J Behav Educ (2019) 28:27–53
https://doi.org/10.1007/s10864-018-9299-6

ORIGINAL PAPER

Maintenance and Generalization in Functional Behavior Assessment/Behavior Intervention Plan Literature

Brittany Pennington[1] · Jessica Simacek[1] · Jennifer McComas[1] · Kristen McMaster[1] · Marianne Elmquist[1]

Published online: 28 April 2018

Abstract School practitioners increasingly use functional behavior assessments to design behavior intervention plans when students engage in challenging behavior that is not improved by classroom management or standard interventions. Behavior intervention plans should aim to produce behavior change that maintains across time and generalizes to relevant contexts. In applied terms, maintenance occurs when a behavior continues to occur after the intervention is fully or partially withdrawn; generalization occurs when a learner engages in a trained behavior in an untrained context. This review examined the current status of maintenance and generalization in the school-based FBA/BIP literature and found that many school-based researchers continue to implement the intervention during the maintenance phase or in the generalization contexts. Implications for practice and directions for future research are discussed.

Keywords Behavior · Assessment · Intervention · Maintenance · Generalization

Introduction

Schools increasingly use functional behavior assessments (FBAs) to develop individualized behavior intervention plans (BIPs) when students engage in challenging behavior (Horner et al. 1990; Kincaid et al. 2015). FBAs include direct and indirect assessments that inform BIPs by determining the variables that maintain behavior (Cooper et al. 2007). FBA-based BIPs reduce challenging behavior and increase appropriate behavior across a variety of settings, participant ages, and disability categories (Goh

✉ Brittany Pennington
penni156@umn.edu

[1] Department of Educational Psychology, University of Minnesota, 250 Education Sciences, 56 East River Road, Minneapolis, MN 55455, USA

and Bambara 2012) and improve behavior more effectively than non-function-based interventions (Ingram et al. 2005). Because FBA-based BIPs reflect best practices, researchers should examine ways to design BIPs to increase their likelihood of long-term success.

Practitioners should aim to write BIPs that produce lasting behavior change (Baer et al. 1968). Teachers may achieve lasting behavior change in different ways. One way to produce lasting change may involve continuing to implement an effective intervention over a long period of time. For example, a teacher may improve a student's behavior by seating a student near a preferred peer (e.g., Losinski et al. 2015). The teacher may be satisfied implementing that intervention for the entire school year. Another way to produce lasting change involves designing the BIP so that it produces maintenance (i.e., the behavior change continues even after the BIP is no longer in place) and generalization (i.e., the behavior change occurs in contexts in which the BIP is not directly implemented). For example, a teacher might use prompts to increase compliance and then fade the prompts gradually over time to increase the likelihood that the results maintain after the BIP is no longer in place (e.g., Campbell and Tincani 2011). However, determining how to effectively design BIPs to produce maintenance and generalization is challenging, and few researchers have examined maintenance and generalization phases across school-based FBA/BIP studies (Goh and Bambara 2012).

Goh and Bambara (2012) analyzed a number of variables related to FBA-based interventions conducted in schools, including the extent to which interventions facilitated maintenance and generalization. They found a median of 100% non-overlapping data (PND) between baseline and maintenance and 94% PND between baseline and generalization, suggesting lasting behavior change; however, they suggested caution in interpreting their results because only 17 studies measured maintenance and six measured generalization. Additionally, Goh and Bambara (2012) did not consider how authors defined maintenance and generalization. Thus, whether maintenance phases in their included studies measured continued intervention effectiveness or behavior change after removal of the intervention remains unclear. Aside from this example, to our knowledge, no researcher has examined school-based FBA/BIP studies to determine whether maintenance and generalization occurred. Further, to our knowledge, no researcher has analyzed how school-based researchers defined, measured, or programmed for maintenance and generalization.

The purpose of the current review was to identify the state of maintenance and generalization in school-based FBA/BIP literature. This review addressed the following research questions: (1) How have researchers defined maintenance and generalization? (2) How have researchers measured maintenance and generalization? (3) What behavioral principles and strategies to produce maintenance and generalization have researchers included in their BIPs? (4) To what extent have maintenance and generalization occurred?

Method

Search Procedures

To locate recent school-based FBA/BIP studies that included maintenance and/or generalization data, we searched the titles, keywords, abstracts, and full texts of articles in the electronic databases of Education Resource Information Center (ERIC), PsycINFO, and PsycARTICLES using the phrases *functional behavior assessment*, *functional analysis*, *assessment-based intervention*, *behavior intervention*, *structural analysis*, and *function-based intervention* with the terms *school* and *classroom* for articles published in peer-reviewed journals between January 2005 and May 2016. This search identified 573 studies. Coders conducted a title review to eliminate articles that included words in the title indicating exclusion. During the title review, coders eliminated studies if the title indicated: (1) the study was an RCT, qualitative study, or review, (2) the setting was clearly non-school-based (e.g., clinic or daycare setting), (3) the participants were adults, or (4) the study was unrelated (e.g., studies promoting health behavior). The title review eliminated 270 articles.

Coders evaluated the remaining articles to determine whether they met six criteria. First, to evaluate interventions in public school settings, the study must have been conducted in a public school and existing school staff (e.g., teachers, paraprofessionals) must have participated in assessment and/or intervention. Second, participants must have been students in grades K-12. Studies in which the primary participant was the teacher (e.g., studies on coaching) were excluded. Third, the study must have included a functional behavior assessment, defined as any assessment designed to identify the function maintaining the target behavior (e.g., functional assessment interview, ABC data collection, functional analysis). Fourth, in order to evaluate the effects of FBA-based BIPs for individual students, the study must have included intervention data for individual students in a single-subject experimental design. Fifth, the study must have graphed follow-up, maintenance, or generalization data. Sixth, articles that measured only academic behavior were excluded. If some participants met the inclusion criteria and others did not, only participants who met criteria were included. Twenty-one articles met the inclusion criteria.

Next, coders conducted an ancestral search of included articles and literature reviews from the initial search. Three additional articles were identified during the ancestral search. Then, coders conducted a hand search of *Journal of Applied Behavior Analysis, Behavioral Disorders, Beyond Behavior, Education and Treatment of Children, Behavior Modification, Intervention in School and Clinic, Journal of Behavioral Education, Journal of Emotional and Behavioral Disorders, Journal of Positive Behavior Interventions,* and *Preventing School Failure* from 2005 to 2016. Two additional articles were identified during the hand search. The number of articles located meeting our inclusion criteria is consistent with previous research; of 83 studies evaluated by Goh and Bambara (2012) that included function-based BIPs in school settings, 17 measured maintenance and six measured generalization.

Search Inter-rater Agreement

In addition to the primary coder, two additional coders, both PhD students in special education with experience in behavior analysis, served as the second coder for at least 15% of articles such that two coders independently screened a total of 42% of articles for the title review, 43% of articles for the full-text review, 31% of articles for the ancestral search, and 40% of articles for the hand search. For each step in the search process, IRA was calculated as agreements over agreements plus disagreements. IRA averaged 95% for the title review (range 92–100%), 94% for the full-text review (range 87–100%), 100% for the ancestral search, and 100% for the hand search.

