *Journal of Applied Behavior Analysis*, 40(4), 659–672. <https://doi.org/10.1901/jaba.2007.659-672>
- McLay, L. K., France, K. G., Knight, J., Blampied, N. M., & Hastie, B. (2019). The effectiveness of function-based interventions to treat sleep problems, including unwanted co-sleeping, in children with autism. *Behavioral Interventions*, 34(1), 30–51. <https://doi.org/10.1002/bin.1651>
- Mendres-Smith, A. E., Borrero, J. C., Castillo, M. I., Davis, B. J., Bercraft, J. L., & Hussey-Gardner, B. (2020). Tummy time without the tears: The impact of parent positioning and play. *Journal of Applied Behavior Analysis*, 53(4), 2090–2107. <https://doi.org/10.1002/jaba.715>
- Moore, T. R., & Amado, R. S. (2021). A conceptual model of treatment adherence in a behavior analytic framework. *Education and Treatment of Children*, 44(1), 1–17. <https://doi.org/10.1007/s43494-020-00032-0>
- Morris, C., Conway, A. A., Bercraft, J. L., & Ferrucci, B. J. (2022). Toward an understanding of data collection integrity. *Behavior Analysis in Practice*, 15(4), 1361–1372. <https://doi.org/10.1007/s40617-022-00684-x>
- Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, 18(1), Article 143. <https://doi.org/10.1186/s12874-018-0611-x>
- Najdowski, A. C., Tarbox, J., & Wilke, A. E. (2012). Utilizing antecedent manipulations and reinforcement in the treatment food selectivity by texture. *Education and Treatment of Children*, 35(1), 101–110. <https://doi.org/10.1353/etc.2012.0004>
- Northup, J., Vollmer, T. R., & Serrett, K. (1993). Publication trends in 25 years of the *Journal of Applied Behavior Analysis*. *Journal of Applied Behavior Analysis*, 26(4), 527–537. <https://doi.org/10.1901/jaba.1993.26-527>
- Oxford Centre for Evidence-Based Medicine. (2011). The Oxford 2011 levels of evidence. <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebml-levels-of-evidence>
- Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, 15(4), 477–492. <https://doi.org/10.1901/jaba.1982.15-477>
- Reimers, T. M., Wacker, D. P., Cooper, L. J., & de Raad, A. O. (1992). Acceptability of behavioral treatments for children: Analog and naturalistic evaluations by parents. *School Psychology Review*, 21(4), 628–643.
- Riviere, V., Becquet, M., Peltret, E., Facon, B., & Darcheville, J.-C. (2011). Increasing compliance with medical examination requests directed to children with autism: Effects of a high-probability request procedure. *Journal of Applied Behavior Analysis*, 44(1), 193–197. <https://doi.org/10.1901/jaba.2011.44-193>
- Roane, H. S., & DeRosa, N. M. (2014). Reduction of emergent dropping behavior during treatment of elopement. *Journal of Applied Behavior Analysis*, 47(3), 633–638. <https://doi.org/10.1002/jaba.136>
- Sivaraman, M., & Fahmie, T. A. (2020). Evaluating the efficacy and social validity of a culturally adapted training program for parents and service providers in India. *Behavior Analysis in Practice*, 13(4), 849–861. <https://doi.org/10.1007/s40617-020-00489-w>
- Slaton, J. D., Hanley, G. P., & Raftery, K. J. (2017). Interview-informed functional analyses: A comparison of synthesized and isolated components. *Journal of Applied Behavior Analysis*, 50(2), 252–277. <https://doi.org/10.1002/jaba.384>
- Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, 10(2), 349–367. <https://doi.org/10.1901/jaba.1977.10-349>
- Sun, X. (2022). Behavior skills training for family caregivers of people with intellectual or developmental disabilities: A systematic review of literature. *International Journal of Developmental Disabilities*, 68(3), 247–273. <https://doi.org/10.1080/20473869.2020.1793650>
- Tanner, A., & Andreone, B. E. (2015). Using graduated exposure and differential reinforcement to increase food repertoire in a child with autism. *Behavior Analysis in Practice*, 8(2), 233–240. <https://doi.org/10.1007/s40617-015-0077-9>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L.,

- Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., & Lewin, S. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467–473. <https://doi.org/10.7326/M18-0850>
- Tsami, L., & Lerman, D. C. (2020). Transfer of treatment effects from combined to isolated conditions during functional communication training for multiply controlled problem behavior. *Journal of Applied Behavior Analysis*, *53*(2), 649–664. <https://doi.org/10.1002/jaba.629>
- Unholz-Bowden, E., McComas, J. J., McMaster, K. L., Girtler, S. N., Kolb, R. L., & Shipchandler, A. (2020). Caregiver training via telehealth on behavioral procedures: A systematic review. *Journal of Behavioral Education*, *29*(2), 246–281. <https://doi.org/10.1007/s10864-020-09381-7>
- Van Camp, C. M., & Berth, D. (2018). Further evaluation of observational and mechanical measures of physical activity. *Behavioral Interventions*, *33*(3), 284–296. <https://doi.org/10.1002/bin.1518>
- Walton, K.M., & Ingersoll, B.R. (2012). Evaluation of a sibling-mediated imitation intervention for young children with autism. *Journal of Positive Behavior Interventions*, *14*(4), 241–253. <https://doi.org/10.1177/1098300712437044>
- Wiskow, K. M., Donaldson, J. M., & Matter, A. L. (2017). An evaluation of generalization of compliance across response types. *Behavior Analysis: Research and Practice*, *17*(4), 402–420. <https://doi.org/10.1037/bar0000087>
- Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, *11*(2), 203–214. <https://doi.org/10.1901/jaba.1978.11-203>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Becraft, J. L., Hardesty, S. L., Goldman, K. J., Shawler, L. A., Edelstein, M. L., & Orchowicz, P. (2024). Caregiver involvement in applied behavior-analytic research: A scoping review and discussion. *Journal of Applied Behavior Analysis*, *57*(1), 55–70. <https://doi.org/10.1002/jaba.1035>