Incorporating feasibility protocols in intervention research

Daniel L. Gadke [a,*], Thomas R. Kratochwill [b], Maribeth Gettinger [b]

[a] *Department of Counseling, Educational Psychology, and Foundations, Mississippi State University, United States*
[b] *Department of Educational Psychology, University of Wisconsin, Madison, United States*

ARTICLE INFO

ABSTRACT

In this paper, we review some dimensions of feasibility research. Feasibility research focuses on the intervention process and addresses questions about whether and how an intervention can be evaluated and implemented. Feasibility studies are implemented prior to conducting an outcome-focused pilot study or full-scale evaluation to test the effectiveness of an intervention. We propose a feasibility framework that includes 10 possible dimensions to evaluate in a feasibility trial, including (a) recruitment capability, (b) data collection procedures, (c) design procedures, (d) social validity, (e) practicality, (f) integration into existing systems, (g) adaptability, (h) implementation, (i) effectiveness, and (j) generalizability. Among these dimensions we offer some priorities that researchers can consider in establishing feasibility. Although feasibility investigations can advance evidence-based practice in psychology and education, we review current challenges for researchers to consider when incorporating a feasibility protocol into their intervention research agenda.

A focus on evidence-based practice (EBP; Kratochwill & Shernoff, 2004) as well as recent legislative requirements (Every Student Succeeds Act, 2015) have contributed to an increase in intervention research in school psychology. Intervention research is aimed at generating evidence of effective practice, primarily through demonstrating improved outcomes under controlled experimental conditions. A focus on real-world implementation of EBP to achieve similar outcomes, however, is often lacking (Sanetti & Collier-Meek, 2019). Increasingly, researchers are being challenged to address questions related not only to treatment effectiveness, but also the feasibility of evaluating and implementing interventions. Feasibility is a concept that encapsulates ideas about whether it is possible to do something. Feasibility studies, in particular, have been defined as "pieces of research done before a main study in order to answer the question 'Can this study be done?'" (National Institute for Health Research, 2012). Feasibility studies are designed to maximize the internal and external validity of intervention research by determining whether key elements (e.g., ability to recruit and randomize participants) are feasible before a controlled evaluation study is conducted. Despite the importance of feasibility studies for intervention development, the literature provides limited guidance for researchers looking to include a systematic evaluation of feasibility in their research programs. The purpose of this paper is to describe how to incorporate feasibility protocols into the process of intervention development and research in psychology and education.

1. Importance of Feasibility Research

Historically, intervention researchers in psychology and education have applied both single-case design (Kratochwill et al., 2010)

and group design methodologies (What Works Clearinghouse, 2017) to evaluate treatment outcomes. Multiple pilot studies are often completed prior to conducting a randomized controlled trial (RCT), with the aim of "readying" an intervention for application in practice settings (Becker, Park, Boustani, & Chorpita, 2019; McKenney, Page, Lakota, Niekra, & Thompson, 2019). Researchers often assume that interventions supported by high-quality research will be incorporated seamlessly into school practice leading to similar positive outcomes. In practice, however, there are often difficulties in translating demonstrable benefits of interventions into real-world contexts, such as schools and classrooms. One reason for this disconnect is that pilot studies and RCTs focus on intervention outcomes (Does this work?), not the intervention process (Can this work? How does this work?; Orsmond & Cohn, 2015). Feasibility research is a critical component of intervention development because it focuses explicitly on the implementation process. Feasibility studies consider whether intervention development and research can be accomplished to inform implementation and to support the design of larger studies, such as RCTs (Tickle-Degnen, 2013).

The importance of feasibility for intervention research in psychology and education mirrors the critical role that feasibility research currently plays in the medical literature, where feasibility studies are conducted early and often to inform drug trials and cancer treatments, and to investigate public health behaviors (Whitehead et al., 2014). In fact, feasibility work is so common in medical research that Biomed Central (BMC) recently launched a multi-disciplinary online and open-access journal, entitled *Pilot and Feasibility Studies*, dedicated exclusively to describing and reporting this work (Lancaster, 2015). Unfortunately, feasibility research is rarely done (or reported) in psychology or education. Moreover, there is a history in psychology and education of evaluating feasibility from an outcome perspective; that is, assessing treatment outcomes rather than testing the intervention procedures, outcome measures, and/or research techniques to be applied in a main study. Multiple reviews of published intervention research (including research in public health and mental health domains) have found that studies labeled as "feasibility trials" have an inappropriate focus on treatment outcomes rather than methodological issues. Arian et al. (2010), for example, reported that more than 80% of studies characterized as having a feasibility focus actually incorporated hypothesis testing and included tests of effectiveness as the primary emphasis. Similarly, Shanyinde, Pickering, and Weatherall (2011) found minimal, if any, coverage of feasibility issues or focus on feasibility objectives in more than 75% of studies labeled as "feasibility studies." Collectively, these findings suggest that many published feasibility studies emphasize treatment effectiveness, not feasibility, as the primary objective.

A focus on outcomes often leads to feasibility studies being conducted as small-scale models of the planned evaluation, asking "Does the intervention work with a small sample?" If the answer is "yes," then the research progresses to a full-scale study, such as a RCT. Unfortunately, this approach is problematic and costly. RCTs may be undermined by problems of acceptability, compliance, delivery of the intervention with integrity, and recruitment and retention that could have been addressed through a feasibility study (Bowen et al., 2009; Tickle-Degnen, 2013). When feasibility studies focus primarily on outcomes, researchers may not understand the process and mechanisms by which the intervention is working until *after* the project is completed. For example, data collection procedures may not be sensitive to change, or teachers may not implement an intervention because of time constraints. When these dimensions are not evaluated and addressed *prior to* conducting a large-scale evaluation, researchers are left to discover such issues along the way or after the fact, even with a treatment that "worked" in terms of outcomes.

2. Benefits of Feasibility Research

There are several benefits of incorporating feasibility studies into intervention research. Feasibility studies, for example, can allow exploration of the acceptability and feasibility of an intervention and evaluation design to inform decisions about progressing to a full-scale effectiveness or efficacy study. According to Eldridge et al. (2016), when researchers fail to conduct this vital preparatory work, there is a negative impact on the implementation and evaluation of interventions. The negative impact can be seen in evaluations that are undermined by problems related to the delivery of the intervention such as low compliance, integrity, or acceptability; issues with recruitment, retention, and randomization; inadequate measurement of key outcomes; and smaller-than-expected effect sizes – many of which could have been predicted and remediated through feasibility studies. Thus, the benefit of feasibility research lies in the potential contribution for both maximizing "real-world" implementation of EBP and enhancing the quality of experimental evaluation designs.

With respect to implementation of EBP, feasibility studies have the potential to close the well-documented research-to-practice gap. By focusing on the intervention process, feasibility studies explicitly target implementation (Orsmond & Cohn, 2015). Feasibility protocols afford researchers the opportunity to test their interventions in practice settings, consider real-world barriers and facilitators to implementation, and address issues of cultural or linguistic relevance, thereby creating a bridge between research and practice (Bowen et al., 2009). Newly-developed interventions should be tested prior to an effectiveness study to address inherent "uncertainties," such as optimal intervention content and mode of delivery; acceptability of the intervention for key stakeholders and end users; and capacity of teachers or interventionists to deliver the intervention within the structure, routine, available resources, and time constraints of their settings. Depending on the results of a feasibility study, further work may be necessary to refine and prepare the intervention for a large-scale evaluation and, ultimately, increase the likelihood of its implementation.

Feasibility studies also enable researchers to generate useful data to improve the quality of a full-scale evaluation study (Moore et al., 2018). Feasibility research allows investigators to explore a number of design issues such as recruitment capability, sample size, and retention; willingness of participants to be randomized to conditions; choice of outcome measures that are valid, reliable, and sensitive to change; and viability of a specified research design. Conducting a thorough review of these study dimensions allows researchers to eradicate issues and make adequate adjustments to the research design prior to conducting a costly and time-intensive evaluation study. Doing so increases the value of research by helping to avoid methodological design flaws and reducing the burden of research waste.

Although feasibility research does incur additional expenses in terms of time and resources for intervention researchers, the initial burden is offset by ultimately minimizing the potential waste of research investment that may occur when the feasibility of key elements (e.g., ability to recruit participants) is not assessed before conducting a full-scale trial. Furthermore, even when controlled evaluation studies do demonstrate positive outcomes, they may not generalize to real-world settings, rendering the benefits of an evidence-based intervention irrelevant.

3. Conceptualizing Feasibility within Intervention Development

Feasibility is a key component of the iterative process for intervention development (Diamond & Powell, 2011; Zucker et al., 2019; see Fig. 1). Feasibility studies focus on issues beyond efficacy potential to determine whether to proceed with controlled experimentation. In asking, "Can this be done," feasibility research takes into consideration multiple variables needed to conduct an eventual full-scale outcome study (National Institute for Health Research, 2012).

Feasibility research can be contextualized within an overall phased approach to intervention development and evaluation (see Fig. 1). In a phased approach, feasibility studies are implemented prior to full-scale evaluation to test the effectiveness of an intervention. As shown in Fig. 1, intervention development is not a fixed linear process. Answers to questions addressed in feasibility trials are used to either guide further development or refinement of interventions (returning to the Intervention Development phase) or contribute to the design of subsequent testing (proceeding to the Intervention Pilot Study phase). Feasibility studies clarify the critical elements of interventions; identify implementation needs in terms of infrastructure, professional development, or resources; and determine the practicality and usability of interventions prior to conducting pilot studies or RCTs (Kazdin, 2018). By answering meaningful intervention process questions prior to focusing on intervention outcomes, feasibility research may also help to minimize the growing concern regarding the limited use of EBPs and provide a clear foundation for future pilot and large-scale studies (Tickle-Degnen, 2013).