Coding Procedures

A coding manual was developed to answer the research questions. All questions were coded at the level of the participant. The following sections summarize the coding procedures, and the complete manual is available from the first author upon request.

Maintenance or Follow-Up Probe

Coders recorded which term the authors used. If both terms described the same phase, they coded *Both, same phase*. If both terms were used to describe different phases, they coded *Both, different phases.*

Did the Intervention Continue in the Maintenance/Follow-Up Phase?

If the full intervention was implemented in the maintenance or follow-up phase, coders recorded *Yes*, if the intervention was not implemented, they coded *No*, and if part of the intervention was implemented, they coded *partial.*

How Much Time Elapsed Between the Last Intervention Probe and First Maintenance or Follow-Up Probe?

Coders recorded the amount of time in days, weeks, or months (as reported in the study) between the last intervention probe and first maintenance/follow-up probe.

How Many Probes Did the Researchers Collect?

Coders recorded the total number of maintenance or follow-up probes.

What Behavioral Principles Did Researchers Include in BIPs?

Coders recorded the presence or absence of behavioral momentum (e.g., high-probability request sequence), differential reinforcement, extinction, feedback, MO manipulation, NCR, prompting, punishment, and teaching a skill.

Did Researchers Program for Maintenance?

To identify strategies to promote maintenance, we identified strategies listed in two seminal articles and two behavior analysis textbooks (Cooper et al. 2007; Mayer et al. 2011; Stokes and Baer 1977; Stokes and Osnes 1989). We located the articles cited as examples in the seminal articles and textbooks to determine whether they promoted maintenance over time or generalization to novel contexts. Table 1 outlines the strategies we identified.

Coders recorded the presence of any of the following strategies as described in Table 1: (1) *delay reinforcement*, (2) *schedule thinning*, (3) *demand fading*, (4) *introduce the behavior to naturally maintaining contingencies*, (5) *use intermittent reinforcement*, (6) *teach self-management*, (7) *teach behaviors to a level likely to encounter reinforcement*, (8) *teach the learner to recruit reinforcement*, and (9) *partial/sequential withdrawal*. If authors stated that they programmed for maintenance but none of the above strategies were included in the intervention, coders recorded *Yes/unclear*. If the authors did not state that they programmed for maintenance and none of the above strategies were apparent, coders recorded *No/unclear.*

Do the Authors Conclude that Maintenance Occurred?

If the authors stated or implied that maintenance occurred, coders recorded *Yes*. If the authors stated or implied that maintenance did not occur (e.g., on-task behavior decreased during the maintenance phase), coders recorded *No*. If the authors stated their results without interpretation, coders recorded *Unclear*.

Generalization

Was the Intervention Implemented in the Generalization Context?

If the intervention was implemented in the generalization context, coders recorded *Yes*, if the intervention was not implemented, they coded *No*, and if part of the intervention was implemented, they coded *Partial*.

What was the Generalization Context?

If generalization was measured with a different adult, coders recorded *person*. If generalization was measured in a different location, they coded *setting*. If generalization was measured with a different activity, they coded *task*. If multiple elements were different in the generalization context, coders recorded each element that was different, and if the context was unclear, they coded *unclear*.

Table 1 Strategies to facilitate maintenance and generalization

Citation	Strategy	Definition
Maintenance		
Broden et al. (1971)	Introduce naturally maintaining contingencies	Gradually introduce or condition contingencies in the natural environment and/or more natural contingencies
Freeland and Noell (2002)	Delayed Reinforcement	Increase the time until reinforcement is delivered
Hoch et al. (2002)	Use intermittent reinforcement	Reinforce on a variable schedule
Holman and Baer (1979)	Teach self-management	Teach the individual to monitor his/her own progress
Koegel and Rincover (1977)	Schedule thinning	Gradually decrease the rate of reinforcement
Mank and Horner (1987)	Teach the learner to recruit reinforcement	Teach the learner strategies to obtain reinforcement post-intervention
Paine et al. (1982)	Teach behaviors likely to encounter reinforcement	Teach behaviors that will be reinforced post-intervention to the level necessary to contact natural reinforcement
Pace et al. (1994)	Demand fading	Reduce then gradually increase the amount of work required to obtain reinforcement
Rusch and Kazdin (1981)	Partial/sequential withdrawal	Gradually withdraw intervention components
Generalization		
Broden et al. (1971)	Ask people in the generalization setting to reinforce	Ask people already in the generalization setting to reinforce the target behavior
Dunlap and Johnson (1985)	Use intermittent reinforcement	Reinforce on a variable schedule
Holman and Baer (1979)	Teach self-management	Teach the individual to monitor his/her own progress
Kirby and Bickel (1988)	Teach loosely	Vary non-critical elements of the training setting
Ninness et al. (1991)	Instruct to generalize	Remind the learner to generalize behavior to the generalization setting
Schwarz and Hawkins (1970)	Delay reinforcement	Deliver reinforcement after a period of time has elapsed
Seymour and Stokes (1976)	Teach the learner to recruit reinforcement	Teach the learner strategies to obtain reinforcement in the generalization setting
Stokes et al. (1974)	Teach sufficient examples	Teach the learner to respond under different stimulus conditions
Van den Pol et al. (1981)	Contrive a mediating stimulus	Bring the target behavior under control of a stimulus in the training setting that can be added to the generalization setting
Van den Pol et al. (1981)	Program common stimuli	Make the training and generalization contexts similar
Wahler (1969)	Sequential modification	Train in subsequent contexts if generalization does not occur

Was the Generalization Context Probed During Baseline?

Coders recorded *Yes* if the generalization context was probed during baseline, and *No* if the generalization context was not probed during baseline.

How Many Intervention Sessions Occurred Prior to the First Generalization Probe?

Coders counted the total number of probes other than baseline that occurred prior to the first generalization probe. If an intervention probe and generalization probe appeared on the same day or session on the graph, coders recorded "0."

How Many Probes Did the Researchers Collect?

Coders counted the total number of generalization probes.

What Behavioral Principles Did Researchers Include in BIPs?

Coders recorded the presence or absence of behavioral momentum (e.g., high-probability request sequence), differential reinforcement, extinction, feedback, MO manipulation, NCR, prompting, punishment, and teaching a skill.

Did Researchers Program for Generalization?

We identified strategies to promote generalization using the procedures described above for maintenance. Coders recorded which of the following strategies were included, as described in Table 1: (1) *delay reinforcement*, (2) *teach sufficient exemplars*, (3) *program common stimuli*, (4) *teach loosely*, (5) *ask people in the generalization setting to reinforce the behavior*, (6) *teach the learner to recruit reinforcement*, (7) *contrive a mediating stimulus*, (8) *teach self-management*, (9) *instruct to generalize*, (10) *sequential modification*, (11) *teach behaviors to a level likely to encounter reinforcement*, or (12) *use intermittent reinforcement*.

Did the Authors Conclude Generalization Occurred?

If the authors stated or implied that generalization occurred, coders recorded *Yes*. If the authors stated or implied that generalization did not occur, coders recorded *No*. If the authors stated their results without interpretation, coders recorded *Unclear*.

