



Measuring Implementation of Evidence-Based Programs Targeting Young Children at Risk for Emotional/Behavioral Disorders

Conceptual Issues and Recommendations

Kevin S. Sutherland

Bryce D. McLeod

Virginia Commonwealth University, Richmond, VA, USA

Maureen A. Conroy

University of Florida, Gainesville, FL, USA

Julia R. Cox

Virginia Commonwealth University, Richmond, VA, USA

Young children with and at risk for emotional/behavioral disorders (EBD) present challenges for early childhood teachers. Evidence-based programs designed to address these young children's behavior problems exist, but there are a number of barriers to implementing these programs in early childhood settings. Advancing the science of treatment integrity measurement can assist researchers and consumers interested in implementing evidence-based programs in early childhood classrooms. To provide guidance for researchers interested in assessing the integrity of implementation efforts, we describe a conceptual model of implementation of evidence-based programs designed to prevent EBD when applied in early childhood settings. Next, we describe steps that can be used to develop treatment integrity measures. Last, we discuss factors to consider when developing treatment integrity measures with specific emphasis on psychometrically strong measures that have maximum utility for implementation research in early childhood classrooms.

Keywords: *prevention, treatment implementation, treatment integrity*

Authors' Note: Kevin S. Sutherland, Department of Special Education and Disability Policy, Virginia Commonwealth University; Bryce D. McLeod, Department of Psychology, Virginia Commonwealth University; Maureen A. Conroy, Department of Special Education, School Psychology, and Early Childhood Studies, University of Florida; Julia R. Cox, Department of Psychology, Virginia Commonwealth University. This research was supported by grants from the U.S. Department of Education, Institute of Education Sciences (R324A080074) and the Presidential Research Initiation Program at Virginia Commonwealth University. The opinions expressed by the authors are not necessarily reflective of the position of or endorsed by the U.S. Department of Education or Virginia Commonwealth University. Correspondence concerning this article should be addressed to Kevin S. Sutherland, Department of Special Education and Disability Policy, Virginia Commonwealth University, Richmond, VA, 23284. Email: kssuther@vcu.edu

Nationally, there is an increasing trend for states to offer early educational programs aimed at preschool-age children, especially for children who are at risk for school failure. For example, in 2011, the U.S. Departments of Education and Health and Human Services created the Race to the Top—Early Learning Challenge state grants to assure that young children (particularly those at risk) have access to high-quality early childhood programs. Many young children who attend these federal and state funded programs display high levels of problem behaviors that interfere with their own learning, affect their classmates, and affect interactions with their teachers (Driscoll & Pianta, 2010; Quesenberry, Hemmeter, & Ostrosky, 2011). The complex array of risk factors to which these young children are exposed increases their risk for the development of emotional/behavioral disorders (EBD) associated with poor developmental outcomes (Berlin, Brooks-Gunn, McCarton, & McCormick, 1998; Nelson, Stage, Duppong-Hurley, Synhorst, & Epstein, 2007).

Several evidence-based programs (EBPs) exist that target young children at risk for EBD. They are typically comprised of multiple components and intervention strategies; for example, *PK—Promoting Alternative Thinking Strategies (PK PATHS)* (Domitrovich, Cortes, & Greenberg, 2007) and *Incredible Years* (Webster-Stratton, Reid, & Hammond, 2004). Although there is evidence supporting the efficacy of these programs when delivered in controlled settings, there are fewer studies indicating their effectiveness when implemented by teachers in authentic early childhood settings (Domitrovich, Gest, Jones, Gill, & DeRousie, 2010). Translating EBPs into typical early childhood settings can be difficult for researchers and educators. Indeed, early childhood educators often struggle to implement EBPs designed to prevent and ameliorate chronic problem behaviors, while researchers struggle to identify the variables that can facilitate implementation (Domitrovich et al., 2010; Durlak, 2010).

Historically, the approach toward translating EBPs designed to prevent and ameliorate problem behaviors of children at risk for EBD has been to “train” teachers in a large group didactic format and “hope” that they return to their classrooms and implement the program with integrity (i.e., skillfully deliver the procedures prescribed by the EBP), and this approach to translation of research into practice has often failed in obtaining sustained implementation by classroom teachers (Becker & Domitrovich, 2011; Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005; Joyce & Showers, 2002). Failure to implement a program effectively may be due, in part, to the differences in contextual variables (e.g., child, teacher, and setting characteristics) in research and everyday early childhood settings that might influence the integrity of implementation (i.e., extent to which the EBP was delivered as designed; McLeod, Southam-Gerow, Tully, Rodriguez, & Smith, 2013). By measuring the contextual variables that influence implementation integrity, it may be possible to better understand how to effectively transport programs from research settings to authentic settings. However, achieving this translational goal requires the appropriate methods and tools to evaluate implementation integrity in research and practice.

We propose that treatment integrity frameworks developed in the treatment technology field (Carroll & Nuro, 2002; McLeod, Southam-Gerow, et al., 2013) can be applied to assess the implementation integrity of EBPs in classrooms for young children with EBD. Treatment integrity refers to the degree to which an EBP was delivered as intended and is composed of four dimensions: treatment adherence, treatment differentiation, competence, and relational factors (McLeod, Southam-Gerow, & Weisz, 2009). *Treatment adherence*

refers to the extent to which the teacher delivers the program as designed (i.e., components prescribed by the EBP). *Treatment differentiation* refers to the extent to which interventions under study differ along appropriate lines defined by the program's protocol (e.g., presence of components proscribed by the EBP). *Competence* refers to the level of skill and degree of responsiveness demonstrated by the teacher when delivering the components prescribed by the EBP. Finally, *relational factors* involve aspects of treatment receipt, including the quality of the child–teacher relationship and level of child involvement. Each dimension captures a unique aspect of intervention delivery that is important to implementation research (Carroll & Nuro, 2002; McLeod, Southam-Gerow, et al., 2013).

In this article, we provide guidance for researchers in the development of treatment integrity measures for EBPs targeting the prevention and amelioration of EBD within an implementation science framework. We begin with a brief description of implementation science and then highlight the importance of measuring the four dimensions of treatment integrity in implementation research. After describing these dimensions, a conceptual model of program implementation for young children at risk for EBD served in early childhood settings is provided. We then present steps for developing psychometrically strong treatment integrity measures that have maximum utility for implementation research of EBP targeting this high-risk population. We conclude with heuristics that can help the field of early intervention advance the science of treatment implementation.

Implementation Science and Early Intervention and Prevention of EBD

Implementation science has been defined as “the scientific study of methods to promote the systematic uptake of research findings and other evidence-based practices into routine practices” (Eccles & Mittman, 2006, p. 1). Implementation science focuses on transferring efficacious programs and practices into authentic settings. While some programs designed to prevent or ameliorate the problem behaviors demonstrated by young children at risk for EBD have been identified (e.g., *PK PATHS*, Domitrovich et al., 2007; *Incredible Years*, Webster-Stratton et al., 2004), a major challenge for the field is large-scale implementation of these programs by early childhood practitioners in authentic early childhood settings (Domitrovich et al., 2010; Domitrovich, Moore, & Greenberg, 2012).