Feasibility research is often confused with pilot research (Arian, Campbell, Cooper, & Lancaster, 2010; Donald, 2018). Despite similarities, the two types of research have distinct characteristics. Feasibility studies are concerned with process, sometimes considering only one aspect of a large-scale outcome evaluation study. In contrast, pilot studies are best characterized as small-scale versions of what could become a large-scale study, retaining a focus on outcomes (effectiveness) and, most often, incorporating all aspects of a larger study (National Institute for Health Research, 2012; Thabane et al., 2010; Tickle-Degnen, 2013). When feasibility research indicates a project can be implemented, then a pilot study is conducted to determine if the intervention is effective on a small scale; if the pilot study yields promising results, then an RCT may be in order (see Fig. 1).
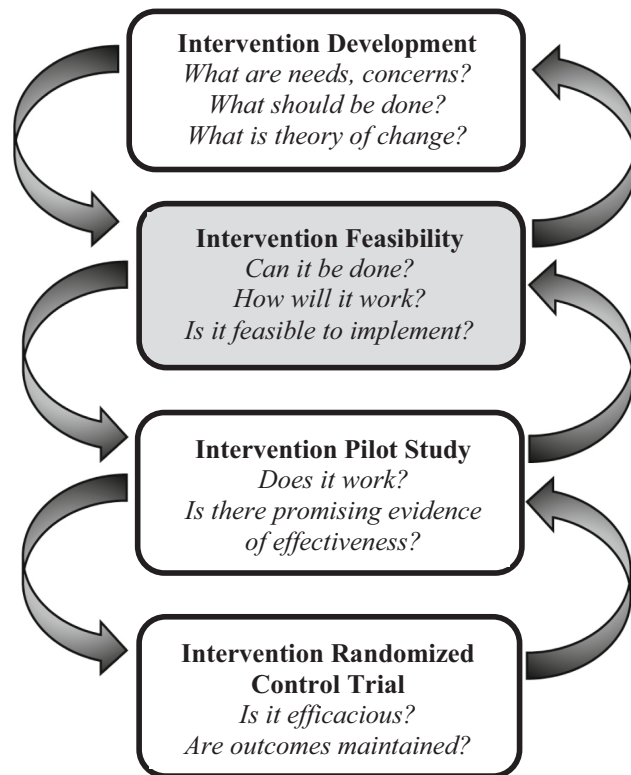

Fig. 1. Phased intervention development context for conducting feasibility research.

4. Feasibility Study Dimensions

Over the last decade, several researchers in medical and occupational health fields have contributed significantly to an understanding of the nature and design of feasibility studies, including measurement procedures, research design, and dimensions of feasibility to evaluate. For example, Bowen et al. (2009) delineated eight dimensions of feasibility for medical intervention research: acceptability, demand, implementation, practicality, adaptation, integration, expansion, and limited-efficacy testing. Acceptability considers how individuals participating in an intervention react to it. Demand relates to estimating how often an intervention is likely to be used. Implementation is how likely practitioners will implement the intervention as designed in an uncontrolled setting. Similar to implementation, practicality considers whether the intervention can be used given environmental constraints such as time, resource availability, and practitioner commitment. Adaptation focuses on the extent to which an intervention can be meaningfully modified to fit the needs of different situations (e.g., new delivery format or population). On a larger scale, integration also considers the degree to which systems-level change may be needed in the targeted intervention setting. As it implies, expansion is the extent to which an already successful intervention can be used effectively with a new population or in a different environment. Finally, limited-efficacy testing refers to evaluating treatment outcomes in limited ways, for example, with convenience samples or with weak statistical power. In recent years, intervention researchers from multiple disciplines have opted to focus on one or more of these dimensions, depending on the relevance of each dimension for their intervention research agenda.

Similar to Bowen et al., Tickle-Degnen (2013) described four aspects of feasibility to evaluate before designing a RCT in rehabilitation intervention research – specifically, process, resources, management, and scientific assessment. Process assessment focuses the number of available participants and the number of participants likely to remain in the project, adhering to its demands. Resource assessment considers factors such as physical space demands, technology requirements, timelines, motivation of participating
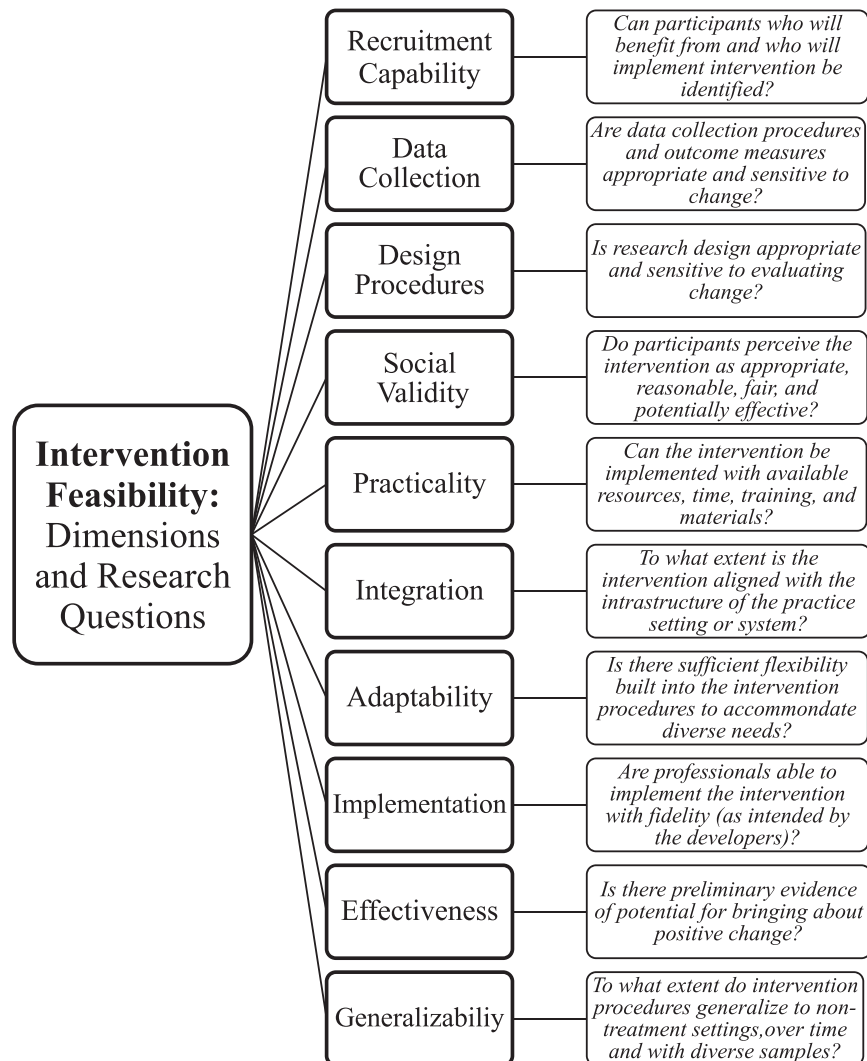

Fig. 2. Intervention feasibility dimensions and research questions.

institution members to carry out the work, and back-up plans. Management assessment focuses on the primary researchers and the research team, including skills to oversee the project, expertise and experience relative to the project goals, appropriate data management skills, and ethical adherence. Finally, scientific assessment considers procedural safety; frequency, intensity, and duration of procedures; reliability and validity of procedures; criteria for meaningful or significant change; and characteristics of populations that will benefit from the procedures.

Building on the work of Bowen et al. (2009) and Tickle-Degnen (2013), Orsmond and Cohn (2015) identified five objectives of feasibility work within the context of behavioral research, specifically to evaluate recruitment capacity, data collection procedures, intervention acceptability, availability of resources, and effectiveness potential. An evaluation of recruitment capacity gauges the likelihood of being able to recruit participants with targeted characteristics into the study. Whereas an evaluation of data collection procedures focuses on the appropriateness, sensitivity, and feasibility of data collection, intervention acceptability concerns the suitability and feasibility of the intervention per se. An evaluation of resources determines whether a research team has the resources available to complete a project. Finally, effectiveness potential considers promising evidence to support a large-scale evaluation of an intervention.

Taken together, the recommendations and guidelines for feasibility research proposed by Bowen et al. (2009), Tickle-Degnen (2013), and Orsmond and Cohn (2015) share the same conceptual underpinnings, each focusing on preliminary questions for researchers to address prior to conducting a large-scale outcome evaluation. Aside from Osmond and Cohn's focus on behavior interventions, however, there are few guidelines for conducting feasibility studies for psychological and educational interventions. Recently, Kazdin (2018) addressed feasibility within the context of dissemination of EBPs, emphasizing the need to consider physical characteristics of the intervention, infrastructure needs (e.g., technology, training needs, cost, time), and practicality of the intervention prior to dissemination. Guidance for feasibility work in the context of intervention development, however, remains largely absent from the literature.

5. Bringing feasibility work to intervention research in psychology and education

To address the gap in the literature on feasibility, we present a framework for conducting feasibility research that emphasizes not only efficacy potential but also intervention-specific issues such as recruitment capacity, practicality, and acceptability, among other variables. Building on the foundation for feasibility work related to medical and health-related interventions cited above, we propose a model with 10 dimensions of feasibility that are relevant for intervention development and evaluation (see Fig. 2). These dimensions include recruitment capability, data collection procedures, design procedures, social validity, practicality, integration into existing systems, adaptability, implementation, effectiveness, and generalizability.

The determination of 10 feasibility dimensions was guided by the work of multiple researchers who have contributed to the conceptualization of feasibility and delineation of critical dimensions of feasibility research (Bowen et al., 2009; Eldridge et al., 2016; Kazdin, 2018; Orsmond & Cohn, 2015; Tickle-Degnen, 2013). It is important to note that some feasibility dimensions in Fig. 2 are also assessed in intervention outcome research, sometimes with similar measurement procedures. An important distinction, however, lies in the interpretive context for measures used in intervention outcome versus feasibility research. For example, intervention research may include a measure of social validity (acceptability) to assess factors such as ease of intervention implementation, the likelihood that participants will use the intervention in the future, and perceptions of effectiveness and fairness. In this application context, social validity indices are used and interpreted as outcomes – specifically the outcome of participant satisfaction (Lochman et al., 2017). Although important to include in outcome-based research, a measure of social validity is also critical for feasibility research. When incorporated as part of a feasibility protocol, the purpose of assessing acceptability shifts from gauging post-implementation satisfaction to determining a priori whether stakeholders will consider using the intervention and, in turn, adapting the intervention as needed to promote implementation.

In the following sections, we describe each dimension in Fig. 2 from a feasibility perspective. We address the relevance of each dimension for feasibility vis-à-vis intervention development and offer considerations for assessing and evaluating the dimension.

6. Recruitment Capability

The first feasibility dimension relates to recruitment capability, or the extent to which researchers are capable of successfully recruiting study participants. Multiple factors influence recruitment capability. Take, for example, the decision to target a rural versus urban sample. According to the United States Census Bureau (2012), 81% of the population lives on only 3% of the land; the other 19% is spread across 97% of the country. Moreover, there are 249 universities with programs in school psychology (Gadke, Valley-Gray, & Rossen, 2019); however, some rural states have no programs whereas densely populated states have as many 35 (California). These types of population statistics are important to consider in a feasibility study when targeted samples with specific characteristics are needed for the investigation. Several questions relating to feasibility may arise if an intervention is designed to target a rural population. For example, how accessible is this population? How many schools are needed to recruit a sufficient number of participants? How much travel to rural sites is involved? How many other researchers are also interested in accessing the same rural sample of participants?