Calculation of Intervention Effectiveness

Percentage of non-overlapping data (PND) is a measure of overlap between phases recommended for single-subject research (Scruggs et al. 1987). PND is calculated by determining the percentage of data in the phase of interest, typically

the intervention phase, that do not overlap with the highest baseline data point if the goal is to increase behavior or the lowest baseline data point if the goal is to decrease behavior (Scruggs et al. 1987). We used PND because single-subject data violate the assumption of independence required by parametric approaches and in order to facilitate comparison between our findings and the findings in a previous meta-analysis (Goh and Bambara 2012). We calculated the effectiveness of the intervention in terms of producing maintenance by calculating PND on overall behavior change from baseline to maintenance following the procedures described by Goh and Bambara (2012). For studies that included generalization data, we only calculated PND for studies that provided baseline data in the generalization context. If a study reported generalization data for more than one context, we reported the average PND across contexts.

Coding Inter-rater Agreement

A second coder, a PhD student with experience in behavior analysis, coded 36% of studies that included maintenance data (38% of participants) and 63% of studies that included generalization data (57% of participants). Agreement was scored when both coders recorded the same code for the same participant. Coders practiced on articles that were not used for inter-rater agreement, and then, the first author randomly selected articles to code for inter-rater agreement. IRA was calculated as agreements over agreements plus disagreements. For maintenance, agreement was 81% for term, 100% for whether the intervention was withdrawn, 73% for time lapse to first probe, 100% for number of probes, 95% for maintenance strategies, 86% for behavioral principles, 100% for whether authors concluded maintenance occurred, and 100% for PND. Agreement was presumably low for time lapse to first probe because of a lack of clarity in many manuscripts: authors often reported data in sessions rather than days, or reported the number of sessions per week without specifying on which days those sessions occurred. For generalization, agreement was 100% for whether the intervention was implemented in the generalization context, 100% for whether the generalization context was probed during baseline, 92% for number of probes, 100% for number of probes prior to first generalization probe, 100% for generalization context, 100% for generalization strategies, 100% for generalization principles, 92% for whether authors concluded generalization occurred, and 100% for PND. Coders resolved disagreements by consensus for the final coding.

Results

Fifty-four participants in 23 studies met the inclusion criteria for maintenance, and 14 participants in eight studies met the inclusion criteria for generalization. Studies were published in 14 journals; the journals most frequently included were *Journal of Positive Behavioral Interventions, Behavioral Disorders*, and *Beyond Behavior*.

Maintenance

Table 2 summarizes the studies that included maintenance or follow-up data. The first two categories, which term was used and whether the intervention continued into the maintenance/follow-up phase, are discussed at the study level, and the remaining five categories are discussed at the participant level.

Authors of eight studies (35%) used the term maintenance, six studies (26%) used the term follow-up, eight studies (35%) used both terms to describe the same phase, and one study (4%) included both a follow-up phase and a maintenance phase. Researchers implemented the intervention in the maintenance/follow-up phase in 12 studies (52%), partially withdrew the intervention in the maintenance/follow-up phase in four studies (17%), and withdrew the intervention in the maintenance/follow-up phase in four studies (17%). Coders could not determine whether the intervention continued in the maintenance phase in three studies (13%). Neither term appeared to consistently indicate that the intervention was or was not withdrawn.

For 16 participants (30%), coders could not determine how much time elapsed between the last intervention probe and the first maintenance probe. For the remaining 40 participants, between 1 day and 5 months elapsed between intervention and maintenance (median = 15 days). Researchers collected between one and 63 maintenance probes (median = 4).

Forty participants' BIPs included differential reinforcement (74%), 22 included prompting (41%), 18 included skill training (33%), 18 included MO manipulation (33%), 15 included extinction (28%), two included non-contingent reinforcement (4%), one included behavioral momentum (2%), and one included restraint (2%). Coders could not identify a strategy to facilitate maintenance in 31 BIPs (55%). Among the BIPs that included identified strategies, 18 included self-monitoring (32%), nine included partial-sequential withdrawal (16%), nine included teaching the learner to recruit reinforcement (16%), and five included introducing naturally maintaining contingencies (9%). Some BIPs included more than one strategy.

Authors concluded that maintenance occurred for 52 participants (93%) and did not occur for two participants (4%). For one participant, the authors concluded that maintenance did not occur because the teacher's procedural fidelity was low, but did occur after fidelity improved (Aitken et al. 2011). For one participant, coders could not determine whether the authors concluded that maintenance occurred. PND ranged from 0 to 100%, with a median of 100% and a mean of 80%.

Generalization

Table 3 summarizes the studies that included generalization data. Two categories, whether researchers implemented the intervention in the generalization context and whether researchers probed the generalization context during baseline, are discussed at the study level; the remaining categories are discussed at the participant level.

Researchers in three studies (38%) implemented the intervention in the generalization context; researchers in three studies (38%) did not. In one study (13%),

Table 2 Summary of studies that included maintenance data

Citation	Participant	RQ #1		RQ #2		RQ #3		RQ #4	
		Term	Continued intervention?	Time lapse	No. of probes	Principles	Programmed?	Occurred?	PND (%)
Aitken et al. (2011)	Caleb	M	Partial	29 days	6	DR, extinction, prompt, skill	Self-monitoring	No, so "refresher" then yes	100
Athens and Vollmer (2010)	George	Both, same	Yes	1 month	5	DR	No/unclear	Yes	100
	Clark	Both, same	Yes	2 months	5	DR	No/unclear	Yes	0
Campbell and Tincani (2011)	Lori	Maintenance	No	1 day	14	Skill, prompt	Partial/sequential withdrawal	No	21
	James	Both, different	No	1 day (maintenance probe), 8 weeks (follow-up)	12	Skill, prompt	Partial/sequential withdrawal	Yes	92
	Shawn	Both, different	No	1 day (maintenance probe), 8 weeks (follow-up)	7	Skill, prompt	Partial/sequential withdrawal	Yes	86
Cannella-Malone et al. (2011)	William	Both, same	Unclear	3 weeks	3	MO	Self-monitoring, partial/sequential withdrawal	Yes	100
	Reese	Both, same	Unclear	3 weeks	3	MO	Self-monitoring, partial/sequential withdrawal	Yes	100
	Lonny	Both, same	Unclear	3 weeks	3	MO	Self-monitoring, partial/sequential withdrawal	Yes	100
Casey and Merical (2006)	Karl	F/U	Yes	5 months	3	DR, prompt	No/unclear	Yes	100

Table 2 (continued)

Citation	Participant	RQ #1		RQ #2		RQ #3		RQ #4	
		Term	Continued intervention?	Time lapse	No. of probes	Principles	Programmed?	Occurred?	PND (%)
Casey et al. (2008)	Joey	F/U	Yes	1 month	63	Behavioral momentum, DR, extinction, NCR, prompt, restraint	No/unclear	Yes	83
Cho Blair et al. (2007)	Minsu	F/U	No	Unclear	6	DR, extinction, skill, MO	No/unclear	Yes	100
DeJager and Filter (2015)	Hank	Both, same	Yes	4 weeks	1	DR, extinction, prompt, skill	No/unclear	No	100
	Charlie	Both, same	Yes	1 week	1	DR, extinction, prompt, skill	No/unclear	Yes	100
Germer et al. (2011)	David	M	Yes	12 days	3	DR, prompt	Self-monitoring	Unclear	100
Janney et al. (2012)	Hugo	F/U	Yes	Unclear	3	DR, extinction, MO	Partial/sequential withdrawal (intervention reinstated)	Yes	100
	Tomas	F/U	Yes	Unclear	6	DR, extinction, prompt	Partial/sequential withdrawal (intervention reinstated)	Yes	100
	Eric	F/U	Yes	Unclear	4	DR, extinction, MO	Partial/sequential withdrawal (intervention reinstated)	Yes	100