Implementation of EBPs for children at risk for EBD can be difficult due to the complexity of the intervention programs and the contexts in which they are implemented (Durlak, 2010). An important focus of implementation research is to understand how contextual issues influence the delivery of EBPs across a variety of settings (Mendel, Meredith, Schoenbaum, Sherbourne, & Wells, 2008; Schoenwald & Hoagwood, 2001). A number of factors in early childhood settings can influence program implementation (McLeod, Southam-Gerow, et al., 2013; Schoenwald et al., 2011), including level and type of teacher training (Pianta & Rimm-Kaufman, 2006), teachers' instructional ability (Domitrovich et al., 2010; Hamre et al., 2010), and the quality of teacher–child relationships (Driscoll & Pianta, 2010; Vo, Sutherland, & Conroy, 2012). In addition, more proximal factors such as individual teacher characteristics (e.g., years of experience, licensure/credentials), child characteristics (e.g., risk factors, level and type of problem behavior), and organizational variables at the program level are important (e.g., administrative support, program type/requirements [Head Start vs. community

child care], Baker, Kupersmidt, Voegler-Lee, Arnold, & Willoughby, 2010; Domitrovich et al., 2010; Durlak, 2010; Han & Weiss, 2005). Given the potential influence of these contextual factors, it is critical to study the relationship between contextual factors, EBP implementation, and child outcomes.

Implementation research has used models from different research traditions to study how contextual factors influence EBP implementation and outcomes. One such model is the quality of care framework (Donabedian, 1988; Mendel et al., 2008). Broadly speaking, quality of care research seeks to improve the outcomes of individuals who access health care across a variety of settings by studying how *structural elements* (e.g., contextual elements including attributes of settings, clients, and providers) and *processes of care* (e.g., activities and behaviors associated with giving and receiving care) influence *outcomes*. To improve outcomes, quality of care research attempts to establish causal links between each element (Donabedian, 1988). This framework represents a logical starting point for conceptualizing and studying the relation between the various levels of the program/school system in which EBPs are delivered (see, for example, Garland, Bickman, & Chorpita, 2010; Knox & Aspy, 2011). The quality of care framework emphasizes that implementation research must study how intervention programs are delivered and received. Consequently, this research moves beyond a primary focus on outcomes, as is the case with efficacy research, and instead places an emphasis on assessing program implementation integrity (i.e., extent to which the EBP was delivered as designed; McLeod, Southam-Gerow, et al., 2013).

Well-validated treatment integrity measures are critical to implementation research (Durlak, 2010; Southam-Gerow & McLeod, 2013). Unfortunately, the early intervention field is lacking measures suitable for assessing implementation integrity in general (Wolery, 2011), including the implementation of EBPs by teachers designed for children at risk for EBD. As the field of early intervention moves into implementation research, guidelines for measures are needed (Kratochwill et al., 2012). It is possible that treatment integrity research from the treatment technology field (i.e., treatment development and evaluation research, see Carroll & Nuro, 2002, for a discussion) may be useful in addressing the conceptual and methodological gap for measuring implementation integrity—allowing us to systematically identify the difference between what we *think* and what we *know* affects implementation and program effectiveness.

Measuring Implementation Integrity in Programs Targeting Early Intervention and Prevention of EBDs

Treatment integrity research can be used to inform the measurement of implementation integrity of EBPs targeting early intervention and prevention of EBDs. The term *treatment integrity* refers to the extent to which an intervention was delivered as intended (Sanetti & Kratochwill, 2009; McLeod et al., 2009; Southam-Gerow & McLeod, 2013). The education field as a whole (including early intervention and prevention of EBD) has yet to settle on a single definition of treatment integrity (Sanetti & Kratochwill, 2009). A number of different terms have been proposed: *treatment fidelity*, *treatment adherence*, and *intervention integrity* (Dane & Schneider, 1998; Jones, Clarke, & Power, 2008; McLeod et al., 2009; Sanetti & Kratochwill, 2009). To advance the study of implementation integrity in

early intervention, it will be useful for the field to adopt a common definition. We believe that the four-dimension treatment integrity model pulled from the treatment integrity field described above and defined in greater detail below can be used as a starting point to help the field move toward a common definition (see McLeod, Southam-Gerow, et al., 2013, for a discussion).

Treatment technology researchers have asserted that to support the development and evaluation of interventions, it is important for programs to have (a) a standardized treatment model (e.g., treatment protocol), (b) documented procedures for training and supervising interventionists, and (c) tools to monitor the four dimensions of treatment integrity (treatment adherence, competence, treatment differentiation, relational factors; Carroll & Nuro, 2002). While each of these elements is needed to interpret findings from randomized clinical trials (RCTs), each element is also important to have in implementation research to maximize the effectiveness of programs in authentic settings (i.e., sustainability; McLeod, Southam-Gerow, et al., 2013). Many EBPs targeting early prevention and intervention of children with or at risk for EBD have only two of these elements, standardized treatment protocol and training procedures. However, most early intervention and prevention programs for EBD have not developed the tools to monitor the four dimensions of treatment integrity. As discussed below, this may represent a potential barrier to measurement of the implementation and effectiveness of an intervention as the field moves toward implementation research.

In general, the measurement of treatment integrity is underdeveloped in the school-based prevention field, with few studies adequately measuring the integrity of program implementation (Sanetti, Gritter, & Dobey, 2011). Recent reviews in the special education and school psychology fields indicate that fewer than half of the studies reporting on interventions include integrity data (Harn, Parisi, & Stoolmiller, 2013; Sanetti et al., 2011; Swanson, Wanzek, Haring, Ciullo, & McCulley, 2011). Moreover, nearly all the studies focus on adherence, leaving teacher competence, treatment differentiation, and relational factors relatively unstudied (Sanetti et al., 2011). Similarly, treatment integrity has been narrowly defined and measured in the early intervention field (Wolery, 2011), and the same pattern is seen in studies of EBPs targeting the prevention and intervention of EBDs. In the following section, we define the four treatment integrity dimensions and describe how each can be applied to measuring the integrity of teacher-delivered programs targeting EBDs in young children. To help illustrate each point, we provide examples from our own work in the BEST in CLASS research project, which is a Tier 2 manualized program for young children at risk for EBD (Conroy, Sutherland, Vo, Carr, & Ogston, 2013; Vo et al., 2012), and provide additional examples from other programs.

Treatment Adherence

Treatment adherence refers to the extent to which a teacher delivers the components contained within an EBP as designed (e.g., delivers the prescribed intervention components contained within a specific treatment protocol). Measuring adherence to a treatment protocol is the most commonly measured dimension of treatment integrity in the education field (Sanetti, Dobey, & Gritter, 2012). Typically, researchers use indirect assessment, such as checklists (e.g., Hamre et al., 2012) to evaluate adherence. Using a dichotomous “present”

or “absent” system, an observer uses the checklist to determine whether each component within an EBP was present or not. Although checklists are cost-effective and easy to use, they have some limitations, namely, checklists do not assess the dosage of specific prescribed intervention components (Hogue, Liddle, & Rowe, 1996; Wolery, 2011). Because the delivery of components may vary across teachers, classrooms, and schools/programs, capturing variability in the delivery of particular components of the program is important in implementation research (Durlak, 2010). For example, simply counting the frequency of the delivery of a component can misrepresent the therapeutic process by giving a higher weight to components that are used more often but may not fairly weigh those used in a more thorough manner (McLeod, Islam, & Wheat, 2013). Therefore, when measuring adherence, it is critical to use rating schemes that capture the variability or “extensiveness” of intervention delivery (McLeod, Islam, et al., 2013; Wolery, 2011).