In a similar vein, recruitment may be impacted by targeting low versus high incidence concerns. Some researchers, for example, study social-emotional presentation of children with rare genetic disorders (e.g., CHARGE Syndrome) that may require a national recruitment net, whereas others study more readily available populations such as students with specific learning disabilities in reading. Orsmond and Cohn (2015) discussed their particular difficulty in recruiting individuals with autism spectrum disorder (ASD) due to

the number of different recruitment projects in the same region that also focused on ASD. Orsmond and Cohn further noted that the targeted goal of their intervention (sustaining existing friendships) had an unintentional impact on recruitment. Many children with ASD whose families expressed interest in their participation did not have existing friends and, as such, were not eligible to participate. Notably, recruitment capacity has a "domino effect" on other dimensions of a feasibility study, as each is influenced by the researcher's capacity to recruit participants.

Recruitment capability, for example, may have a direct impact on research design. As an illustration, one of the current authors was involved in intervention outcome research on selective mutism, a low incidence childhood disorder that occurs in less than 1% of the population (Bergman, 2013). Although the research design was conceptualized as a single-case concurrent multiple baseline design, the research team underestimated the amount of time necessary to recruit participants. As a consequence, the study was delayed for several months (which posed a problem for baseline participants), and it became necessary to implement a non-concurrent design. Moreover, because funding was linked to participant recruitment, the research team was at risk of losing their financial support. In this case, a feasibility investigation assessing recruitment capability would have led to modifying not only the recruitment process but also the methodology used to investigate the intervention outcome.

7. Data Collection Procedures

A second feasibility dimension relates to data collection procedures, including outcome measurement. To assess this feasibility dimension, researchers must ask relevant questions (e.g., Are the data collection procedures easily understood and implemented? How much data are needed to draw meaningful conclusions?; Tickle-Degnen, 2013). These questions guide researchers to consider the appropriateness of their data collection procedures and outcome measures for both interventionists and the target population (Orsmond & Cohn, 2015).

There are a number of issues to consider related to data collection, including the need for pre- and post-measures, whether to conduct direct observation, determining data sources (parents, teachers, students), identifying data collectors (researcher, research aides, teachers, parents, etc.), and developing appropriate outcome measures. Once decisions about data collection procedures are made, then additional questions arise: How much time is required to complete the measures? Are the instructions sufficient? Will the sample need help completing the measures? What are the psychometric properties of the measures for the sample? What is the response format? (Orsmond & Cohn, 2015).

These considerations may vary slightly when considering a large-group versus single-case design study. In a large-group study, for example, the outcome (dependent variable) may be adequately measured at one or two time points with a single measure; for a single-case design, however, data collection is daily or weekly over a week, month, or year. Moreover, in single-case methodology there may be unique feasibility considerations associated with repeated assessment of the dependent variables, such as reactivity or observer drift and fatigue.

8. Design Procedures

The feasibility of the research design or methodology is related to data collection procedures and outcome measures. After determining what procedures will be used to collect data, it is essential to consider the overall research design that is informed by clear research questions and specified independent variables. A helpful first step for researchers in thinking about design is to develop a logic model for the investigation (see Funnell & Rogers, 2011, for information on logic models and theory of change). The development of a logic model not only guides a feasibility study but also addresses the dimensions of a pilot or intervention outcome investigation.

Another consideration in design selection is the availability of established guidelines or protocols with criteria for how the study should be conducted. As an illustration, the What Works Clearinghouse (2017) has standards for intervention research designs, including RCTs and quasi-experimental designs, regression discontinuity designs, and single-case designs. Consider, for example, researchers who are exploring the feasibility of adopting a single-case design for their intervention evaluation study. To meet WWC design standards (*Standards 4.1*, 2020), they will need to include three replications of the intervention effect on the designated unit of analysis (single participant, classroom, school, etc.). As part of their feasibility work, researchers must determine the feasibility of their proposed single-case design, including recruitment, logistics of the timing of the intervention, or delay in treatment for participants on baseline. Some researchers may argue that adopting a single-case design can help remove the challenging control, no-treatment condition circumstance in conventional RCT investigations. Nevertheless, even in single-case designs there is typically a baseline phase requirement thereby necessitating that participants "wait" to receive the intervention (sometimes a circumstance exacerbated in multiple baseline designs). Feasibility considerations will vary across different classes of designs or design hybrids. These issues can be elucidated in feasibility research prior to conducting an outcome study.

9. Social Validity

Social validity refers to the social significance or relevance of intervention goals, the importance of intervention outcomes, and the acceptability of intervention procedures (Carter & Wheeler, 2019). Acceptability is the most traditional way of conceptualizing and, in turn, measuring social validity. Acceptability refers to the perceived appropriateness, fairness, reasonableness, and intrusiveness of an intervention for addressing a specific concern (Kazdin, 1981; Nastasi & Truscott, 2000; Reimers, Wacker, Cooper, & DeRaad, 1992). The success of intervention development depends, in part, on the acceptability of the research protocol and intervention practices for key stakeholders (Sekhon, Cartwright, & Francis, 2017). Acceptability assessment examines the extent to which an intervention meets

the needs of the interventionists, participants, and treatment setting (Carter, 2007). Although often part of a pilot study, a focus on acceptability during the feasibility phase maximizes the degree to which promising interventions are likely to be evaluated rigorously and, in turn, implemented in practice settings. In effect, acceptability functions as a "gatekeeper" for intervention implementation. Successful interventions hinge on the acceptability of the treatment to multiple stakeholders, including professionals who implement interventions, individuals who receive or participate in interventions, as well as the broader community or system within which the intervention is delivered (Mautone et al., 2009; Sterling-Turner & Watson, 2002).

Several conceptual models have contributed to the treatment acceptability literature in psychology and education (Lennox & Miltenberger, 1990; Reimers, Wacker, & Koeppl, 1987; Witt & Elliott, 1985). Most relevant to feasibility research is Lennox and Miltenberger's model, which underscores the importance of examining a range of factors related to acceptability to substantiate a decision to select one particular treatment. Taking into consideration multiple influences on acceptability, Carter (2008) proposed a distributed model of treatment acceptability. In this model, overall acceptability is distributed among the system, interventionists, and participants, thus necessitating an assessment of acceptability within each subdivision.

Carter (2008) distributive model provides a useful framework for examining treatment acceptability. By acknowledging different subdivisions of acceptability, researchers are able to make sense of contradictory acceptability data and render informed decisions that balance the varying perspectives of acceptability prior to designing a large-scale evaluation study. For example, a classroom teacher may endorse a behavior intervention that relies on punitive strategies (high teacher acceptability); however, the shared perception about the nature of the intervention within the larger school community may reveal low overall acceptability. Similarly, the same treatment could be viewed as acceptable by teachers but unacceptable by students. Moreover, a treatment may be highly acceptable among one cultural group but not another. The demographic characteristics of school cultures (urban versus rural, linguistic diversity, level of poverty, racial-ethnic composition) often influence the possible types of interventions that are viewed as acceptable, irrespective of effectiveness or acceptability by individual teachers.

Assessing acceptability is a critical aspect of feasibility research, yet it lacks an agreed-upon set of evidence-based methods. Treatment acceptability research is characterized by the use of rating scales to measure acceptability and subsequent analysis of variables (e.g., time demands, complexity of the intervention) associated with high versus low acceptability (Carter, 2007; Finn & Sladeczek, 2001). Two widely used traditional scales for measuring intervention acceptability are the *Treatment Evaluation Inventory* (Kazdin, 1980) and *Intervention Rating Profile* (Witt & Elliott, 1985); a more recent scale, *Usage Rating Profile–Intervention–Revised* (Briesch, Briesch, & Chafouleas, 2014), focuses specifically on acceptability and use of classroom management strategies.

Although relevant for research and evaluation purposes, the utility of rating scales for feasibility research is somewhat limited. Instead, Ayala and Elder (2011) recommended utilizing formative research methods such as focus groups and interviews to assess the acceptability of intervention materials and procedures in terms of cultural appropriateness, content, presentation, packaging, and delivery. Focus groups and interviews provide considerable opportunity for discussion between researchers and intended interventionists, intervention recipients, and school-level staff. Compared to rating scales, these methods allow researchers to probe further on topics as they arise during discussion, which results in a deeper understanding of reactions to the intervention that can impede or facilitate further research and implementation.

10. Practicality

Several authors have provided conceptual guidance on the practicality dimension of feasibility (Bowen et al., 2009; Kazdin, 2018; Orsmond & Cohn, 2015; Tickle-Degnen, 2013). Similar to the related dimensions of implementation and social validity, practicality considerations involve determining whether the intervention can be used given contextual and environmental constraints such as time, resource availability, and practitioner commitment (Bowen et al., 2009).

According to Tickle-Degnen (2013), a comprehensive evaluation of practicality calls for both resource assessment and management assessment, involving an a priori review of the institution or agency conducting the research. Whereas resource assessment is designed to consider materials and resources that affect practicality (e.g., physical space and technology requirements), management assessment focuses specifically on members of the research team, including their experience and expertise to conduct the study. Additional variables related to practicality include the scale of the intervention (e.g., single tier versus multi-tiered program), training resources (e.g., web-based versus traditional face-to-face training), and cost analysis (e.g., benefits of the intervention relative to implementation costs; Green, 2006; Kazdin, 2018).

A study conducted by one of the current authors to evaluate outcomes of a program entitled *Family and Schools Together* (FAST) illustrates the importance of the practicality dimension in feasibility research. FAST is a family engagement program designed to assist children and families to overcome mental health barriers and promote positive academic and social outcomes (McDonald & Howard, 1998). The site of the study was a large school district that originally agreed to be involved in the investigation with 60 schools participating. Unique characteristics of the school district (e.g., size, location) contributed to significant challenges and barriers to family participants (e.g., travel to the school, work schedules) and school personnel involvement (e.g., space for services, school security schedules). These unanticipated challenges created practical constraints to implementing the intervention and limited participation in the intervention program, thus undermining the integrity of the RCT. In this example, conducting a feasibility trial (with a specific focus on practicality) prior to implementation of the RCT would have identified some of the practical challenges in this context that were not anticipated at the outset. Moreover, this example illustrates the context-specificity of practicality and the importance of examining practicality within the targeted setting where the full-scale intervention trial is designated to occur.

11. Integration into Existing Systems

As illustrated in the above example, every context targeted for treatment implementation may be unique. Building on this feature, a related feasibility dimension to consider is integration. This dimension refers to the extent to which a proposed intervention is aligned with the unique features of the practice setting so as to facilitate integration into the current service delivery approach and, in turn, maximize implementation (Bowen et al., 2009; Carroll et al., 2007). Because of the documented link between intervention "fit" and implementation fidelity, integration is a critical focus of feasibility research (Harn, Parisi, & Stoolmiller, 2013).