Table 2 (continued)

Citation	Participant	RQ #1		RQ #2		RQ #3		RQ #4	
		Term	Continued intervention?	Time lapse	No. of probes	Principles	Programmed?	Occurred?	PND (%)
Lane et al. (2007)	Aaron	M	Partial	7 weeks	3	DR, extinction, MO, prompt	Yes/unclear	Yes	0
Lane et al. (2009)	Derek	M	Unclear	1 month	3	DR, extinction, MO	Self-monitoring, teach learner to recruit SR+	Yes	0
	Mark	M	Unclear	1 month	2	DR, extinction	Self-monitoring, teach the learner to recruit SR+	Yes	100
Lo and Cartledge (2006)	Ted	M	Partial	Unclear	1	DR, skill	Self-monitoring, teach learner to recruit SR+	Yes	100
	Adam	M	Partial	Unclear	3	DR, skill	Self-monitoring, teach learner to recruit SR+	Yes	67
	Chad	M	Partial	Unclear	7	DR, skill	Self-monitoring, teach learner to recruit SR+	Yes	0
	Sam	M	Partial	Unclear	1	DR, skill	Self-monitoring, teach learner to recruit SR+	Yes	100
Losinski et al. (2015)	Alexandra	M	Yes	Unclear	8	MO	No/unclear	Yes	88
	Brenda	M	Yes	Unclear	4	MO	No/unclear	Yes	75
	Larry	M	Yes	Unclear	5	MO	No/unclear	Yes	40
	Hannah	M	Yes	Unclear	5	MO	No/unclear	Yes	80

Table 2 (continued)

Citation	Participant	RQ #1		RQ #2		RQ #3		RQ #4	
		Term	Continued intervention?	Time lapse	No. of probes	Principles	Programmed?	Occurred?	PND (%)
Majeika et al. (2011)	Andrew	M	Partial	5 weeks	2	DR, extinction, MO, prompt	Self-monitoring	Yes	50
Mueller and Nkosi (2007)	Bruce	F/U	Yes	5 weeks	23	DR, prompt	No/unclear	Yes	100
Payne et al. (2007)	Amy	M	Yes	1 week	1	DR	No/unclear	Yes	100
	Julie	M	Yes	1 week	1	DR	No/unclear	Yes	100
Reeves et al. (2013)	Ron	Both, same	Yes	4 days	3	DR, prompt	Self-monitoring, introduce to naturally maintaining contingencies, teach learner to recruit SR+	Yes	100
	Sam	Both, same	Yes	2 days	3	DR, prompt	Self-monitoring, introduce to naturally maintaining contingencies, teach learner to recruit SR+	Yes	100
	Joe	Both, same	Yes	2 days	3	DR, prompt	Self-monitoring, introduce to naturally maintaining contingencies, teach the learner to recruit SR+	Yes	100

Table 2 (continued)

Citation	Participant	RQ #1		RQ #2		RQ #3		RQ #4	
		Term	Continued intervention?	Time lapse	No. of probes	Principles	Programmed?	Occurred?	PND (%)
Restori et al. (2007)	A1	Both, same	Unclear	15 days	4	DR, prompt	Self-monitoring	Yes	100
	A2	Both, same	Unclear	15 days	4	DR, prompt	Self-monitoring	Yes	100
	A3	Both, same	Unclear	15 days	4	DR, MO manipulation	No/unclear	Yes	100
	A4	Both, same	Unclear	15 days	4	DR, MO manipulation	No/unclear	Yes	100
	C1	Both, same	Unclear	15 days	4	DR, extinction	No/unclear	Yes	0
	C2	Both, same	Unclear	15 days	4	DR, extinction	No/unclear	Yes	50
	C4	Both, same	Unclear	15 days	4	DR, extinction	No/unclear	Yes	100
Skinner et al. (2009)	Austin	M	Yes	"at least 1 month"	3	NCR	No/unclear	Yes	100
Sprague and Perkins (2009)	Participant 1	Both, same	No	1 day	10	DR, feedback, skill, punishment	No/unclear[a]	Yes	100
	Participant 2	Both, same	No	1 day	11	DR, feedback, skill, punishment	No/unclear[a]	Yes	100
	Participant 3	Both, same	No	1 day	4	DR, Feedback, Skill, Punishment	No/unclear[a]	Yes	100
	Participant 4	Both, same	No	1 day	6	DR, feedback, skill, punishment	No/unclear[a]	Yes	100
Stichter et al. (2009)	Jason	Both, same	No	Unclear	4	MO	No/unclear	Yes	50
	Nigel	Both, same	No	1 week	3	Prompt	No/unclear	Yes	17

Table 2 (continued)

Citation	Participant	RQ #1		RQ #2		RQ #3		RQ #4	
		Term	Continued intervention?	Time lapse	No. of probes	Principles	Programmed?	Occurred?	PND (%)
Strain et al. (2011)	Caleb	Both, same	No	"2–4 weeks"	6	MO	No/unclear	Yes	0
	Josh	Both, same	Yes	Unclear	4	DR, prompt, skill	Self-monitoring	Yes	100
	Alex	Both, same	Yes	Unclear	3	DR, MO, prompt, skill	Introduce to natural maintaining contingencies	Yes	100
	Rosalie	Both, same	Yes	Unclear	3	DR, prompt, skill	Introduce to natural maintaining contingencies	Yes	100

[a]Length of intervention session was extended, but it was unclear whether reinforcement schedule or demands were affected (schedule thinning or demand fading)

Table 3 Summary of studies that included generalization data

Citation	Participant	RQ #1		RQ # 2			RQ # 3		RQ #4	
		Intervention implemented?	Context	Probed in BL?	No. of probes	Sessions before gen	Principles	Programming	Occurred?	PND
Cho Blair et al. (2006)	Minsu	Partial	Setting and person	Yes	7	5	MO, skill	Train sufficient exemplars	Yes	100%
	Nari	Partial	Setting and person	Yes	6	6	MO, skill	Train sufficient exemplars	Yes	100%
	Hyun	Partial	Setting and person	Yes	7	3	MO, skill	Train sufficient exemplars	Yes	100%
Casey et al. (2008)	Joey	Yes	Setting and person	No	Unclear	78	Behavioral momentum, DR, extinction, NCR, prompt, restraint	Train sufficient exemplars	Yes	[a]
Davis et al. (2012)	Mary	Yes	Task and person	No	5	22	DR, prompt	Delay SR +, ask people in the gen setting to reinforce	Yes	[a]
Germer et al. (2011)	David	Unclear	Task	No	3	5	DR, prompt	Self-monitoring	Unclear	[a]
Groskreutz et al. (2014)	Brian	No	Task	Yes	30	0	DR, extinction, Prompt	Sequential modification, train sufficient exemplars, teach the learner to recruit SR+	Yes	33%