We measured adherence to the BEST in CLASS program by assessing teachers’ implementation of each prescribed intervention practice included in the model through the use of a Likert-type extensiveness scale (Sutherland, McLeod, Conroy, Abrams, & Smith, 2013). Likert-type extensiveness scales consider the breadth and depth of intervention delivery when generating scores (see Carroll et al., 2000, and Hill, O’Grady, & Elkin, 1992, for exemplars). This type of system provides an estimate of the extent to which a teacher delivers a specific intervention practice during an observational period when implementing the BEST in CLASS program. To illustrate, one practice in the BEST in CLASS program is the use of “behavior-specific praise” with focal children. To evaluate adherence, we use a 7-point Likert-type scale (1 = *never*, 3 = *some*, 5 = *considerable*, 7 = *very extensive*) to rate the teacher’s use of behavior-specific praise, rather than a dichotomous scale (i.e., present/absent). When completing the adherence rating, the observer considers the frequency (i.e., the number of times throughout the observation that the teacher provides behavior-specific praise) and the thoroughness (i.e., the persistence with which the teacher uses behavior-specific praise to achieve desirable child behavior [i.e., engagement]). Using this adherence measurement approach can help researchers examine teachers’ implementation of the individual intervention practices that comprise a program and produce a separate score for each intervention practice, thus allowing a more nuanced assessment of treatment adherence. With a similar approach to measuring adherence, Bierman, Nix, Greenberg, Blair, and Domitrovich (2008) used a 4-point Likert-type scale to assess the extensiveness with which teachers in the Head Start REDI project covered the core components of the PK PATHS.

Treatment Differentiation

Whereas treatment adherence assesses whether a teacher follows a particular approach, treatment differentiation evaluates whether (and “to where”) teachers deviate from the program (McLeod et al., 2009). It is important to conduct treatment differentiation checks when comparing two active treatments (e.g., EBP vs. business as usual [BAU]). In such cases, it is essential to establish whether intervention practices prescribed in the intervention condition are found in the BAU condition (i.e., assess for treatment diffusion). In addition, it is also important to determine the presence of other intervention practices that could interfere with the effectiveness of the EBP if present. Most treatment differentiation checks

to date have assessed for treatment diffusion and/or whether treatment adherence is consistent across sites (e.g., Odom et al., 2010). Differentiation checks have typically used adherence measures in the intervention and BAU conditions (Wolery, 2011), which is how we are assessing differentiation in our BEST in CLASS efficacy trial. While these methods are useful for measuring treatment diffusion in efficacy research, they are not sufficient for implementation research because adherence measures are designed to measure only the core components of an EBP and fail to measure other programs or intervention practices, not part of the EBP that could be naturally occurring in classrooms.

To assess for treatment diffusion, differentiation checks must (a) establish whether any undesirable components were delivered in the EBP condition and (b) characterize the intervention components delivered in the BAU condition, components prescribed and proscribed (i.e., undesirable components) by the EBP. When EBPs are found to have more positive results than BAU, it is important to clarify what the intervention practices were in the BAU condition to help interpret findings. In those cases where BAU outperforms an EBP or there are no differences, it is important to clarify what BAU practices may have contributed to the positive or neutral outcomes. This approach enhances the informational value of BAU and is critical in interpreting findings from implementation research.

Given that intervention practices used in early childhood settings are not well-characterized (Weiland, Ulvestad, Sachs, & Yoshikawa, 2013), measuring treatment diffusion in BAU classrooms presents unique challenges, especially for programs that target children with problem behaviors. Not surprisingly, teachers often use various combinations of intervention practices and/or programs to address children's problem behaviors in their classrooms (Sutherland, Lewis-Palmer, Stichter, & Morgan, 2008). To fully assess treatment diffusion, differentiation checks must measure a wide array of intervention practices, including intervention practices not included in the specific EBP (Waltz, Addis, Koerner, & Jacobson, 1993). At present, a measure designed to assess for the wide array of social, behavioral, cognitive, emotional, and pre-academic intervention components likely to be delivered in early childhood classrooms does not exist, which represents a significant gap for measuring differentiation in the field. As a result, research studies to date have failed to address treatment differentiation. In fact, Bierman et al. (2008) noted their randomized controlled trial of Head Start REDI lacked differentiation data, which limited their ability to interpret intervention effects. Clearly, it will be important to address this measurement gap to move the field forward in implementation research.

Competence

Measuring *how well* an EBP is implemented is crucial in implementation research. Competence is the quality of intervention delivery and is hypothesized to play an instrumental role in intervention research (Harn et al., 2013). Whereas treatment adherence focuses on whether a teacher delivers specific prescribed intervention components contained within an EBP, competence focuses on whether a teacher knows when and how to deliver an intervention for maximum impact (Barber, Sharpless, Klostermann, & McCarthy, 2007). The competent delivery of intervention components requires a teacher to adapt specific components to meet the unique characteristics of the classroom and the individual child. Measuring teacher competence of these individual intervention components is of

particular relevance to implementation research, given the influence of factors such as teacher training (Pianta & Rimm-Kaufman, 2006) and teaching ability (Domitrovich et al., 2010; Hamre et al., 2010) on intervention outcomes.

While there have been recent calls for researchers to reliably measure competence in authentic settings and further examine its relationship with treatment integrity (e.g., Harn et al., 2013), competence has proven difficult to define and measure. The few existing measures are designed to assess competence pertaining to the delivery of intervention components contained in specific programs (called “technical” or “limited-domain” competence; Barber et al., 2007). These measures focus on the level of skill and degree of responsiveness a teacher displays when delivering the specific components contained within an EBP and are fairly narrow in focus.

Different strategies can be used to measure teacher competence (McLeod, Southam-Gerow, et al., 2013; Southam-Gerow & McLeod, 2013). Researchers most often use observational methods (e.g., a Likert-type scale; Bierman et al., 2008) completed by an independent rater. For competence measures, exemplary scoring strategies involve ratings on a Likert-type scale that estimate the technical quality of teachers’ use of intervention components (skillfulness), their timing of intervention components, and the appropriateness of use of the intervention components for the given child and situation (teacher responsiveness). This scoring strategy has been used in exemplary competence coding systems that evidence strong psychometric characteristics (e.g., Carroll et al., 2000; Hogue, Dauber, et al., 2008) and was used in the Head Start REDI (Bierman et al., 2008) and PK PATHS (Domitrovich et al., 2007) studies. Bierman et al. (2008) reported that Head Start REDI trainers rated teachers’ implementation quality monthly on a 6-point Likert-type scale (1 = *poor* to 6 = *exemplary*), while in the PK PATHS study, observers rated implementation quality once per month on a 4-point scale. To measure technical competence in the BEST in CLASS intervention, we determine how well a teacher delivers the prescribed intervention practices using a 7-point Likert-type scale (1 = *very poor*, 7 = *excellent*). This allows us to evaluate the teachers’ competence in implementation of each of the intervention components that comprise our model.

Relational Factors

Whereas treatment adherence, treatment differentiation, and teacher competence focus on the technical aspects of treatment delivery (how the intervention practices of an EBP are delivered), the relational factors focus on treatment receipt (i.e., how those intervention practices are received by the child). Traditional definitions of treatment integrity have not included relational factors (e.g., Perepletchikova & Kazdin, 2005). We assert that this domain is critical in implementation research. Adherence to an EBP protocol is not sufficient if a child does not participate in the program. Some characteristics of young children with or at risk for EBD may affect child engagement (Qi & Kaiser, 2003; Quesenberry et al., 2011). Furthermore, a program that actively engages a homogeneous sample of children may fail to engage a more diverse sample of children that may be found in typical early childhood classrooms. Simply focusing on whether the technical aspects of an EBP are delivered therefore may miss important information needed for interpreting study findings (e.g., the EBP failed to engage the children so the delivery of the program needs to be modified).