In school settings, teachers may be asked by researchers to carry out new interventions while still implementing their own curriculum, adhering to their personal beliefs or standards, and enacting school-wide programs. Staff may not readily implement an intervention when there are overlapping or conflicting programs already in place, or when competing goals and incongruent philosophies about students' learning and development exist (Durlak & DuPre, 2008). Therefore, a goal of feasibility research is to gauge how well an intervention fits within the organizational structure, physical environment, or existing service delivery approach. If integrating a new intervention requires extensive change to the existing system, then further testing for effectiveness may not be warranted unless researchers are able to eliminate or modify components that are not well aligned with the implementation setting.

A focus on integration requires researchers to understand the interplay between implementation features of the intervention and structural characteristics of the practice setting (Durlak & DuPre, 2008). In terms of structural integration, feasibility assessment examines existing interventions and service delivery models to determine whether a new intervention can be integrated into the current structure. Consider the example in which researchers are interested in evaluating the benefits of an evidence-based supplemental (Tier 2) behavior support intervention. Sustained implementation of Tier 2 interventions relies, in part, on the extent to which a school has a multi-tiered system of academic and behavior supports (MTSS). Therefore, one aspect of integration to assess prior to a large-scale evaluation of the Tier 2 behavior intervention is the extent to which MTSS is already in place in participating schools, with behavior expectations aligned with those of the intervention to be evaluated (Gettinger, Kratochwill, Lindner, Eubanks, & Foy, 2019).

Although necessary, evidence of structural compatibility and integration may not be sufficient to conclude that an intervention is feasible for further development or outcome evaluation. Equally important is the fit between the culture of the school, classroom, and teachers and the orientation or underlying philosophy of the intervention itself (referred to as the "culture of the intervention"; Finnan, 2000). Researchers must concern themselves with the alignment or congruence with individual- and system-level beliefs or orientations (Donnell & Gettinger, 2015). The opinions and beliefs of individual teachers about new instructional interventions, in particular, have been shown to moderate fidelity levels (Haney, Lumpe, Czerniak, & Egan, 2002). When teachers' instructional philosophy matches the instructional approach of the intervention, they implement the intervention with higher fidelity than do teachers for whom a match does not exist.

The extent to which a new intervention is aligned with teachers' perceptions has not been studied as extensively as alignment with system infrastructure. This gap is due, in part, to the challenge associated with assessing and responding to a lack of belief congruence. Finnan (2000) recommended developing visual displays of compatibility versus incompatibility between the assumptions underlying an intervention and those of interventionists; this type of comparison promotes discussion and reflection prior to conducting efficacy studies and implementation. As interventions are developed, researchers need to evaluate the degree of fit between assumptions, intervention components, and delivery features with the school structure and culture. Prior to investing time and resources into evaluating the effectiveness of interventions, researchers should determine whether gaps between assumptions and structural features of the intervention and those of the practice setting are too wide, or whether there is sufficient alignment to facilitate integration (e.g., alignment between teachers' current approach to teaching reading and the instructional emphasis in a reading intervention to be evaluated).

12. Adaptability

Largely in line with the concepts of practicality and integration is the dimension of adaptability. Whereas practicality and integration reflect the capacity to deliver an intervention using existing resources within a system, adaptability relates to whether an intervention can be meaningfully modified to fit the needs of different situations, for example, using an alternative delivery format or with different populations (Lyon et al., 2019). When considering the adaptability of an intervention, it is essential to ask: Are the tools needed for implementation universal? Can the program be easily adjusted to fit across various school settings? Is it necessary to adhere to a rigid set of requirements to implement the intervention?

According to Bowen et al. (2009), an adaptable intervention should achieve comparable outcomes irrespective of the specific topography of its delivery format or the unique characteristics of the target population. This dimension can be tested relatively easily by comparing the outcomes of an intervention when it is implemented using a different format or with different populations. The *Good Behavior Game* (GBG; Barrish, Saunders, & Wolf, 1969) offers an example of a classroom intervention that has proven to be adaptable over the course of time. The GBG has demonstrated positive effects for elementary (Bowman-Perrot, Burke, Zaini, Zhang, & Vannest, 2016), middle school (Flower, McKenna, Bunuan, Muething, & Vega, 2014), and high school students (Stratton, Gadke, & Morton, 2018), suggesting it can be adapted to maintain fidelity across different setting needs. Moreover, a review of the literature reveals the GBG intervention has also been implemented with positive outcomes using a variety of delivery formats (Flower et al., 2014; Tingstrom, Sterling-Turner, & Wilczynski, 2006).

Another illustration of the importance of determining adaptability comes from a research project designed to evaluate *The Incredible Years* (Webster-Stratton & Hancock, 1998), an evidence-based program for improving parents' behavior management skills for their young children. The original research protocol required parents to travel to their community Head Start center to view a series

of videotapes on parent management skills. The research team, however, learned that most families did not have transportation options. Feasibility assessment determined, first, that virtually all families owned a TV and tape player, and, second, that intervention delivery could be easily adapted by providing families with their own set of tapes for home viewing.

13. Implementation

The level of implementation (also termed treatment integrity, or fidelity of implementation) is the extent to which interventionists (e.g., teachers, therapists, families) enact the procedures of an intervention as intended and as designed by the investigators (Century, Rudnick, & Freeman, 2010; O'Donnell, 2008; Sanetti and Kratochwill, 2014). A focus on implementation is aimed at identifying both variations in the degree to which an intervention is implemented with integrity, as well as the conditions under which implementation is likely to succeed. Within the context of a feasibility trial, researchers consider what it would take for an intervention to be, first, evaluated (in a pilot study, single-case design study, or RCT outcome study) and, subsequently, implemented and sustained in naturalistic settings (Bowen et al., 2009).

Historically, researchers have gauged treatment integrity by assessing implementation of the structural components of an intervention, typically through observation or self-report (Dusenbury, Brannigan, Falco, & Hansen, 2003; Schulte, Easton, & Parker, 2009). Adherence (number/percentage of intervention components implemented as designed by researchers) has been the primary structural dimension used by researchers to measure implementation (Gresham, Gansle, Noell, Cohen, & Rosenblum, 1993; Peterson & McConnell, 1996; Wickstrom et al., 1998). Although adherence remains a widely applied element of implementation, more recent research has shifted from a narrow focus on adherence to broader, multidimensional models of treatment integrity (Century et al., 2010; Lakin & Shannon, 2015; Sanetti, Fallon, & Collier-Meek, 2011; Schulte et al., 2009). The construct of implementation has been expanded to include additional structural dimensions such as exposure (amount or dosage of treatment in terms of number, frequency or length of intervention sessions), as well as process dimensions such as quality of delivery (skill, enthusiasm, understanding, and interpersonal strengths of the interventionist) and participant responsiveness (engagement, enjoyment, and attentiveness of the individuals who receive the intervention). Assessing adherence is straightforward, relying heavily on observable implementation of research procedures and intervention activities. An assessment of process dimensions, however, is more difficult; it often involves the use of subjective ratings to evaluate the quality of implementation and the nature of interactions between interventionists and participants. A focus on all aspects of implementation, both structural and process, enhances the quality of feasibility data to guide the design of subsequent pilot studies and implementation in naturalistic settings (Schulte et al., 2009).

Methods to collect implementation data range from self-reports and interviews to observations and analyses of participant artifacts (O'Donnell, 2008). Understandably, researchers tend to customize their implementation measures to align with the specific content and procedures of the intervention being studied. Measuring implementation as part of a feasibility study is not without certain limitations that can undermine the accuracy and validity of conclusions about implementation integrity. For example, self-report is subject to social desirability; interventionists may report higher or inaccurate levels of use, quality, understanding, or responsiveness to an intervention than what actually occurs (e.g., Gettinger et al., 2019). Observation techniques represent a more rigorous measurement of implementation. Some dimensions of implementation, however, are difficult to assess through observation. Observation procedures tap structural dimensions (e.g., adherence) but do not adequately capture relevant process dimensions (e.g., level of knowledge or understanding of the intervention). Furthermore, interventionists may demonstrate greater adherence or enthusiasm when they are being observed, thus inflating the level of implementation. Finally, ratings of process dimensions (e.g., participant engagement) can be highly subjective and variable depending on how raters perceive the dimension and its relevance for implementation.

To circumvent these measurement concerns, Ruiz-Primo (2006) recommended a multi-method and multi-source approach for studying implementation. Ruiz-Primo developed a scheme for categorizing implementation measurement procedures according to four dimensions along a high-to-low continuum – specifically, level of judgment required, directedness, sensitivity, and alignment to the intervention program. Ruiz-Primo also identified three main sources of information, including individuals who deliver an intervention, intervention participants, and independent data collectors. By design, an evaluation of intervention implementation should include multiple measures that involve differing levels of judgment (some objective and some subjective), directedness (some capturing implementation more directly than others), levels of sensitivity (some more sensitive than others), and levels of alignment (some more aligned to the intervention being studied). Moreover, measurement should include multiple information sources including interventionists (e.g., teachers), participants (e.g., students), and independent observers (e.g., trained observers or raters). Diverse methods and sources allow for the triangulation of data and provide a rich source of information about implementation issues relative to feasibility.

14. Effectiveness

It is the effectiveness dimension where feasibility research protocols and pilot and outcome studies begin to blend. Although the primary goal of feasibility research is to focus on the process, it is nevertheless prudent to consider intervention outcome effectiveness to some degree. Bowen et al. (2009) viewed effectiveness as part of feasibility assessment, considering what constitutes meaningful change and what populations are benefiting. Orsmond and Cohn (2015) also identified a preliminary evaluation of effectiveness as one objective of feasibility research (i.e., determining whether the intervention shows promising evidence of positive outcomes with the population for which it is intended). Sometimes this information can be assessed from prior research; however, when the objective is to test a new intervention or scale up an existing evidence-based program, then feasibility work is enlightening.