Table 3 (continued)

Citation	Participant	RQ #1		RQ # 2			RQ # 3		RQ #4	
		Intervention implemented?	Context	Probed in BL?	No. of probes	Sessions before gen	Principles	Programming	Occurred?	PND
Lo and Cartledge (2006)	Ted	No	Setting	Yes	28	0	DR, skill	Self-monitoring, teach the learner to recruit SR+	Yes	45%
	Adam	No	Setting	Yes	32	0	DR, skill	Self-monitoring, teach the learner to recruit SR+	Yes	0%
	Chad	No	Setting	Yes	22	0	DR, skill	Self-monitoring, teach the learner to recruit SR+	Yes	0%
	Sam	No	Setting	Yes	26	0	DR, skill	Self-monitoring, teach the learner to recruit SR+	Yes	42%
Majeika et al. (2011)	Andrew	No	Task, person, and task	Yes	4	3	DR, skill	Self-monitoring	Unclear	100%**
Mueller and Nkosi (2007)	Bruce	Yes	Setting and person	Yes	101	63	DR, extinction, MO, prompt	Sequential modification, ask people in the gen setting to reinforce	Yes	100%

Table 3 (continued)

Citation	Participant	RQ #1		RQ # 2			RQ # 3		RQ #4	
		Intervention implemented?	Context	Probed in BL?	No. of probes	Sessions before gen	Principles	Programming	Occurred?	PND
	Meredith	Yes	Setting, person, and task	Yes	59	12	DR, Prompt	Sequential modification, ask people in the gen setting to reinforce	Yes	25%

[a]PND not calculated for studies that did not probe the generalization setting during baseline

**Calculated PND in one generalization setting that had baseline data, but not another context that had no baseline data

researchers partially implemented the intervention in the generalization context, and in one study (13%), coders could not determine whether researchers implemented the intervention in the generalization context.

Researchers in five studies (63%) probed the generalization context during baseline, and researchers in three studies (38%) did not. The number of generalization probes ranged from 3 to 101 (median = 7). In one study, the number of probes was unclear because the graph did not indicate which data were collected in the generalization setting. The number of intervention probes prior to the first baseline probe ranged from zero to 78 (median = 5). Researchers measured generalization to a different setting for four participants (29%) and to a different task for two participants (14%). Five participants' generalization contexts involved a different setting and person (36%); two involved a different person, setting, and task (14%); and one involved a different person and task (7%).

Eight students' BIPs included differential reinforcement (57%), seven included skill training (50%), four included MO manipulation (29%), four included prompting (29%), two included extinction (14%), one included behavioral momentum (7%), and one included restraint (7%). Six students' BIPs included self-monitoring (43%), five included training sufficient exemplars (36%), and five included teaching the learner to recruit reinforcement (36%). Three included sequential modification (21%), three included asking people in the generalization setting to reinforce the desired behavior (21%), and one included delayed reinforcement (7%). Researchers concluded generalization occurred for 12 participants (86%), and coders could not determine researchers' conclusions for two participants (14%). No researcher stated generalization did not occur. PND ranged from 0 to 100%, with a median of 45% and a mean of 50%.

Discussion

We examined how researchers defined, measured, and programmed for maintenance and generalization and whether they concluded maintenance or generalization occurred in studies on FBA-based BIPs in school settings. Overall, findings reveal that, among the included studies, 69% of researchers implemented all or part of the intervention in the maintenance phase and 51% implemented all or part of the intervention in the generalization context. Some researchers specifically programmed for maintenance and generalization in their school-based BIPs, but many did not, leaving a number of strategies from the applied behavior analytic literature that support maintenance and generalization apparently untested in recent published school-based research (e.g., teaching loosely, programming indiscriminable contingencies, schedule thinning, demand fading, and programming common stimuli).

In their meta-analysis, Goh and Bambara (2012) found little overlap between baseline and maintenance data and between baseline and generalization data in school-based FBA/BIP research, which they interpreted as indicating generalized outcomes occurred. Like Goh and Bambara, we found a median of 100% PND between baseline and intervention for maintenance, and unlike Goh and Bambara, we found a median of 45% PND between baseline and generalization. This lower

finding might be related to probing generalization across multiple stimulus contexts (e.g., setting and person) or to a lower percentage of articles implementing all or part of the intervention in the generalization context(s). Goh and Bambara (2012) did not consider how researchers defined maintenance and generalization. By contrast, we coded how researchers defined maintenance and generalization. Based on our findings, it appears difficult to draw conclusions about the extent to which generalized outcomes occurred in school-based studies because researchers measured different outcomes in their maintenance phases. In the sections below, we discuss issues, limitations, and future directions related to definitions, measurement, and strategies.

Variation in Definitions

To determine how researchers defined maintenance and generalization, we examined whether they continued to implement intervention in the maintenance/follow-up phase or in the generalization context. In the majority of studies in this review, the researchers implemented the intervention in the maintenance phase or generalization context. Thus, many researchers measured whether the intervention continued to work across time or in multiple contexts, rather than whether the target behavior occurred absent the intervention. This approach is inconsistent with the rigorous applied behavior analytic definition of maintenance and generalization as trained behavior occurring in the absence of all or part of the intervention responsible for the behavior change (Cooper et al. 2007). Similarly, Cuvo (2003) reported that stimulus generalization has been defined as behavior reinforced in the presence of one stimulus occurring in the presence of another stimulus under extinction conditions. However, studies in the applied literature have reported prompting and reinforcement occurring in generalization conditions (Cuvo 2003). Cuvo (2003) recommended resolving this disparity by adhering to standard definitions of terms.

In the studies included in this review, some researchers used the term *follow-up* instead of or in addition to the term maintenance. However, across studies no relation was apparent between term and definition; neither the term *maintenance* nor the term *follow-up* consistently indicated the intervention was withdrawn. We propose the term *maintenance* be reserved for measuring behavior following the full withdrawal of the intervention, the term *fading* be used to describe partially withdrawing the intervention, and the term *follow-up* be used for measuring continued intervention effectiveness. Adherence to the above definitions would offer several benefits. First, this adherence would require researchers to make a clear choice about which outcome they desire to measure. For example, a researcher would have to decide whether a study goal was for a teacher to sustain intervention implementation over time (follow-up), or for the effects of the BIP to continue after removing the intervention (maintenance). Second, it would require researchers to clarify which outcome they were measuring, which would facilitate examining results across studies to determine which strategies continue to produce results while they are in place and which strategies produce lasting post-intervention results. Third, it would encourage programming for maintenance rather than relying on continued intervention.

Similarly, although researchers may want to know whether an intervention works in multiple contexts, we assert that the term *generalization* should be used when measuring behavior in untrained contexts.