While the association between relational factors and child outcomes has not typically been a focus of integrity measurement in programs targeting EBDs, lessons learned from other areas of research (e.g., youth therapy) may be useful as the field advances, particularly for interventions that have a social, emotional, and/or behavioral emphasis. To illustrate, therapist–client alliance and client responsiveness are linked to symptom reduction in youth psychotherapy (Karver, Handelsman, Fields, & Bickman, 2006; McLeod, 2011). A therapist’s ability to (a) cultivate a relationship with the client (child or parent) marked by warmth and trust (alliance) and (b) promote the child’s participation in therapeutic activities (involvement) is considered instrumental in promoting positive outcomes (Chu et al., 2004; Chu & Kendall, 2004; McLeod, 2011). Similarly, research in the field of early childhood suggests that teacher–child relationships characterized by the closeness between a teacher and a child are related to desirable developmental outcomes (Driscoll & Pianta, 2010; Pianta & Stuhlman, 2004). Research also suggests that increasing positive teacher–child interactions (as opposed to coercive interactions) may be particularly salient for improving outcomes for high-risk children (Burchinal, Howes, & Kontos, 2002). Thus, we suggest that relational factors are likely to be an important aspect of treatment integrity measurement to advance implementation research in EBPs targeting early intervention and prevention of EBD.

As one example of the measurement of relational factors, Domitrovich et al. (2010) rated children’s engagement in intervention classrooms only during PK PATHS lessons of the Head Start REDI program on a 4-point Likert-type scale. While significant increases across time were not noted, mean ratings of child engagement during PK PATHS lessons were generally high, ranging from 3.3 to 3.55. In our work, we have included two items to assess relational factors in our integrity measures (i.e., *Child responsiveness to teacher behavior* and *Child engagement*). Observers code the extensiveness of these behaviors on a 7-point Likert-type scale (1 = *not at all*; 7 = *very extensive*) in BEST in CLASS and BAU classrooms.

Summary

The measurement of these four dimensions of treatment integrity is underdeveloped in the early intervention field, particularly in research on teacher-delivered EBPs that address the needs of young children at risk for EBDs. To illustrate, a scan of the literature identified six EBPs evaluated in eight RCTs that met the following criteria: (a) target children aged 3 to 4 at risk for EBDs, (b) teachers delivered instructional practices that targeted child problem behaviors and/or pre-academic outcomes, and (c) children randomly assigned to condition. Our abbreviated review included the following EBPs: *Chicago School Readiness Project* (CSRP; Raver et al., 2009), *Incredible Years* (Webster-Stratton, Reid, & Hammond, 2001, 2004; Webster-Stratton, Reid, & Stoolmiller, 2008), *Preschool PATHS* (Domitrovich et al., 2007), Head Start Research-Based, Developmentally Informed (REDI; Bierman et al., 2008), Reaching Educators, Children, and Parents (RECAP; Han, Catron, Weiss, & Marciel, 2005), and *Tools of the Mind* (Barnett et al., 2008). Table 1 provides an overview of each dimension measured by each program. As seen, only three of the eight studies (37.5%) reported on teacher adherence and competence of delivery. One study (12.5%) reported on relational factors, and none of the studies reported on treatment differentiation.

Table 1
Abbreviated Review of Treatment Integrity Measures of Evidence-Based Programs

Program	Study	Adherence	Competence	Differentiation	Relational factors
CSRP	Raver et al. (2009)	Yes	Yes	No	No
Incredible Years	Webster-Stratton, Reid, and Hammond (2001)	No	No	No	No
	Webster-Stratton, Reid, and Hammond (2004)	No	No	No	No
	Webster-Stratton, Reid, and Stoolmiller (2008)	No	No	No	No
REDI	Bierman, Nix, Greenberg, Blair, and Domitrovich (2008)	Yes	Yes	No	Yes
Promoting Alternative Thinking Strategies	Domitrovich, Cortes, and Greenberg (2007)	No	Yes	No	No
RECAP	Han, Catron, Weiss, and Marciel (2005)	No	No	No	No
Tools of the Mind	Barnett et al. (2008)	Yes	No	No	No

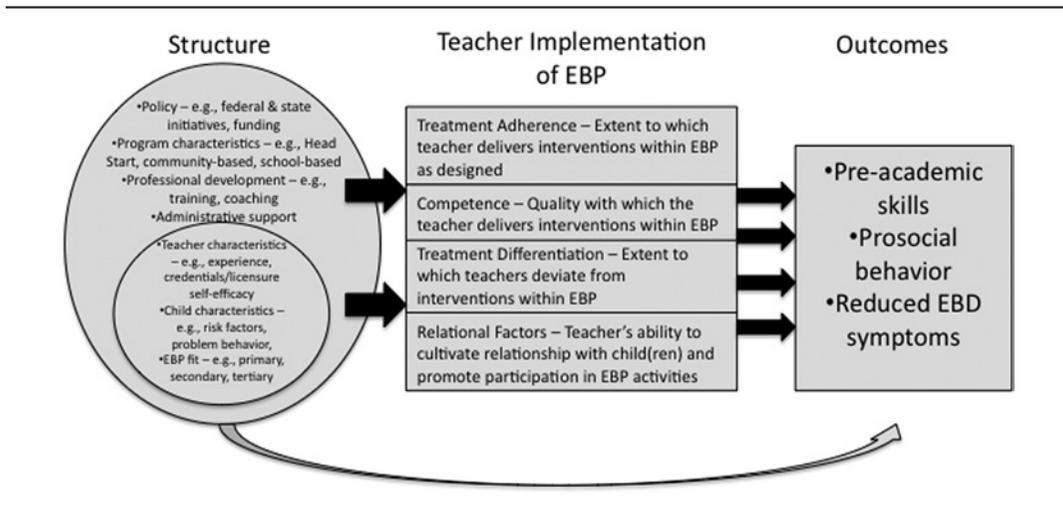
Note. CSRP = Chicago School Readiness Project; REDI = Research-Based, Developmentally Informed; RECAP = Reaching Educators, Children, and Parents.

Not surprisingly, adherence and competence tend to be the dimensions of treatment integrity most often assessed, but most studies do not report on both dimensions. In addition, treatment differentiation and relational factors that affect implementation are understudied in this literature. To advance the field's efforts in implementation science, we believe a comprehensive approach to treatment integrity is essential. The following section describes a treatment implementation framework that integrates treatment integrity measurement into the quality of care model (Donabedian, 1988; Mendel et al., 2008). We believe this framework can be used to guide the development of integrity measures that will have maximum applicability and utility for use in implementation research of EBPs addressing the needs of young children at risk for EBD.