With the intent of understanding the potential effectiveness of an intervention, researchers can explore several elements of feasibility. First is whether the dependent variable(s) will be sensitive to change. Although this information might also be gleaned from prior research, there will likely be significant contextual variables that have a bearing on sensitivity (e.g., variation in participants, measures, or assessors). Second, feasibility issues often surround the method of collection of outcome measures. Testing the assessment protocol, method of training data collectors, or logistical issues in collecting data are just a few of the procedural variables to be evaluated. Finally, there are concerns related to the independent variable. Researchers will need to consider such intervention features

ABC Support feasibility study dimensions, questions, measurement, and decisions
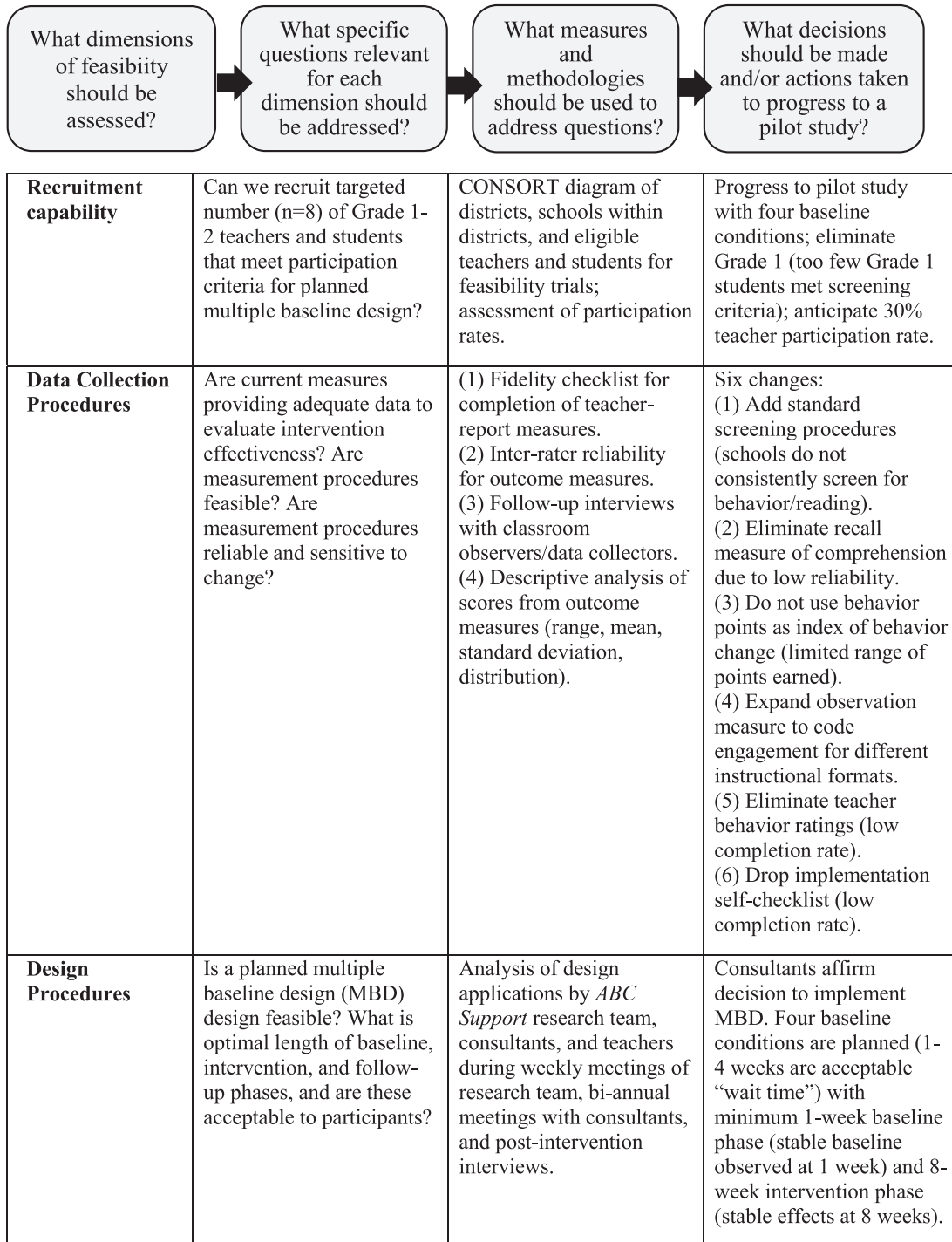
What dimensions of feasibiity should be assessed?	→	What specific questions relevant for each dimension should be addressed?	→	What measures and methodologies should be used to address questions?	→	What decisions should be made and/or actions taken to progress to a pilot study?

Recruitment capability	Can we recruit targeted number (n=8) of Grade 1-2 teachers and students that meet participation criteria for planned multiple baseline design?	CONSORT diagram of districts, schools within districts, and eligible teachers and students for feasibility trials; assessment of participation rates.	Progress to pilot study with four baseline conditions; eliminate Grade 1 (too few Grade 1 students met screening criteria); anticipate 30% teacher participation rate.
Data Collection Procedures	Are current measures providing adequate data to evaluate intervention effectiveness? Are measurement procedures feasible? Are measurement procedures reliable and sensitive to change?	(1) Fidelity checklist for completion of teacher-report measures. (2) Inter-rater reliability for outcome measures. (3) Follow-up interviews with classroom observers/data collectors. (4) Descriptive analysis of scores from outcome measures (range, mean, standard deviation, distribution).	Six changes: (1) Add standard screening procedures (schools do not consistently screen for behavior/reading). (2) Eliminate recall measure of comprehension due to low reliability. (3) Do not use behavior points as index of behavior change (limited range of points earned). (4) Expand observation measure to code engagement for different instructional formats. (5) Eliminate teacher behavior ratings (low completion rate). (6) Drop implementation self-checklist (low completion rate).
Design Procedures	Is a planned multiple baseline design (MBD) design feasible? What is optimal length of baseline, intervention, and follow-up phases, and are these acceptable to participants?	Analysis of design applications by *ABC Support* research team, consultants, and teachers during weekly meetings of research team, bi-annual meetings with consultants, and post-intervention interviews.	Consultants affirm decision to implement MBD. Four baseline conditions are planned (1-4 weeks are acceptable "wait time") with minimum 1-week baseline phase (stable baseline observed at 1 week) and 8-week intervention phase (stable effects at 8 weeks).

Fig. 3. ABC Support feasibility study dimensions, questions, measurement, and decisions.

Social Validity (Acceptability)	How do teachers and students rate the acceptability of *ABC Support*?	(1) Post-intervention teacher acceptability measure (IRP scale format with items specific to *ABC Support*); 3.5 criterion for acceptable. (2) Student acceptability measure; 4.5 for acceptable.	Mean teacher acceptability rating = 3.6, (above criterion for progressing to pilot study). Four item ratings were < 3.5 (criterion) and required modification (see **Practicality** dimension for discussion of items below criterion). Mean student acceptability rating = 4.7 (above criterion).
Practicality	Do schools have resources to implement *ABC Support*? Does format of materials support implementation? Do content and amount of training/coaching support implementation? Does research team have skills and resources to conduct pilot study?	(1) Post-intervention teacher interview to focus on items from social validity measure that fell below criterion for acceptability. (2) Research team meetings to address implementation issues and personnel/resource needs for pilot study.	Teacher interview responses led to recommended changes in: (1) organization and format of intervention materials (3.2 rating on acceptability measure). (2) cultural/linguistic relevance of intervention materials (3.1 rating). (3) duration of intervention (3.4 rating). (4) time required for implementation and completion of measures (3.4 rating).
Integration into Existing System	What is current level of MTSS implementation in schools, and how does *ABC Support* fit into MTSS framework? What is nature of core reading instruction and behavior support, and how are reading and behavior supports in *ABC Support* aligned?	(1) Teacher rating of schoolwide MTSS implementation; ≤ 85% of all elements "fully/mostly in place" is criterion for pilot study participants. (2) Classroom observation "snapshot" of Tier 1 reading and behavior support; ≤ 85% core instruction elements aligned with *ABC Support* elements are "mostly/somewhat present" (e.g., behavior expectations) is criterion. (3) Post-intervention teacher interview to assess alignment of *ABC Support* with classroom routines and MTSS framework.	Teachers rated 90% of MTSS elements for behavior and academic skills as "fully" or "mostly" in place; adequate level of MTSS implementation to proceed to pilot study. Observations revealed evidence-based core classroom instruction is "mostly" (75%) or "somewhat" (25%) present; criterion met for pilot study. Despite content alignment, scheduling of *ABC Support* implementation did not consistently fit classroom routines. Pilot study requires flexible scheduling of Tx sessions for teachers.

Fig. 3. (*continued*).

as training for intervention agents, dosage, duration, reactivity, methods of checking integrity, and whether booster sessions are needed for implementation integrity.

One way to assess potential outcome effectiveness in a feasibility trial is through the use of effect size measures. For example, a researcher might implement a non-experimental single-case AB design and calculate one or more of the many effect size measures

Implementation	Do teachers report they implement *ABC Support* in accordance with manualized procedures? Based on direct observations, do teachers implement *ABC Support* as intended?	(1) Teacher implementation self-checklist. (2) Fidelity of implementation observation checklist and global rating; $\leq$ 85% average session integrity is criterion.	Retain self-checklist in pilot study as implementation guide only (teachers did not complete). Observed implementation fidelity scores across students = 87.5%; criterion met for pilot study. Global rating of implementation fidelity = 3.8 (out of 4.0).Training and coaching is adequate to promote implementation fidelity.
Effectiveness	How do teachers and observers rate overall effectiveness of *ABC Support*? Do intervention data show promise of effectiveness?	(1) Implementation observation form ratings of student performance. (2) Post-intervention teacher acceptability measure (items that focus on perceived effectiveness) and interview. (3) Student performance data.	Three indicators show promise of effectiveness and support moving to pilot study evaluation: (1) High observer ratings for student success in reading (3.5/4.0) and behavior (3.3/4.0) and overall effectiveness (3.9/4.0). (2) High teacher ratings of effectiveness of *ABC Support* for behavior (3.7/4.0) and reading fluency (4.0/4.0). Interview responses indicate increases in reading confidence and engagement during core reading instruction. (3) Visual analysis of graphs of student performance data (non-overlapping data points) indicate positive change overall from baseline to intervention. No effect sizes were calculated.

Fig. 3. (*continued*).

available for this class of designs (Pustejovsky & Ferron, 2017; Shadish, Hedges, & Pustejovsky, 2014). These effect size measures can be compared to the benchmark targeted for outcome effects by the researcher and compared to prior effect size metrics reported in the literature. Although researchers may use effect sizes from a feasibility trial as an informal estimate of what to expect in a main trial, it is important to note that, due to small samples in most feasibility trials, such estimates may lack precision and could result in inappropriate decisions to proceed or not proceed to a main trial (Sim, 2019).

15. Generalizability

Although similar to adaptability, the dimension of generalizability relates to feasibility on a broader scope. At the heart of generalizability is the degree to which an intervention can maintain positive effects when transitioned from a highly controlled experimental setting to an educational or other real-world setting. When considering generalizability, researchers ask: Can similar intervention effects be achieved in settings that are different from the experimental setting, when scaled-up for implementation in large systems, or when implemented with diverse populations? Relatedly, generalization also calls for consideration of how well intervention effects are maintained across time.