Maintenance, fading, and follow-up, as defined above, may each be appropriate in different situations. For example, if an intervention involves teaching a student to use a planner to monitor work completion, that student might continue to use the planner throughout school and after graduation, and follow-up would be an appropriate measure of long-term effectiveness. Alternatively, a teacher may use frequent consequences to increase on-task behavior and then may choose to fade the consequences to a more feasible level without ever removing them entirely. We expect that many practitioners aim to produce behavior change that extends beyond the intervention, and maintenance, as defined above, is often the most appropriate measure of long-term change. We suggest future researchers examine different ways to achieve long-term change. For example, a future review could compare contexts in which continuing to implement the intervention produces long-term change to contexts in which fading the intervention is necessary to produce long-term behavior change. This review could contribute to refining the definition of maintenance (e.g., identifying components of interventions that are sustainable over time), which would inform practitioners as they program BIPs.

Variation in Measurement

The number of maintenance and generalization probes collected varied across studies. The number of probes should differ depending on each particular context and on the stability of the data. Decisions about measurement should be purposeful and guided by data, but although researchers presumably make purposeful and data-based decisions, the rationale for decisions about measurement was not always clear. By explaining measurement decisions, researchers can provide information to guide practitioners. Baer (1981) recommended practitioners plan for generalized outcomes by identifying the contexts in which the learner should use the target behavior and how long the target behavior should continue after the intervention. Researchers should create a similar plan, and the plan should guide decisions about how long to measure maintenance and how many probes to collect. Additionally, this review reveals a need for studies that measure long-term maintenance; no researchers showed maintenance data across school years on their graph, and only one researcher measured maintenance more than a year after the intervention. (One excluded study measured maintenance across school years, Kern et al. 2006.)

The amount of time that elapsed between the last intervention probe and first maintenance probe also varied across studies. In one study (Casey and Merical 2006), researchers waited 5 months, while in another study (Campbell and Tincani 2011), researchers began maintenance probes the following day. Researchers who wish to investigate whether treatment effects maintain after intervention should conduct the initial maintenance probe shortly after withdrawing the intervention; that way, if maintenance fails to occur, they can reinstate the intervention or adjust the plan. Similarly, the number of intervention sessions that occurred prior to the first

generalization probe varied across studies. Some researchers (e.g., Lo and Cartledge 2006) continuously probed in both the intervention and generalization contexts, while other researchers (e.g., Mueller and Nkosi 2007) probed the generalization context after demonstrating successful behavior change in the initial context. Researchers may wish to probe regularly in the generalization context to identify when generalization occurs, but if resources are limited, we recommend researchers or practitioners to probe the generalization context shortly after achieving behavior change in the initial context. By probing shortly after achieving behavior change, researchers and practitioners can determine whether generalization occurred quickly in order to modify the intervention to increase the likelihood of generalization if it does not occur. A limitation of the present review is that we did not code the amount of time between probes or the total amount of time for which researchers measured maintenance, follow-up, or generalization. Future researchers might consider the total lengths of maintenance and generalization phases.

Measuring behavior over long periods of time or across multiple contexts requires considerable time. Researchers and practitioners may consider including *challenges* to assess maintenance and generalization (see Wacker et al. 2011). For example, after demonstrating maintenance occurred, researchers or practitioners could present a difficult task, have the teacher step out of the classroom, or have a new adult play the role of substitute teacher. Although behavior during challenges might not represent behavior under typical conditions, challenges are a promising approach for measuring maintenance and generalization efficiently. More research is needed to evaluate the utility of challenges to assess maintenance and generalization in school contexts.

More than one-third of the researchers who measured generalization measured behavior in the generalization context only after intervention started. Future researchers should probe the generalization context during baseline (Cooper et al. 2007). Behavior must be measured in the generalization context prior to intervention to determine whether generalization occurred; otherwise, whether the learner engaged in the target behavior in the generalization context prior to intervention remains unknown.

Strategies to Facilitate Maintenance and Generalization

The behavioral principles most frequently incorporated into BIPs were differential reinforcement, skill training, prompting, and MO manipulation. Of the studies that included strategies to facilitate generalized outcomes, most included self-monitoring. Self-monitoring is potentially the most effective strategy to facilitate generalized behavior change, because learners themselves are always present and self-monitoring can facilitate both maintenance and generalization (Cooper et al. 2007). However, Baer et al. (1984) cautioned that teaching self-monitoring does not ensure it will be used and that the act of self-monitoring is an additional response the learner must maintain and generalize. Although self-monitoring holds promise, additional strategies must be considered.

Other than self-monitoring, few strategies to facilitate maintenance were examined. Of the 58 participants for whom maintenance was measured, nine participants' interventions included a strategy that involved systematically withdrawing the intervention. In two studies, researchers taught learners to recruit reinforcement by raising their hands to obtain attention (Lo and Cartledge 2006; Reeves et al. 2013). Teaching students to recruit reinforcement could be viewed as related to introducing naturally maintaining contingencies.

In the current review, coders recorded "introduce to naturally maintaining contingencies" if researchers explained a plan to teach behavior likely to encounter natural, functional reinforcement to the level necessary to obtain that reinforcement; only 9% of BIPs included such a plan. Many researchers increased on-task behavior. Although increasing on-task behavior is often appropriate, and natural reinforcers like grades and assignment completion follow on-task behavior, researchers should not assume natural reinforcers will maintain on-task behavior after intervention. Instead, prior to training, researchers should observe the classroom to evaluate existing contingencies and plan their intervention accordingly.

Researchers who measured generalization frequently used sequential modification and teaching sufficient exemplars. Sequential modification, described by Stokes and Baer (1977), involves measuring behavior continuously in multiple contexts. If behavior fails to generalize, the interventionist trains in subsequent contexts. However, Stokes and Baer (1977) noted that studies that use this method typically exhaust all contexts by training in each context; thus, researchers are unable to determine whether behavior generalizes to untrained contexts. Later researchers stated sequential modification might be a description of a multiple-baseline design rather than a procedure to facilitate generalization and should be used only after other attempts at programming for generalization have failed (Stokes and Osnes 1989; Tiger et al. 2008). In order to measure generalization to untrained contexts using sequential modification, researchers must withhold training in at least one context (e.g., Groskreutz et al. 2014). One study involved making the reinforcement contingency difficult to discriminate (delayed reinforcement; Davis et al. 2012), and two involved asking people in the generalization setting to reinforce the behavior (Casey et al. 2008; Mueller and Nkosi 2007). This may be a useful strategy in schools (e.g., asking a cafeteria worker to reinforce hand raising), but researchers and practitioners should consider whether it is plausible to reinforce the behavior in all relevant settings.

It is difficult to draw conclusions across studies about the extent to which tested strategies facilitated maintenance and generalization due to the variation in researchers implementing and not implementing the intervention in the maintenance and generalization phases. Thus, although some strategies have been tested numerous times in recent literature in this review (e.g., self-monitoring), and although a majority of researchers have concluded those strategies produced maintenance, further research is needed to evaluate the extent to which various strategies produce maintenance absent intervention and which produce continued intervention effectiveness. Additionally, the effects of many strategies to support maintenance and generalization (i.e., teaching loosely, programming indiscriminable contingencies, schedule thinning, demand fading, and programming common stimuli) have not been tested

in recent school-based FBA/BIP literature. Thus, the extent to which many promising strategies produce maintenance and generalization in FBA-based BIPs in schools remains unknown. Future researchers should study strategies like those listed above (see Table 1 for definitions) that have not been examined in recent school-based literature.