Treatment Implementation Model

The measurement of implementation integrity must occur within frameworks defined by theoretical and empirical work (Kratochwill et al., 2012; Sanetti & DiGennaro Reed, 2012). Figure 1 provides a framework for approaching implementation science for EBPs in early childhood settings that emphasize prevention and intervention of EBDs. The model is designed to help promote an understanding of how factors present at different levels of the program or school context may influence the implementation and outcome of programs for young children at risk for EBD. Placed within the quality of care framework, the model draws from multiple lines of research and integrates facets of treatment integrity from the treatment technology field (McLeod, Southam-Gerow, et al., 2013; Sanetti & Kratochwill, 2009) and treatment implementation models (Aarons, Hurlburt, & Horwitz, 2011) with the

Figure 1
Conceptual Model of Treatment Implementation



theory and findings from therapy process research focused on how EBPs produce change (Doss, 2004; McLeod, Islam, et al., 2013). The end product is a model that details the potential relations between the structural elements of program/school settings, the implementation of interventions, and child outcomes.

A conceptual model (illustrated in Figure 1) based on the quality of care framework may be useful to guide research focused on the implementation of programs for young children at risk for EBD in authentic settings. The left side of the model focuses on some characteristics of settings in which EBPs are implemented that might influence treatment implementation and outcome (Aarons et al., 2011). This part of the model identifies macro (e.g., policy, characteristics of programs) and micro (e.g., teacher, child, EBP fit) contexts, which may, either in isolation or in combination, influence implementation and outcomes of EBPs. The middle section includes the four dimensions of treatment integrity that represent the critical aspects of treatment implementation. Each dimension captures a unique technical (what the teacher does) or relational (teacher–child relationship, child engagement) aspect of program implementation. Finally, the right portion of the diagram represents desirable child outcomes (pre-academic skills, prosocial behavior, reduced EBD symptoms) associated with EBPs targeting the prevention and amelioration of EBD (e.g., Domitrovich et al., 2007; Vo et al., 2012; Webster-Stratton et al., 2004).

Because the model is based on empirical and conceptual work outlined in implementation research, it represents an ideal framework to inform the development of integrity measures within the field of early intervention for young children at risk for EBD. Developing measures to assess the dimensions of EBP implementation identified in the model will produce tools that allow researchers to study (a) the integrity of program implementation in early childhood classrooms and (b) how contextual elements influence program implementation and outcomes (Han & Weiss, 2005; Southam-Gerow & McLeod, 2013). In addition, grounding measure development in a conceptual framework addresses concerns raised by researchers about the limited theoretical basis for existing integrity

measures (Sanetti & DiGennaro Reed, 2012). Next, we describe how lessons learned from existing treatment integrity research can help inform the development of integrity measures designed to address the measurement gaps within this framework. Again, we provide examples from our work developing the BEST in CLASS treatment integrity measure (i.e., the BEST in CLASS Adherence and Competence Scale [BiCACS]; Sutherland et al., 2013).

Treatment Integrity Measure Development

As noted by Durlak (2010), intervention development that ignores treatment integrity is incomplete. Unfortunately, the measurement of treatment integrity in the field related to interventions focused on the prevention of EBDs delivered by early childhood teachers has several key gaps: (a) lack of agreed on definition of integrity, (b) lack of measurement of different dimensions of integrity, and (c) lack of theoretical frameworks used to inform integrity measurement. Developing treatment integrity measures that map on to the “active ingredients” or “core components” of an intervention allows researchers to interpret study findings with greater precision and meaning (Durlak, 2010; Wolery, 2011). By developing integrity measures in a systematic fashion, researchers can produce integrity measures suitable for implementation research. We next describe steps that can assist researchers in developing integrity measures that align with their own intervention development work.

Scale and Subscale Focus

The first step is to identify the components and sub-components of the EBP that should be assessed to determine treatment integrity; these will become the subscales of the integrity assessment measure. As illustrated previously in our discussion of the BEST in CLASS treatment integrity measures, the scale and subscales should be defined according to conceptual (e.g., behavioral, transactional) and/or integrity (e.g., adherence vs. competence) domains. Depending on the focus of the intervention (e.g., primary, secondary, or tertiary), it may also be important to disentangle group- and individual-focused intervention components. For example, some prevention programs may be more universal in nature, with teachers focusing intervention efforts on whole groups or classes of children with the purpose of preventing problem behaviors (e.g., Webster-Stratton et al., 2008). Other interventions, such as BEST in CLASS (Vo et al., 2012), may be more targeted in nature, with teachers focusing intervention efforts on selected children who are displaying elevated levels of problem behavior. Failure to distinguish between the two levels may obscure important individual differences in treatment integrity (e.g., a teacher may evidence different levels of integrity in delivering the intervention components directed at specific children vs. those that target larger groups of children). Thus, because BEST in CLASS is a Tier 2 or secondary intervention, we focused our integrity measurement on teacher delivery of BEST in CLASS components to focal children identified as at risk for EBD. To illustrate, if a teacher delivered a BEST in CLASS intervention component (e.g., presenting opportunities to respond) to a child in the classroom who was not identified as at risk, the teacher’s delivery of the intervention component would not be coded by the observer as an indicator of treatment integrity because the focal child was not a targeted recipient of the specific component.

Item Development

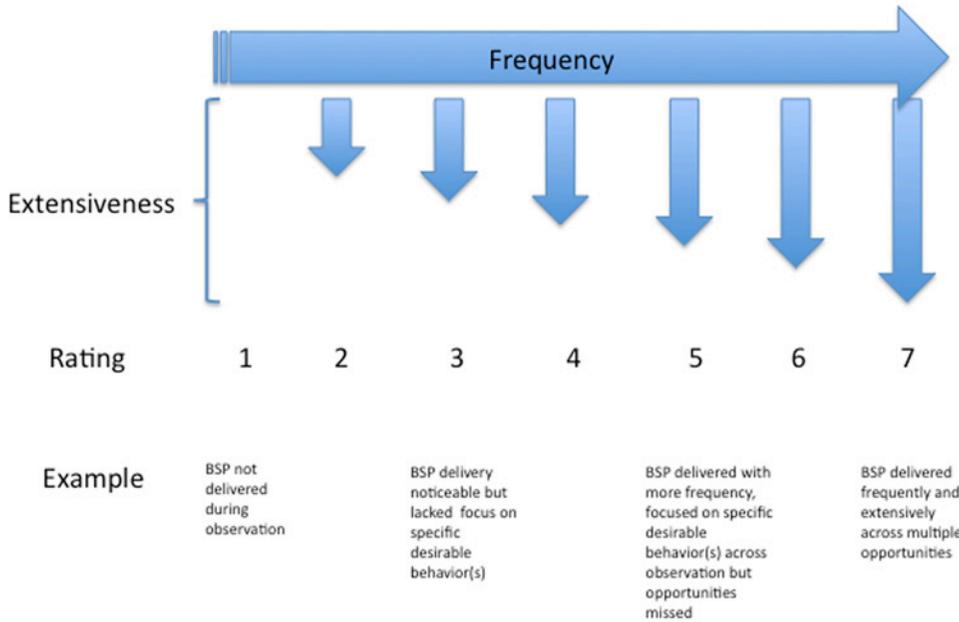
The next step is to create items for each scale or subscale. In generating items, it is necessary to determine an appropriate level of inference (McLeod, Southam-Gerow, et al., 2013). Goldfried and Padawer (1982) proposed a framework for scoring, from more to less specific, that consists of three levels: technique (e.g., specific intervention practice, such as, behavior-specific praise), therapeutic strategy (e.g., teacher's implementation of behavior-specific praise), and theoretical level (e.g., the relation between use of behavior-specific praise and child behavior). Some researchers have identified the middle level of inference, therapeutic strategy, as the most promising level for implementation research (McLeod, Southam-Gerow, et al., 2013). Defined as the goal or general principle that guides an intervention (e.g., altering environmental contingencies), therapeutic strategies address the domain of child functioning targeted by the teacher (e.g., problem behavior, social skills, emotion regulation). Focusing on this level allows researchers to test specific process-outcome relations (e.g., whether the promotion of appropriate behavior through the implementation of behavior-specific praise leads to reductions in behavioral problems). Developing items to assess the use of specific therapeutic strategies also increases utility by allowing investigators to group related intervention components under a single item. For example, effective components such as behavior-specific praise and differential reinforcement of alternative behaviors can be grouped under one item called *individual reinforcement*. Combining intervention components in this manner produces a manageable list of items that could be used to assess implementation integrity in early childhood classrooms, help characterize the interventions used in BAU classrooms, and aid in the development of treatment differentiation measures.