Within a research context the notion of intervention generalizability brings to focus the need to consider internal versus external validity. When considering generalizability, an experimental setting with a high degree of internal validity offers rigorous control and maximizes confidence regarding causality of effects; however, it may minimize the external validity, or generalizability, of the intervention (Green & Glasgow, 2006). In feasibility research, it is essential for researchers to incorporate mechanisms to enhance generalizability; an intervention that does not hold up outside of the experimental setting has arguably limited use. The generalizability of an intervention has significant implications for expansion and scaling up of the intervention. A generalizable intervention is one that can be applied outside of the experimental setting with different populations; it should fit easily with various goal structures across settings, have primarily positive effects in multiple settings, and offer little disruption to the system in which it is implemented over time (Bowen et al., 2009).

An example of one approach to conceptualizing generalizability is a generalization map (Drabman, Hammer, & Rosenbaum, 1979). Building on the work of Drabman and colleagues, Allen, Tarnowski, Simonian, Elliott, and Drabman (1991) conceptualized a generalization map as an assessment framework to categorize the generalized effects of child and adolescent behavioral interventions. The map represents a useful framework for considering components of generalization that can be assessed in a feasibility trial or when the intervention is implemented.

16. Multidimensional Example of Applying Feasibility Dimensions to Intervention Development Research

In this section, we illustrate the application of the framework presented in Fig. 1 to assess the feasibility of a newly developed, school-based intervention called *Academic and Behavior Combined Support* (*ABC Support*; Gettinger et al., 2019). The example is "multidimensional" because the researchers considered eight feasibility dimensions in contrast to other examples where only a few dimensions were considered. *ABC Support* is an integrated, supplemental intervention for early elementary students that merges a combined focus on developing reading fluency skills and strengthening positive classroom behaviors (engagement and compliance with behavior expectations). The development and evaluation of *ABC Support* followed the phased approach depicted in Fig. 1 over a three-year project period. Partnerships with school-based personnel (administrators, classroom teachers, and school-wide interventionists) were central to the development and feasibility testing of *ABC Support*. As part of the iterative phased development process, the *ABC Support* research team conducted a series of feasibility trials (during Year 2) over a 9-month period during which six teachers (Grades 1 and 2) "tried out" the intervention with 18 students. The goal was to examine eight dimensions of feasibility of *ABC Support* through multiple quantitative and qualitative procedures with the aim of (a) deriving a final iteration of the intervention to be evaluated, and (b) designing a pilot outcome study to test the intervention using single-case design methodology.

Prior to initiating the feasibility trials, the *ABC Support* research team developed the logic model shown in Fig. 3 that involved: (a) consensus on the critical feasibility dimensions to assess prior to an outcome study, (b) determination of the relevant questions to address for each feasibility dimension, (c) development and/or selection of methods (formal and informal procedures) to obtain information relative to each question, and (d) specification of decisions and actions resulting from the feasibility assessment to enact prior to launching the outcome study (to be conducted during Year 3). Although a detailed description of the decision-making process is beyond the scope of this paper, the team relied heavily on two data sources during Year 1 of the project (initial development of *ABC Support*) to reach consensus on the dimensions, questions, measures, and actions summarized in Fig. 3. First, the team conducted a systematic review of outcome-oriented research with similar intervention procedures to describe and categorize the limitations/ challenges cited by authors in conducting their outcome studies. The challenges that occurred with high frequency across intervention studies (e.g., participant attrition, low fidelity of implementation, limited maintenance) helped to prioritize feasibility dimensions to assess and to frame the relevant feasibility questions. Second, the first iteration of *ABC Support* was developed during Year 1 in collaboration with potential end users (classroom teachers and interventionists) and school-based intervention consultants. These collaborators provided input on the specific practicality, implementation, and social validity issues that tend to undermine implementation of EBP in schools. The critical perspectives of end users also guided the team's prioritization of feasibility dimensions to assess.

As shown in Fig. 3, the feasibility questions were addressed through a variety of quantitative and qualitative measurement procedures, including teacher interviews, ratings, and self-report measures; direct observations; student performance on outcome measures and self-report; and feedback from consultants and end user partners. The resulting decisions and actions were determined based on information derived from multiple sources relative to pre-determined criteria for moving forward with the outcome study. The research team utilized a "traffic light" system proposed by Avery et al. (2017) to guide its decision-making about the outcome study. This system delineates three decision pathways relative to engaging in full-scale outcome research. First, feasibility findings may provide strong evidence for the utility, impact, sustainability, and long-term use of an intervention, thus supporting progression to a full-scale evaluation ("green light"). Second, a feasibility study may indicate the need to make meaningful changes prior to an evaluation study ("yellow light"). For example, outcome measures may lack sensitivity to change; key stakeholders may not find certain features of the intervention to be acceptable, thereby potentially impacting intervention integrity; or, the long-term sustainability of an intervention may be questionable due to time and resource constraints. Guided by feasibility findings, researchers fine-tune interventions and/or research designs before moving to an experimental study, thus maximizing both internal validity (scientific robustness) and external validity (generalizability to real-world settings) of large-scale studies. Finally, based on findings from a feasibility study, researchers may opt to not conduct future studies ("red light"). For example, a feasibility study may reveal that investigators will not be able to recruit the sample needed to detect meaningful change, or that the intervention is too complex to be practical in an educational setting.

Describing the application of Avery et al.'s (2017) system by the research team to every feasibility dimension in Fig. 3 is beyond the

scope of this example; however, an illustration of each type of decision is provided. First, the *ABC Support* research team found that teacher acceptability ratings (social validity) for the majority of intervention elements met the designated criterion (3.5/4.0) for inclusion in the final iteration of *ABC Support* to be evaluated ("green light"). Four elements, however, received acceptability ratings that fell below the criterion (< 3.5), calling for changes to be made prior to the outcome study ("yellow light") as summarized in Fig. 2. Finally, the inability of Grade 1 students to engage appropriately in the reading fluency tasks embedded in the intervention protocol signaled to the research team that it was not feasible to move forward with an outcome study for students below Grade 2 ("red light"). As shown in Fig. 3, the findings relative to nearly all dimensions of feasibility indicated that *ABC Support* showed promise and warranted pilot testing. The feasibility results, particularly information obtained through teacher interviews, provided the research team with direction for further development and modification of the intervention procedures (e.g., improved organization and design of intervention materials; greater clarity in manualized procedures), as well as revisions to the outcome measures. The researchers gleaned two additional lessons through the feasibility study. First, partnerships with teachers and school personnel were critical for developing an intervention with high potential for social validity and real-world application. Second, careful collection and analysis of data to address multiple elements of feasibility – not just efficacy potential (effectiveness) – were critical for proceeding with a pilot study. The information from the feasibility assessment helped the research team identify intervention elements that are vital for social validity and effectiveness as well as the modifications needed to increase fit and responsiveness. These "lessons learned" provide evidence of the value of incorporating feasibility research into intervention development and for considering several feasibility dimensions before proceeding to controlled efficacy testing.

17. Summary and Conclusions with Implications for Future Research and Policy

Feasibility research calls for a focus on the intervention process, addressing questions about whether and how an intervention can be evaluated and implemented. Whereas feasibility work is commonplace in medical research (Orsmond & Cohn, 2015; Tickle-Degnen, 2013; Whitehead et al., 2014), a strong emphasis on feasibility studies has yet to emerge in the fields of psychology and education. Despite an increase in well-controlled intervention studies, there remains a disconnect between intervention research and real-world implementation. The research-to-practice gap is due, in part, to the traditional focus in intervention research on experimental outcomes (Does this work?*)*, not the intervention process. The incorporation of feasibility work into psychology and education offers an opportunity to both improve intervention outcome research and bridge the implementation-science gap.

18. Framework for Selecting Feasibility Dimensions

There are some clear challenges associated with conducting feasibility research. Many intervention researchers have limited time and resources to focus on all dimensions of feasibility. Researchers may simply opt to gamble on whether it is worth the time and effort to gather information on feasibility or conduct a feasibility trial. Although feasibility researchers do not need to assess all 10 dimensions depicted in Fig. 2, it behooves them to consider the most relevant feasibility dimensions for their interventions. Given that the quality of data will likely vary across the feasibility dimensions that are evaluated, researchers will need to triangulate information to make a reasonable and ethical determination as to when to move forward with the next step in intervention development and testing. This process will inevitably be a balancing act (e.g., low recruitment capability balanced with high acceptability and potential effectiveness). A related challenge is that a feasibility trial could indicate the need to cycle back to the Intervention Development phase for further development and/or additional feasibility testing prior to moving to a pilot or full-scale experimental outcome study (see Fig. 1). This circumstance will likely add time to the intervention development process.

To minimize the time and resource burden associated with conducting feasibility studies, we offer a four-step framework (Fig. 4) designed to streamline the process of selecting and prioritizing feasibility dimensions to assess. At the outset researchers must delineate which feasibility dimensions to assess and how to obtain relevant information. A review of prior intervention research in the target area is a good starting point (see Fig. 4). We recommend constructing a checklist of feasibility dimensions addressed in prior studies.
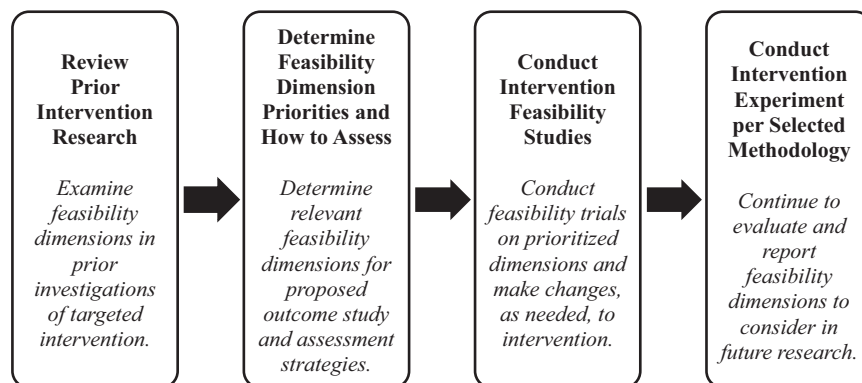

Fig. 4. Framework for selecting and conducting research on feasibility dimensions.

For example, a researcher may discover that a critical aspect of recruitment capability for a similar intervention is offering a strong incentive and providing transportation for participants. As another example, a researcher may learn that the practicality of implementing a web-based intervention protocol (e.g., telepractice or teleconsultation) hinges on schools having in place an electronic infrastructure to be able to participate.

Following a review of existing research, researchers must make some judgments about the priorities for moving forward (see Fig. 4). There is no pre-established ordering of importance of the 10 dimensions in Fig. 1, as this will be contingent on the specific circumstances of the research team and the context of the investigation. Nevertheless, we regard four dimensions – recruitment capability, data collection procedures, design procedures, and implementation – as fundamental to intervention outcome research. Increasingly, standards and guidelines for experimental research are being developed and disseminated; these will become a priority to address if research outcomes are to have scientific credibility (Appelbaum et al., 2018; What Works Clearinghouse, 2020). If these four dimensions can be assessed in a feasibility study, there is a higher chance of conducting a successful outcome study than if these dimensions are not assessed beforehand.