Limitations

The present review has several limitations. This review did not evaluate whether researchers measured fidelity. Fidelity may impact the extent to which maintenance and generalization occur (Wood et al. 2007). Future research is warranted to examine the potential relationship between implementation fidelity and maintenance and/or generalization outcomes. Also, the current review provided PND as a measure of effect size. We selected this metric to replicate the analysis completed in the Goh and Bambara's (2012) review. However, we acknowledge that as effect sizes for single-case design evolve, no consensus exists on the best effect size for single-case data (Chen et al. 2016).

Implications and Conclusion

Findings from the present review highlight the need for practitioners and researchers to consider maintenance and generalization when designing BIPs. Practitioners who desire maintenance and generalization should plan for those outcomes. Given the small number of studies that measured maintenance and generalization absent intervention, practitioners have limited resources for determining how to effectively program a BIP to produce post-intervention behavior change.

Based on the findings of this review, we recommend practitioners to consider maintenance and generalization from the outset when planning BIPs. Practitioners should consider the variety of conditions under which they desire to produce behavior change (e.g., general education classroom, special education classroom, lunch, recess), and they should evaluate the contingencies operating in those environments to determine how to best produce behavior change across contexts and over time. Then, practitioners can determine whether to implement the BIP in all contexts, or to implement the BIP in one context and measure whether behavior change generalizes to other relevant contexts. Additionally, practitioners should try withdrawing or fading the BIP after achieving behavior change to evaluate whether maintenance occurs after the BIP is no longer in place. If practitioners determine behavior change does not maintain after the BIP is faded or withdrawn, they can reinstate the BIP and modify it with strategies likely to produce maintenance. In the words of Baer et al. (1968, p. 97): "In general, generalization should be programmed, rather than expected or lamented."

Funding The research described in this article was supported in part by Grant H325H140001 from the Office of Special Education Programs, U.S. Department of Education. Nothing in the article necessarily reflects the positions or policies of the federal government, and no official endorsement by it should be inferred.

Compliance with Ethical Standards

Conflicts of interest The authors declare no conflicts of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Aitken, A., Harlan, A., Hankins, K., Michels, J., Moore, T. C., Oakes, W. P., et al. (2011). Increasing academic engagement during writing activities in an urban elementary classroom. *Beyond Behavior, 20,* 38–43.

Athens, E. S., & Vollmer, T. R. (2010). An investigation of differential reinforcement of alternative behavior without extinction. *Journal of Applied Behavior Analysis, 43,* 569–589.

Baer, D. M. (1981). *How to plan for generalization*. Austin, TX: Pro-Ed.

Baer, D. M., Stolz, S. B., & Fowler, S. A. (1984). *Self-management tactics for the developmentally disabled*. New York City: Pergamon Press.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91–97.

Broden, M., Hall, R. V., & Mitts, B. (1971). The effect of self-recording on the classroom behavior of two eighth-grade students. *Journal of Applied Behavior Analysis, 4,* 191–199.

Campbell, A., & Tincani, M. (2011). The power card strategy: Strength-based intervention to increase direction following of children with Autism Spectrum Disorder. *Journal of Positive Behavior Interventions, 13,* 1–10.

Cannella-Malone, H. I., Tullis, C. A., & Kazee, A. R. (2011). Using antecedent exercise to decrease challenging behavior in boys with developmental disabilities and an emotional disorder. *Journal of Positive Behavior Interventions, 13,* 230–239.

Casey, S. D., & Merical, C. L. (2006). The use of functional communication training without additional treatment procedures in an inclusive school setting. *Behavioral Disorders, 32,* 46–54.

Casey, S. D., Perrin, C. J., Merical, C. L., Lecomte, J. M., Milligan, J., & Walsh-Czekalski, M. (2008). Increasing bite acceptance and reducing food refusal in a child with Autism: Moving beyond the clinic. *Journal of Behavior Analysis in Health, Sports, Fitness, and Medicine, 1,* 34–44.

Chen, M., Hyppa-Martin, J. K., Reichle, J. E., & Symons, F. J. (2016). Comparing single case design overlap-based effect size metrics from studies examining speech generating device interventions. *American Journal on Intellectual and Developmental Disabilities, 121,* 169–193.

Cho Blair, K. S., Liaupsin, C. J., Umbreit, J., & Kweon, G. (2006). Function-based intervention to support the inclusive placements of young children in Korea. *Education and Training in Developmental Disabilities, 41,* 48–57.

Cho Blair, K. S., Umbreit, J., Dunlap, G., & Jung, G. (2007). Promoting inclusion and peer participation through assessment-based intervention. *Teaching Early Childhood Special Education, 273,* 134–147.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson.

Cuvo, A. J. (2003). On stimulus generalization and stimulus classes. *Journal of Behavioral Education*, *12*, 77–83.

Davis, D. H., Fredrick, L. D., Alberto, P. A., & Gama, R. (2012). Functional communication training without extinction using concurrent schedules of differing magnitudes of reinforcement in classrooms. *Journal of Positive Behavior Interventions, 14,* 162–172.

DeJager, B. W., & Filter, K. J. (2015). Effects of prevent-teach-reinforce on academic engagement and disruptive behavior. *Journal of Applied School Psychology, 31,* 369–391.

Dunlap, G., & Johnson, J. (1985). Increasing the independent responding of autistic children with unpredictable supervision. *Journal of Applied Behavior Analysis, 3,* 227–236.

Freeland, J. T., & Noell, G. H. (2002). Programming for maintenance: An investigation of delayed intermittent reinforcement and common stimuli to create indiscriminable contingencies. *Journal of Behavioral Education, 11,* 5–18.

Germer, K. A., Kaplan, L. M., Giroux, L. N., Markham, E. H., Ferris, G. J., Oakes, W. P., et al. (2011). A function-based intervention to increase a second-grade student's on-task behavior in a general education classroom. *Beyond Behavior, 20,* 19–30.

Goh, A. E., & Bambara, L. M. (2012). Individualized positive behavior support in school settings: A meta-analysis. *Remedial and Special Education, 33,* 271–286.

Groskreutz, N. C., Groskreutz, M. P., Bloom, S. E., & Slocum, T. A. (2014). Generalization of negatively reinforced mands in children with autism. *Journal of Applied Behavior Analysis, 47,* 560–579.

Hoch, H., McComas, J. J., Thompson, A. L., & Paone, D. (2002). Concurrent reinforcement schedules: Behavior change and maintenance without extinction. *Journal of Applied Behavior Analysis, 35,* 155–169.

Holman, J., & Baer, D. M. (1979). Facilitating generalization of on-task behavior through self-monitoring of academic tasks. *Journal of Autism and Developmental Disorders, 9,* 429–446.

Horner, R. H., Dunlap, G., Koegel, R. L., Carr, E. G., Sailor, W., Anderson, J., et al. (1990). Toward a technology of nonaversive behavior support. *Journal of the Association for Persons with Severe Handicaps, 15,* 125–132.

Ingram, K., Lewis-Palmer, T., & Sugai, G. (2005). Function-based intervention planning: Comparing the effectiveness of FBA function-based and non-function-based intervention plans. *Journal of Positive Behavior Interventions, 7,* 224–236.