Given the prescribed (i.e., manualized) nature of the BEST in CLASS model, we have initially focused our item development on specific teaching practices that comprise our program model at Goldfried and Padawer's (1982) strategies level. To illustrate, the current version of our treatment integrity measure (i.e., BiCACS) contains eight items that cover six intervention practices (i.e., *3-5 rules are visible in classroom; Teacher reviews rules, addresses rule violations; Teacher maintains brisk instructional pace; Teacher provides precorrection; Teacher provides opportunities to respond; Teacher provides behavior-specific praise; Teacher provides corrective feedback; Teacher provides instructive feedback*) that are measured on two dimensions, Adherence and Competence. In addition, we have two items to assess Relational Factors (*Child responsiveness to teacher behavior; Child engagement*).

Scoring Strategy

The third step involves determining the appropriate scoring strategy. It is essential to match the scoring strategy to the purpose of the measure. As noted earlier, in implementation research, the scoring strategy should capture the breadth and depth of the specific components that comprise a program. Microanalytic scoring strategies (e.g., frequency counts) may not be a good fit for implementation research because these strategies fail to capture variation in intervention delivery. Here, we focus on exemplar scoring strategies used in previous treatment integrity research to rate adherence (e.g., Carroll et al., 2000; McLeod & Weisz, 2010) and competence (e.g., Carroll et al., 2000; Hogue, Henderson, et al., 2008).

Figure 2
Example of Adherence Rating of Behavior-Specific Praise



For adherence scales, exemplar scoring strategies involve extensiveness ratings of intervention components designed to measure the degree to which teachers use specific components. As discussed, using the BiCACS treatment integrity measure, coders estimate the extent to which teachers implement each component during an entire observation using a 7-point Likert-type scale with the following anchors: 1 = *not at all*, 3 = *somewhat*, 5 = *considerably*, and 7 = *extensively*. Extensiveness ratings are comprised of two key characteristics: thoroughness and frequency (see Figure 2). Thoroughness refers to the depth, complexity, or persistence with which the teacher engages in a given intervention component. Thoroughness is determined by (a) the concentration of effort or commitment the teacher puts into the component, (b) the detail in which the teacher describes the rationale for the component, (c) the depth or intensity of the component, (d) the extent to which the teacher follows through with the component, and/or (e) the extent to which the teacher pursues implementing the component across a session. For example, if a BEST in CLASS focal child is consistently off task, the teacher would receive low ratings on behavior-specific praise if she or he only made one attempt to use behavior-specific praise when the child exhibited desirable behavior and did not persist by identifying and praising any other instances of desirable behavior. The teacher would receive higher ratings if she or he demonstrated a consistent effort to use behavior-specific praise whenever the child was exhibiting desirable behavior, regardless of whether the intervention component resulted in increased child engagement. Frequency refers to the number of times throughout the observation that a given intervention component is executed (regardless of the thoroughness of the component in any particular segment). Thoroughness and frequency are considered in making an extensiveness rating on each item; therefore, extensiveness ratings provide

quantity, or dosage, information about each intervention component. In other words, these ratings determine how much of each intervention component the child is exposed to in a given observation session.

For competence scales, exemplar scoring strategies involve ratings that estimate the technical quality of teacher's delivery of intervention components (skillfulness) and the timing and appropriateness of the teacher's delivery for the given child and situation (teacher responsiveness). To rate competence on the BiCACS, coders use a 7-point Likert-type scale with the following anchors: 1 = *very poor*; 3 = *acceptable*; 5 = *good*; 7 = *excellent*. For each item, coders consider the extent to which a teacher demonstrated the following: (a) expertise, commitment and motivation, (b) clarity of communication, (c) appropriate timing of intervention components (responsiveness), and (d) ability to read and respond to where the child appears to be (responsiveness). For example, on the behavior-specific praise item, coders rate competence based on the quality of the teacher's delivery of specific praise. The dimensions taken into account when rating teacher delivery (skillfulness, timing, and appropriateness) are characteristics of the delivery of the praise statement such as (a) it is sincere, (b) it is contingent on a desirable behavior, and (c) it is focused on the child's effort. Therefore, if a teacher only delivers one behavior-specific praise statement during an observation session, resulting in a low Adherence rating of "2," the Competence rating could still be rated as Excellent ("7") if the statement was delivered promptly after a desirable behavior, was sincere, and was effort-focused.

Summary

These steps may provide guidance and examples for researchers interested in developing integrity measures to assess the range of dosage and quality features of implementation of the core intervention components comprising an EBP. These steps have been followed in psychotherapy research to produce psychometrically strong measures that are predictive of child outcomes and have informed implementation research (Hogue, Henderson, et al., 2008; Southam-Gerow et al., 2010). The preliminary psychometrics of the BiCACS are promising (Sutherland et al., 2013). Specifically, BiCACS items and subscales demonstrated fair to strong reliability, and results also supported the validity of the BiCACS, with the pattern of correlations among the items and subscales in the expected direction and the subscales were distinct from a teacher-report measure of the child-teacher relationship. Analyses also indicated that the BiCACS Adherence subscale was sensitive to changes in adherence over the course of the BEST in CLASS program. It is our view that psychometrically strong integrity measures designed for specific EBPs are an important step in advancing implementation research with the ultimate outcome of improving practices and programs targeting the prevention of EBDs.

Future Directions to Advance the Measurement of Implementation Integrity

Developing measures to assess the integrity of specific EBPs is an important first step in addressing gaps in treatment integrity research in the early intervention field. However

advances are needed in the methods and tools used to establish, maintain, and measure treatment integrity. We propose heuristics to guide the development of integrity measures suitable for implementation research in this field.

First, psychometrically strong measures are needed that assess each of the four integrity dimensions (Sanetti & DiGennaro Reed, 2012). As noted earlier, the competence, differentiation, and relational aspects of treatment integrity have not been measured in early intervention research for young children with EBDs. Measures for each of these dimensions are needed. In developing these measures, reliability at the item level is important to demonstrate so that researchers can assess integrity for each intervention component contained within an EBP. Validity of the items and scales is also critical to insure that researchers can make meaningful comparisons within and across studies using the same measure. Ultimately, the development of psychometrically strong measures will allow for integrity-outcomes analyses that can aid efforts to refine treatment models and guide implementation efforts (Durlak, 2010; Sanetti & Kratochwill, 2009). For example, having valid and reliable adherence and competence measures can help researchers ascertain whether the failure of an EBP to produce expected outcomes in an authentic setting is due to the program (i.e., adherence and competence were strong suggesting that the EBP did not work to address the specific EBD symptoms or class wide behavior challenges) or EBP implementation (i.e., adherence and competence were not strong, so future research efforts need to focus on increasing the effectiveness of teacher training and coaching; Schoenwald et al., 2011).