The next phase in Fig. 4 is likely to require the greatest investment of time and resources because it involves assessing the targeted dimensions of feasibility. Researchers may need to run several feasibility trials with the goal of maximizing the success of an eventual outcome study. As an example, consider the dimension of implementation. Although researchers may gain some insight into strategies that promote implementation integrity from a review of prior research, the selection of a new sample will likely necessitate additional or alternate integrity support beyond what traditionally has been provided (e.g., performance feedback). Having parents implement a behavior intervention protocol at home, for example, may require greater support (e.g., modeling, role-playing, coaching) than what teachers require to be able to implement the same intervention in their classrooms.

The final phase of the feasibility process in Fig. 4 is implementation of the outcome study, while still considering how feasibility dimensions play out under the conditions of the outcome experiment. In other words, assessing feasibility is an ongoing process and does not end with implementation of an experimental study. For example, the impact of feasibility dimensions may have been underestimated or under-investigated during a feasibility trial. Moreover, new and unexpected feasibility issues can surface during an outcome experiment. We recommend that intervention researchers make an effort to share their assessment of feasibility dimensions in research reports with the goal of assisting future researchers and helping to address implementation considerations.

19. Dissemination of Feasibility Investigations

There is often no formal requirement within the research community or from most funding agencies to conduct feasibility research. In the absence of such standards (formal or informal), conducting feasibility studies is without a strong incentive and may cast doubt on the credibility of gathering feasibility data. Although the decision to conduct feasibility assessment rests with individual researchers, there is a growing emphasis on evaluating at least some dimensions of feasibility as a component of funded research. Notably, the Institute for Education Sciences now requires feasibility work as part of intervention development to receive funding. The National Institute of Health has also begun to require feasibility trials. As this funding requirement continues to expand, feasibility work will become a more common requirement in intervention research.

Publication outlets for feasibility research projects are generally limited, at least in psychology and education journals. Some methods adopted by feasibility researchers may fall short of the experimental rigor and formal standards typically applied to outcome research (e.g., internal validity). We offer some dissemination suggestions to school psychology researchers and scientific journals in the field. First, we would advise research teams to clearly identify the purpose of their feasibility investigations and suggest how their findings can guide future intervention outcome research. A section of the research report, for example, could be devoted to presenting this information. Such a strategy will orient journal editors and reviewers examining the scientific report to a clear feasibility purpose that may deviate from the expectations they hold in reviewing conventional intervention-related outcome research. Second, researchers are encouraged to embed information in an intervention outcome report that outlines what dimensions of feasibility were evaluated and how decisions were made to move to experimentally testing the intervention.

We also offer some suggestions for scholarly journals in the field. The hope is that our focus in this paper on incorporating feasibility into intervention research protocols will increase awareness of the importance of this type of research. Methodological developments in other aspects of intervention research have led to acceptance and accommodations for non-conventional reporting. For example, with the replication crisis noted across the field of psychology (see Baker, 2015; Smith & Little, 2018), many journals have begun to create dedicated space for direct and systematic replication research. As noted earlier, the medical literature has been successful with creating a single journal devoted exclusively to pilot and feasibility work. Hopefully, a similar pattern of journal editor responsiveness will emerge for feasibility research in psychology and education, thereby creating an increased demand for the work.

20. Final Perspectives

Despite the challenges associated with conducting feasibility research, there is a compelling need for feasibility work in intervention outcome research to address the persistent disconnect between research and practice. The call for EBPs in psychology and education has increased significantly in recent years. Nonetheless, only 11% of nationally certified school psychologists report implementing EBPs (Hicks et al., 2015) even though 75% of school psychology programs have specific coursework related to EBPs (Reddy, Forman, Stoiber, & Gonzalez, 2017). Researchers are encouraged to incorporate into their research protocols a multi-dimensional feasibility framework (Fig. 2), consider the priorities on dimensions of feasibility, evaluate them in feasibility trials, and, in turn, address critical issues concerning the usability of interventions. Doing so as preparatory work prior to large-scale

evaluation and dissemination will likely contribute to bridging the science-practice gap and maximizing implementation of EBPs.

Declaration of Competing Interest

None.

Acknowledgements

References

Eldridge, S. M., Lancaster, G. A., Campbell, M. J., Thabane, L., Hopewell, S., Coleman, C. L., & Bond, C. M. (2016). Defining feasibility and pilot studies in preparation for randomized controlled trials: Development of a conceptual framework. *PLoS ONE, 11*(3). https://doi.org/10.1371/journal.pone.0150205.

Every Student Succeeds Act. (2015). *20 U.S.C. §§ 6301 et seq*.

Allen, J. S., Tarnowski, K. J., Simonian, S. J., Elliott, D., & Drabman, R. S. (1991). The generalization map revisited: Assessment of generalized treatment effects in child and adolescent behavior therapy. *Behavior Therapy, 22*, 393–405. https://doi.org/10.1016/S0005-7894(05)80373-9.

Appelbaum, M., Klien, R. B., Nezu, A. M., Cooper, H., Mayo-Wilson, E., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA publications and communications board task force report. *American Psychologist, 73*(1), 3–25. https://doi.org/10.1037/amp0000191.

Arian, M., Campbell, M. J., Cooper, C. L., & Lancaster, G. A. (2010). What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Medical Research Methodology, 10*(67), 1–7. https://doi.org/10.1186/1471-2288-10-67.

Avery, K. N. L., Williamson, P. R., Gamble, C., O'Connell Francischetto, E., Metcalfe, C., Davidson, P., … Blazeby, J. M. (2017). Informing efficient randomized controlled trials: Exploration of challenges in developing progression criteria for internal pilot studies. *BMJ Open, 7*, Article e013537. https://doi.org/10.1136/bmjopen-2016-013537.

Ayala, G. X., & Elder, J. P. (2011). Qualitative methods to ensure acceptability of behavioral and social interventions to the target populations. *Journal of Public Health Dentistry, 71*, S69–S79. https://doi.org/10.1111/j.1752-7325.2011.00241.x.

Baker, M. (2015). First results from psychology's largest reproducibility test. *Nature.*. https://doi.org/10.1038/nature.2015.17433.

Barrish, H., Saunders, M., & Wolf, M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis, 2*, 119–124. https://doi.org/10.1901/jaba.1969.2-119.

Becker, K. D., Park, A. L., Boustani, M. M., & Chorpita, B. F. (2019). A pilot study to examine the feasibility and acceptability of a coordinated intervention design to address treatment engagement challenges in school mental health services. *Journal of School Psychology, 76*, 78–88. https://doi.org/10.1016/j.jsp.2019.07.013.

Bergman, R. L. (2013). *Treatment for children with selective mutism: An integrative behavioral approach*. Oxford Press.

Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., Weiner, D., … Fernandez, M. (2009). How we design feasibility studies. *American Journal of Prevention Medicine, 36*, 452–457. https://doi.org/10.1016/j.amepre.2009.02.002.

Bowman-Perrot, I., Burke, M. D., Zaini, S., Zhang, N., & Vannest, K. (2016). Promoting positive behavior using the good behavior game: A meta-analysis of single-case research. *Journal of Positive Behavior Interventions, 18*, 180–190. https://doi.org/10.1177/1098300715592355.

Briesch, A. M., Briesch, J. M., & Chafouleas, S. M. (2014). Investigating the usability of classroom management strategies among elementary school teachers. *Journal of Positive Behavior Interventions, 17*, 5–14. https://doi.org/10.1177/1098300714531827.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*, 1–9. https://doi.org/10.1186/1748-5908-2-40.

Carter, S. L. (2007). Review of recent treatment acceptability research. *Education and Training in Developmental Disabilities, 42*, 301–316.

Carter, S. L. (2008). A distributive model of treatment acceptability. *Education and Training in Developmental Disabilities, 43*, 411–420.

Carter, S. L., & Wheeler, J. J. (2019). *The social validity manual: Subjective evaluations of interventions (2nd ed.)*. Elsevier.

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation, 31*, 199–218. https://doi.org/10.1177/1098214010366173.

Diamond, K. E., & Powell, D. R. (2011). An iterative approach to the development of a professional development intervention for head start teachers. *Journal of Early Intervention, 33*, 75–93. https://doi.org/10.1177/1053815111400416.

Donald, G. (2018). A brief summary of pilot and feasibility studies: Exploring terminology, aims, and methods. *European Journal of Integrative Medicine, 24*, 65–70. https://doi.org/10.1016/j.eujim.2018.10.017.

Donnell, L. A., & Gettinger, M. (2015). Elementary school teachers' acceptability of school reform: Contribution of belief congruence, self-efficacy, and professional development. *Teaching and Teacher Education, 51*, 47–57. https://doi.org/10.1016/j.tate.2015.06.003.

Drabman, R. S., Hammer, D., & Rosenbaum, M. S. (1979). Assessing the generalization in behavior modification with children: The generalization map. *Behavioral Assessment, 1*, 203–219.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350. https://doi.org/10.1007/s10464-008-9165-0.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research: Theory and Practice, 18*, 237–256. https://doi.org/10.1093/her/18.2.237.

Finn, C. A., & Sladeczek, I. W. (2001). Assessing the social validity of behavioral interventions: A review of treatment acceptability measures. *School Psychology Quarterly, 16*, 176–206. https://doi.org/10.1521/scpq.16.2.176.18703.

Finnan, C. (2000, April). *Implementing school reform models: why is it so hard for some schools and easy for others?* Paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA https://files.eric.ed.gov/fulltext/ED446356.pdf.

Flower, A., McKenna, J. W., Bunuan, R. L., Muething, C. S., & Vega, R. J. (2014). Effects of the good behavior game on challenging behaviors in the school settings. *Review of Educational Research, 84*, 546–571. https://doi.org/10.3102/0034654314536781.

Funnell, S., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. Jossey-Bass.

Gadke, D. L., Valley-Gray, S., & Rossen, E. (2019). NASP report of graduate education in school psychology: 2017-2018. *NASP Research Reports, 4*(2), 1–12.

Gettinger, M., Kratochwill, T., Lindner, A., Eubanks, A., & Foy, A. (2019). ABC support: Results and analysis of a feasibility trial. In *Unpublished manuscript*. Wisconsin Center for Education Research, University of Wisconsin, Madison, WI.

Green, L. W. (2006). Public health asks of systems science: To advance our evidence-based practice, can you help us get more practice-based evidence? *American Journal of Public Health, 96*, 406–409. https://doi.org/10.2105/AJPH.2005.066035.