Janney, D. M., Umbreit, J., Ferro, J. B., Liaupsin, C. J., & Lane, K. L. (2012). The effect of the extinction procedure in function-based intervention. *Journal on Positive Behavior Interventions, 15,* 113–123.

Kern, L., Gallagher, P., Starosta, K., Hickman, W., & George, M. (2006). Longitudinal outcomes of functional behavioral assessment based intervention. *Journal of Positive Behavior Interventions, 8,* 67–78.

Kincaid, D., Dunlap, G., Kern, L., Lane, K. L., Bambara, L. M., Brown, F., et al. (2015). Positive behavior support: A proposal for updating and refining the definition. *Journal of Positive Behavior Interventions, 18,* 69–73.

Kirby, K. C., & Bickel, W. K. (1988). Toward an explicit analysis of generalization: A stimulus control interpretation. *The Behavior Analyst, 11,* 115–129.

Koegel, R. L., & Rincover, A. (1977). Research on the difference between generalization and maintenance in extra-therapy responding. *Journal of Applied Behavior Analysis, 10,* 1–12.

Lane, K. L., Eisner, S. L., Kretzer, J., Bruhn, A. L., Funke, L., Lerner, T., et al. (2009). Outcomes of functional assessment-based interventions for students with and at risk for emotional and behavioral disorders in a job-share setting. *Education and Treatment of Children*, *32*, 573–604.

Lane, K. L., Rogers, L. A., Parks, R. J., Weisenbach, J. L., Mau, A. C., Merwin, M. T., et al. (2007). Function-based interventions for students who are nonresponsive to primary and secondary prevention efforts: Illustrations at the elementary and middle school levels. *Journal of Emotional and Behavioral Disorders, 15,* 169–183.

Lo, Y., & Cartledge, G. (2006). FBA and BIP: Increasing the behavior adjustment of African American boys in schools. *Behavioral Disorders, 31,* 147–161.

Losinski, M., Maag, J. W., Katsiyannis, A., & Ryan, J. B. (2015). The use of structural behavioral assessment to develop interventions for secondary students exhibiting challenging behaviors. *Education and Treatment of Children, 38,* 149–174.

Majeika, C. E., Wälder, J. P., Hubbard, J. P., Steeb, K. M., Ferris, G. J., Oakes, W. P., et al. (2011). Improving on-task behavior using a functional assessment-based intervention in an inclusive high school setting. *Beyond Behavior, 20,* 55–66.

Mank, D. M., & Horner, R. H. (1987). Self-recruited feedback: A cost-effective procedure for maintaining behavior. *Research in Developmental Disabilities, 8,* 91–112.

Mayer, G. R., Sulzer-Azaroff, B., & Wallace, M. (2011). *Behavior analysis for lasting change* (2nd ed.). Cornwall-on-Hudson, NY: Sloan Publishing.

Mueller, M. M., & Nkosi, A. (2007). State of the science in the assessment and management of severe behavior problems in school settings: Behavior analytic consultation to schools. *International Journal of Behavioral Consultation and Therapy, 3,* 176–202.

Ninness, H. A. C., Fuerst, J., Rutherford, R. D., & Glenn, S. S. (1991). Effects of self-management training and reinforcement on the transfer of improved conduct in the absence of supervision. *Journal of Applied Behavior Analysis, 3,* 499–508.

Pace, G. M., Ivancic, M. T., & Jefferson, G. (1994). Stimulus fading as treatment for obscenity in a brain-injured adult. *Journal of Applied Behavior Analysis, 27,* 301–305.

Paine, S. C., Hops, H., Walker, H. M., Greenwood, C. R., Fleischman, D. H., & Guild, J. J. (1982). Repeated treatment effects: A study of maintaining behavior change in socially withdrawn children. *Behavior Modification, 6,* 171–199.

Payne, L. D., Scott, T. M., & Conroy, M. (2007). A school-based examination of the efficacy of function-based intervention. *Behavioral Disorders, 32,* 158–174.

Reeves, L. M., Umbreit, J., Ferro, J. B., & Liaupsin, C. J. (2013). Function-based intervention to support the inclusion of students with autism. *Education and Training in Autism and Developmental Disabilities, 48,* 379–391.

Restori, A. F., Gresham, F. M., Chang, T., Lee, H. B., & Laija-Rodriquez, W. (2007). Functional assessment-based interventions for children at-risk for emotional and behavioral disorders. *The California School Psychologist, 12,* 9–30.

Rusch, F. R., & Kazdin, A. E. (1981). Toward a methodology of withdrawal designs for the assessment of response maintenance. *Journal of Applied Behavior Analysis, 14,* 131–140.

Schwarz, M. L., & Hawkins, R. P. (1970). Application of delayed reinforcement procedures to the behavior of an elementary school child. *Journal of Applied Behavior Analysis, 3,* 85–96.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education, 8,* 24–33.

Seymour, F. W., & Stokes, T. F. (1976). Self-recording in training girls to increas work and evoke staff praise in an institution for offenders. *Journal of Applied Behavior Analysis, 9,* 41–54.

Skinner, J. N., Veerkamp, M. B., Kamps, D. M., & Andra, P. R. (2009). Teacher and peer participation in functional analysis and intervention for a first grade student with attention deficit hyperactivity disorder. *Education and Treatment of Children, 32,* 243–266.

Sprague, J., & Perkins, K. (2009). Direct and collateral effects of the first step to success program. *Journal of Positive Behavior Interventions, 11,* 208–221.

Stichter, J. P., Randolph, J. K., Kay, D., & Gage, N. (2009). The use of structural analysis to develop antecedent-based interventions for students with autism. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-009-0693-8.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349–367.

Stokes, T. F., Baer, D. M., & Jackson, R. L. (1974). Programming the generalization of a greeting response in four retarded children. *Journal of Applied Behavior Analysis, 7,* 599–610. https://doi.org/10.1901/jaba.1974.7-599.

Stokes, T. F., & Osnes, P. G. (1989). An operant pursuit of generalization. *Behavior Therapy, 20,* 337–355. https://doi.org/10.1016/S0005-7894(89)80054-1.

Strain, P. S., Wilson, K., & Dunlap, G. (2011). Prevent-teach-reinforce: Addressing problem behaviors of students with autism in general education classrooms. *Behavioral Disorders, 36,* 160–171.

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

van den Pol, R. A., Iwata, B. A., Ivancic, M. T., Page, T. J., Neef, N. A., & Whitley, F. P. (1981). Teaching the handicapped to eat in public places: Acquisition, generalization and maintenance of restaurant skills. *Journal of Applied Behavior Analysis, 14,* 61–69.

Wacker, D. P., Harding, J. W., Berg, W. K., Lee, J. F., Schieltz, K. M., Padilla, Y. C., et al. (2011). An evaluation of persistence of treatment effects during long-term treatment of destructive behavior. *Journal of Applied Behavior Analysis, 96,* 261–282.

Wahler, R. G. (1969). Setting generality: Some specific and general effects of child behavior therapy. *Journal of Applied Behavior Analysis, 2,* 239–246.

Wood, B. K., Umbreit, J., Liaupsin, C. J., & Gresham, F. M. (2007). A treatment integrity analysis of function-based intervention. *Education and Treatment of Children, 30,* 105–120.