Green, L. W., & Glasgow, R. E. (2006). Evaluating the relevance, generalization, and applicability of research: Issues in external validation and translation methodology. *Evaluation & the Health Professions, 29*, 126–153. https://doi.org/10.1177/0163278705284445.

Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychology Review, 22*, 254–272.

Haney, J. J., Lumpe, A. T., Czerniak, C. M., & Egan, V. (2002). From beliefs to actions: The beliefs and actions of teachers implementing change. *Journal of Science Teacher Education, 13*, 171–187. https://doi.org/10.1023/A:1016565016116.

Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools? *Exceptional Children, 79*, 181–193. https://doi.org/10.1177/001440291307900204.

Hicks, T. B., Shahidullah, J. D., Carlson, J. S., & Palejwala, M. H. (2015). Nationally certified school psychologists' use and reported barriers to using evidence-based interventions in schools: The influence of graduate program training and education. *School Psychology Quarterly, 29*, 469–487. https://doi.org/10.1037/spq0000059.

Kazdin, A. E. (1980). Acceptability of alternative treatments for deviant child behavior. *Journal of Applied Behavior, 13*, 259–273. https://doi.org/10.1901/jaba.1980.13-259.

Kazdin, A. E. (1981). Acceptability of child treatment techniques: The influence of treatment efficacy and adverse side effects. *Behavior Therapy, 12*, 493–506. https://doi.org/10.1016/S0005-7894(81)80087-1.

Kazdin, A. E. (2018). *Innovations in psychosocial interventions and their delivery: Leveraging cutting-edge science to improve the world's mental health*. Oxford.. https://doi.org/10.1093/med-psych/9780190463281.001.0001

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_scd.pdf.

Kratochwill, T. R., & Shernoff, E. S. (2004). Evidence-based practice: Promoting evidence-based interventions in school psychology. *School Psychology Review, 33*, 34–48. https://doi.org/10.1521/scpq.18.4.389.27000.

Lakin, J. M., & Shannon, D. M. (2015). The role of treatment acceptability, effectiveness, and understanding in treatment fidelity: Predicting implementation variation in a middle school science program. *Studies in Educational Evaluation, 47*, 28–37. https://doi.org/10.1016/j.stueduc.2015.06.002.

Lancaster, G. A. (2015). Pilot and feasibility studies come of age! *Pilot and Feasibility Studies, 1*(1). https://doi.org/10.1186/2055-5784-1-1.

Lennox, D. B., & Miltenberger, R. G. (1990). On the conceptualization of treatment acceptability. *Education and Training in Mental Retardation, 25*, 211–224.

Lochman, J. E., Boxmeyer, C. L., Jones, S., Qu, L., Ewoldsen, D., & Nelson, W. M. (2017). Testing the feasibility of a briefer school-based preventive intervention with aggressive children: A hybrid intervention with face-to-face and internet components. *Journal of School Psychology, 62*, 33–50. https://doi.org/10.1016/j.jsp.2017.03.010.

Lyon, A. R., Cook, C. R., Locke, J., Davis, C., Powell, B. J., & Waltz, T. J. (2019). Importance and feasibility of an adapted set of implementation strategies in schools. *Journal of School Psychology, 76*, 66–77. https://doi.org/10.1016/j.jsp.2019.07.014.

Mautone, J. A., DuPaul, G. J., Jitendra, A. K., Tresco, K. E., Junod, R. V., & Volpe, R. J. (2009). The relationship between treatment integrity and acceptability of reading interventions for children with attention-deficit/hyperactivity disorder. *Psychology in the Schools, 46*, 919–931. https://doi.org/10.1002/pits.20434.

McDonald, L., & Howard, D. (1998). *Families and schools together*. U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

McKenney, E. L. W., Page, V., Lakota, J., Niekra, N., & Thompson, S. J. (2019). Supporting integrity of discrete trial teaching via tiered consultation: A pilot study. *Journal of Applied School Psychology, 35*, 52–74. https://doi.org/10.1080/15377903.2018.1493555.

Moore, L., Hallingberg, B., Wight, D., Turley, R., Segrott, J., Craig, P., … Moore, G. (2018). Exploratory studies to inform full-scale evaluations of complex public health interventions: The need for guidance. *Journal of Epidemiology and Community Health, 72*, 865–866. https://doi.org/10.1136/hech-2017-210414.

Nastasi, B. K., & Truscott, S. D. (2000). Acceptability research in school psychology: Current trends and future directions. *School Psychology Quarterly, 15*, 117–122. https://doi.org/10.1037/h0088781.

National Institute for Health Research. (2012). *NIHR Evaluation*. Trials and Studies Coordination Centre*: Glossary*. www.netscc.ac.uk/glossary/#glos6/.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum research. *Review of Educational Research, 78*, 33–84. https://doi.org/10.3102/0034654307313793.

Orsmond, G. I., & Cohn, E. S. (2015). The distinctive features of a feasibility study: Objectives and guiding questions. *OTJR: Occupation, Participation, and Health, 35*, 169–177. https://doi.org/10.1177/1539449215578649.

Peterson, C. A., & McConnell, S. R. (1996). Factors related to intervention integrity and child outcome in social skills interventions. *Journal of Early Intervention, 20*, 146–164. https://doi.org/10.1177/105381519602000206.

Pustejovsky, J. E., & Ferron, J. M. (2017). Research synthesis and meta-analysis of single-case designs. In J. M. Kauffman, D. P. Hallahan, & P. C. Pullen (Eds.), *Handbook of special education (2nd ed., pp. 168–186)*. Routledge.

Reddy, L. A., Forman, S. G., Stoiber, K. C., & Gonzalez, J. E. (2017). A national investigation of school psychology trainers' attitudes and beliefs about evidence-based practices. *Psychology in the Schools, 54*, 261–278. https://doi.org/10.1002/pits.21999.

Reimers, T. M., Wacker, D. P., Cooper, L. J., & DeRaad, A. O. (1992). Acceptability of behavioral treatments for children. *School Psychology Review, 21*, 628–643.

Reimers, T. M., Wacker, D. P., & Koeppl, G. (1987). Acceptability of behavioral treatments: A review of the literature. *School Psychology Review, 16*, 212–227. https://doi.org/10.1037/0735-7028.19.1.68.

Ruiz-Primo, M. A. (2006). *A multi-method and multi-source approach for studying fidelity of implementation (CSE Report 677)*. Stanford University Center for research on evaluation standards, and student testing.

Sanetti, L. M. H., & Collier-Meek, M. A. (2019). Increasing implementation science literacy to address the research-to-practice gap in school psychology. *Journal of School Psychology, 76*, 33–47. https://doi.org/10.1016/j.jsp.2019.07.008.

Sanetti, L. M. H., Fallon, L. M., & Collier-Meek, M. A. (2011). Treatment integrity assessment and intervention by school-based personnel: Practical applications based on a preliminary study. *School Psychology Forum: Research in Practice, 5*(3), 87–102.

Sanetti, L. M., & Kratochwill, T. R. (2014). *Treatment integrity: A foundation for evidence-based practice in applied psychology*. American Psychological Association.

Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review, 38*, 460–475.

Sekhon, M., Cartwright, M., & Francis, J. J. (2017). Acceptability of healthcare interventions: An overview of reviews and development of a theoretical framework. *BMC Health Services Research, 17*(88), 1–13. https://doi.org/10.1186/s12913-017-2031-8.

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

Shanyinde, M., Pickering, R. M., & Weatherall, M. (2011). Questions asked and answered in pilot and feasibility randomized controlled trials. *BMC Medical Research Methodology, 11*(117), 1–11. https://doi.org/10.1186/1471-2288-11-117.

Sim, J. (2019). Should treatment effects be estimated in pilot and feasibility studies? *Pilot and Feasibility Studies, 5*(105). https://doi.org/10.1186/s40814-019-0493-7.

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin and Review, 25*, 2083–2101. https://doi.org/10.3758/s13423-018-1451-8.

Sterling-Turner, H. E., & Watson, T. S. (2002). An analog investigation of the relationship between treatment acceptability and treatment integrity. *Journal of Behavioral Education, 11*, 39–50. https://doi.org/10.1023/A:1014333305011.

Stratton, K. K., Gadke, D. L., & Morton, R. M. (2018). Using the good behavior game with high school special education students: Comparing student and teacher-select reinforcers. *Journal of Applied School Psychology, 35*, 105–121. https://doi.org/10.1080/15377903.2018.1509920.

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., … Goldsmith, C. H. (2010). A tutorial on pilot studies: The what, why, and how. *BMC Medical Research Methodology, 10*(1), 1–10. https://doi.org/10.1186/1471-2288-10-1.

Tickle-Degnen, L. (2013). Nuts and bolts of conducting feasibility studies. *American Journal of Occupational Therapy, 67*, 171–176. https://doi.org/10.5014/ajot.2013.006270.

Tingstrom, D. H., Sterling-Turner, H. E., & Wilczynski, S. M. (2006). The good behavior game: 1969-2002. *Behavior Modification, 20*, 225–253. https://doi.org/10.1177/0145445503261165.

United States Census Bureau. (2012). Growth in urban population outpaces rest of nation. https://www.census.gov/newsroom/releases/archives/2010_census/cb12-50.html.

Webster-Stratton, C., & Hancock, L. (1998). Training for parents of young children with conduct problems: Content, methods and therapeutic processes. In J. M. Briesmeister, & C. E. Schaefer (Eds.), *Handbook of parent training (2nd ed., pp. 98–152)*. John Wiley & Sons.

What Works Clearinghouse. (2017). *What Works Clearinghouse: Standards handbook* (Version 4.0). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf.

What Works Clearinghouse. (2020). What Works Clearinghouse: Standards handbook (Version 4.1). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf.

Whitehead, A. L., Sully, B. G. O., & Campbell, M. J. (2014). Pilot and feasibility studies: Is there a difference from each other and from a randomized controlled trial? *Contemporary Clinical Trials, 38*, 130–133. https://doi.org/10.1016/j.cct.2014.04.001.

Wickstrom, K., Jones, K., LaFleur, L., & Witt, J. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 13*, 141–154. https://doi.org/10.1037/h0088978.

Witt, J., & Elliott, S. N. (1985). Acceptability of classroom intervention strategies. In T. Kratochwill (Ed.), *Advances in school psychology (Vol. 4, pp. 251–288)*. Erlbaum.

Zucker, T. A., Carlo, M. S., Landry, S. H., Masood-Saleem, S. S., Williams, J. M., & Bhavsar, V. (2019). Iterative design and pilot testing of the Developing Talkers tiered academic language curriculum for pre-kindergarten and kindergarten. *Journal of Research on Educational Effectiveness, 12*, 274–306. https://doi.org/10.1080/19345747.2018.1519623.