Second, observational and teacher-report measures are needed. Observational assessment is the gold standard in integrity research because it provides objective and highly specific information regarding interventionist performance (McLeod, Southam-Gerow, et al., 2013). However, observational coding may not always be practical when implementing an EBP on a large-scale basis or when assessing sustainability (Hogue, Dauber, & Henderson, 2013). Observational assessment is costly in terms of time and resources. Observations may not be suitable for capturing intervention components that are not used routinely in the classroom, and stakeholders (e.g., teachers, program administrators) do not always support this approach due in part to the perceived intrusiveness of observational methods (Yoder & Symons, 2010). Because psychometrically strong teacher-report measures would address some of these concerns, the development and use of observational and teacher-report integrity measures represent important goals for consideration in the development of treatment integrity measures. Of course, observational measures are needed to facilitate the evaluation of the validity of teacher-report measures given concerns regarding their accuracy (McLeod et al., 2009).

Third, integrity measures capable of assessing variability in treatment implementation are needed. It is expected that teachers may vary in the extent to which they deliver the different intervention components comprising an EBP, so it is important that integrity measures assess the breadth and depth of each intervention component. Durlak (2010) noted that implementation exists on a continuum, is variable across teachers, and is not dichotomous. Given these characteristics, data provided by the dichotomous checklists typically used are of limited use in implementation research. Treatment integrity measures should ideally assess the variability of implementation across the different components of EBPs (Durlak, 2010).

Fourth, tools are needed that measure a wide range of the cognitive, behavioral, emotional, social, and pre-academic intervention components occurring in a classroom (McLeod, Southam-Gerow, et al., 2013). At present, the field does not have tools suitable for characterizing BAU. Without the ability to characterize the practices used by teachers in BAU, it will be difficult to interpret findings generated by effectiveness trials that use BAU comparison conditions.

To maximize the utility of implementation measures, including the characterization of practices used in BAU, items should be designed to measure somewhat broad “therapeutic strategies” (Beutler & Baker, 1998), also called “practice elements” (Chorpita & Daleiden, 2009) or “evidence-based kernels” (Embry & Biglan, 2008). Developing items of this type have several advantages. First, the items are not protocol specific, so the measures can be used by more than one research team (McLeod, Southam-Gerow, et al., 2013). Second, the items would be suitable for assessing treatment differentiation and characterizing BAU. Finally, this process produces a manageable list of items that could be more easily used to assess the implementation integrity efforts in early childhood classrooms.

Conclusion

The development of psychometrically strong integrity measures can contribute to the advancement of implementation research; however, much work remains to be done in the early intervention field. In this article, we have proposed a redefinition of the measurement of treatment integrity with a focus on development of comprehensive measures that focus on adherence, differentiation, competence, and relational factors. In addition, we have provided a conceptual model for directing this research and outlined a process for developing integrity measures. If children’s behavioral and developmental outcomes are going to be maximized by early intervention in authentic settings, then researchers and intervention developers must focus their efforts on developing measures to assess whether EBPs are implemented with integrity and skill as much as they have focused on efforts to assess the outcome produced by their programs.

References

- Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research, 38*, 4-23. doi:10.1007/s10488-010-0327-7
- Baker, C. N., Kupersmidt, J. B., Voegler-Lee, M. E., Arnold, D., & Willoughby, M. T. (2010). Predicting teacher participation in a classroom-based, integrated preventive intervention for preschoolers. *Early Childhood Research Quarterly, 25*, 270-283. doi:10.1016/j.ecresq.2009.09.005
- Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Science and Practice, 38*, 493-500. doi:10.1037/0735-7028.38.5.493
- Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly, 23*, 233-313. doi:10.1016/j.ecresq.2008.03.001

- Becker, K. D., & Domitrovich, C. E. (2011). The conceptualization, integration, and support of evidence-based interventions in the schools. *School Psychology Review, 40*, 582-589.
- Berlin, L. J., Brooks-Gunn, J., McCarton, C., & McCormick, M. C. (1998). The effectiveness of early intervention: Examining risk factors and pathways to enhance development. *Preventive Medicine, 27*, 238-245.
- Beutler, L. E., & Baker, M. (1998). The movement toward empirical validation: At what level should we analyze, and who are the consumers? In K. S. Dobson & K. D. Craig (Eds.), *Empirically supported therapies* (pp. 43-65). Thousand Oaks, CA: SAGE.
- Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and Psychopathology, 20*, 821-843. doi:10.1017/S0954579408000394
- Burchinal, M., Howes, C., & Kontos, S. (2002). Structural predictors of child care quality in child care homes. *Early Childhood Research Quarterly, 17*, 87-105. doi:http://dx.doi.org/10.1016/S0885-2006(02)00132-1
- Carroll, K. M., Nich, C., Sifty, R. L., Nuro, K. F., Frankforter, T. L., Ball, S. A., & Rounsaville, B. J. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research. *Drug and Alcohol Dependence, 57*, 225-238.
- Carroll, K. M., & Nuro, K. F. (2002). One size cannot fit all: A stage model for psychotherapy manual development. *Clinical Psychology: Science and Practice, 9*, 396-406.
- Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology, 77*, 566-579. doi:10.1037/a0014565
- Chu, B. C., Choudhury, M. S., Shortt, A. L., Pincus, D. B., Creed, T. A., & Kendall, P. C. (2004). Alliance, technology, and outcome in the treatment of anxious youth. *Cognitive and Behavioral Practice, 11*, 44-55. doi:http://dx.doi.org/10.1016/S1077-7229(04)80006-3
- Chu, B. C., & Kendall, P. C. (2004). Positive association of child involvement and treatment outcome within a manual-based cognitive-behavioral treatment for children with anxiety. *Journal of Consulting and Clinical Psychology, 72*, 821-829.
- Conroy, M. A., Sutherland, K. S., Vo, A. K., Carr, S. E., & Ogston, P. (2013). Early childhood teachers' use of effective instructional practices and the collateral effects on young children's behavior. *Journal of Positive Behavior Interventions*. Advance online publication. doi:10.1177/1098300713478666
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23-45.
- Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *The Journal of Primary Prevention, 28*, 67-91.
- Domitrovich, C. E., Gest, S. D., Jones, D., Gill, S., & DeRousie, R. M. S. (2010). Implementation quality: Lessons learned in the context of the Head Start REDI trial. *Early Childhood Research Quarterly, 25*, 284-298.
- Domitrovich, C. E., Moore, J. E., & Greenberg, M. T. (2012). Maximizing the effectiveness of social-emotional interventions for young children through high-quality implementation of evidence-based interventions. In B. Kelly & D. F. Perkins (Eds.), *Handbook of implementation science for psychology in education* (pp. 207-229). Cambridge, UK: Cambridge University Press.
- Donabedian, A. (1988). The quality of care: How can it be assessed? *Journal of the American Medical Association, 260*, 1743-1748.
- Doss, B. D. (2004). Changing the way we study change in psychotherapy. *Clinical Psychology: Science and Practice, 11*, 368-386. doi:10.1093/clipsy.bph094
- Driscoll, K. C., & Pianta, R. C. (2010). Banking time in Head Start: Early efficacy of an intervention designed to promote supportive teacher-child relationships. *Early Education and Development, 21*, 38-64. doi:10.1080/10409280802657449
- Durlak, J. A. (2010). The importance of doing well in whatever you do: A commentary on the special section, "Implementation research in early childhood education." *Early Childhood Research Quarterly, 25*, 348-357. doi:10.1016/j.ecresq.2010.03.003
- Eccles, M. P., & Mittman, B. S. (2006). Welcome to implementation science. *Implementation Science, 1*, Article 1.
- Embry, D. D., & Biglan, A. (2008). Evidence-based kernels: Fundamental units of behavioral influence. *Clinical Child and Family Psychology Review, 11*, 75-113. doi:10.1007/s10567-008-0036-x

- Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).
- Garland, A. F., Bickman, L., & Chorpita, B. F. (2010). Change what? Identifying quality improvement targets by investigating usual mental health care. *Administration and Policy in Mental Health, 37*, 15-26. doi:10.1007/s10488-010-0279-y
- Goldfried, M. R., & Padawer, W. (1982). Current status and future directions in psychotherapy. In M. R. Goldfried (Ed.), *Converging themes in psychotherapy: Trends in psychodynamic, humanistic, and behavioral practice* (pp. 3-49). New York, NY: Springer.
- Hamre, B. K., Justice, L., Pianta, R. C., Kilday, C., Sweeny, B., Downer, J., & Leach, A. (2010). Implementation fidelity of the My Teaching Partner literacy and language activities: Associations with preschoolers' language and literacy growth. *Early Childhood Research Quarterly, 25*, 329-347. doi:10.1016/j.ecresq.2009.07.002
- Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., & Scott-Little, C. (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal, 49*, 88-123. doi:10.3102/0002831211434596
- Han, S. S., Catron, T., Weiss, B., & Marciel, K. K. (2005). A teacher-consultation approach to social skills training for pre-kindergarten children: Treatment model and short-term outcome effects. *Journal of Abnormal Child Psychology, 33*, 681-693. doi:10.1007/s10802-005-7647-1
- Han, S. S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology, 33*, 665-679. doi:10.1007/s10802-005-7646-2
- 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.
- Hill, C. E., O'Grady, K. E., & Elkin, I. (1992). Applying the Collaborative Study Psychotherapy Rating Scale to rate therapist adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management. *Journal of Consulting and Clinical Psychology, 60*, 73-79.
- Hogue, A., Dauber, S., Chinchilla, P., Fried, A., Henderson, C., Inclan, J., & Liddle, H. A. (2008). Assessing fidelity in individual and family therapy for adolescent substance abuse. *Journal of Substance Abuse Treatment, 35*, 137-147.
- Hogue, A., Dauber, S., & Henderson, C. E. (2013). Therapist self-report of evidence-based practices in usual care for adolescent problems: Factor and construct validity. *Administration and Policy in Mental Health and Mental Health Services Research*. Advance online publication. doi:10.1007/s10488-012-0442-8
- Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology, 76*, 544-555. doi:10.1037/0022-006X.76.4.544
- Hogue, A., Liddle, H. A., & Rowe, C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy: Theory, Research, Practice, Training, 33*, 332-345.
- Jones, H. A., Clarke, A. T., & Power, T. J. (2008, Spring). Expanding the concept of intervention integrity: A multidimensional model of participant engagement. *In Balance: Society of Clinical Child and Adolescent Psychology Newsletter, 23*(1), 4-5. Retrieved from <http://enewsletter.clinicalchildpsychology.org/InBalance/sites/default/files/Spring%202008.pdf>
- Joyce, B., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Karver, M. S., Handelsman, J. B., Fields, S., & Bickman, L. (2006). Meta-analysis of therapeutic relationship variables in youth and family therapy: The evidence for different relationship variables in the child and adolescent treatment outcome literature. *Clinical Psychology Review, 26*, 50-65.
- Knox, L. M., & Aspy, C. B. (2011). Quality improvement as a tool for translating evidence based interventions into practice: What the youth violence prevention community can learn from healthcare. *American Journal of Community Psychology, 48*, 56-64. doi:10.1007/s10464-010-9406-x
- Kratochwill, T. R., Hoagwood, K. E., Kazak, A. E., Weisz, J. R., Hood, K., Vargas, L. A., & Banez, G. A. (2012). Practice-based evidence for children and adolescents: Advancing the research agenda in schools. *School Psychology Review, 41*, 215-235.

- McLeod, B. D. (2011). The relation of the alliance with outcomes in youth psychotherapy: A meta-analysis. *Clinical Psychology Review, 31*, 603-616. doi:10.1016/j.cpr.2011.02.003
- McLeod, B. D., Islam, N. Y., & Wheat, E. (2013). Designing, conducting, and evaluating therapy process research. In J. Comer & P. Kendall (Eds.), *The Oxford handbook of research strategies for clinical psychology* (pp. 142-164). New York, NY: Oxford University Press.
- McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodriguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychological treatment quality indicator. *Clinical Psychology: Science and Practice, 20*, 14-32. doi:10.1111/cpsp.12020
- McLeod, B. D., Southam-Gerow, M. A., & Weisz, J. R. (2009). Conceptual and methodological issues in treatment integrity measurement. *School Psychology Review, 38*, 541-546.
- McLeod, B. D., & Weisz, J. R. (2010). The Therapy Process Observational Coding System for Child Psychotherapy Strategies Scale. *Journal of Clinical Child and Adolescent Psychology, 39*, 436-443. doi:10.1080/15374411003691750
- Mendel, P., Meredith, L. S., Schoenbaum, M., Sherbourne, C. D., & Wells, K. B. (2008). Interventions in organizational and community context: A framework for building evidence on dissemination and implementation in health services research. *Administration and Policy in Mental Health and Mental Health Services Research, 35*, 21-37. doi:10.1007/s10488-007-0144-9
- Nelson, J. R., Stage, S., Duppong-Hurley, D., Synhorst, L., & Epstein, M. H. (2007). Risk factors predictive of the problem behavior of children at risk for emotional and behavioral disorders. *Exceptional Children, 73*, 367-379.
- Odom, S. L., Fleming, K., Diamond, K., Lieber, J., Hanson, M., Butera, G., & Marquis, J. (2010). Examining different forms of implementation and in early childhood curriculum research. *Early Child Research Quarterly, 25*, 314-328.
- Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology Science and Practice, 12*, 365-383. doi:10.1093/clipsy/bpi045
- Pianta, R. C., & Rimm-Kaufman, S. (2006). The social ecology of the transition to school: Classrooms, families, and children. In K. McCartney & D. Phillips (Eds.), *Blackwell handbooks of developmental psychology* (pp. 490-507). Malden, MA: Blackwell Publishing.
- Pianta, R. C., & Stuhlman, M. W. (2004). Conceptualizing risk in relational terms: Associations among the quality of child-adult relationships prior to school entry and children's developmental outcomes in first grade. *Educational and Child Psychology, 21*, 32-45.
- Qi, C. H., & Kaiser, A. P. (2003). Behavior problems of preschool children from low-income families: Review of the literature. *Topics in Early Childhood Special Education, 23*, 188-216.
- Quesenberry, A., Hemmeter, M. L., & Ostrosky, M. (2011). Addressing challenging behavior in Head Start: A closer look at program policies and procedures. *Topics in Early Childhood Special Education, 30*, 209-220. doi:10.1177/0271121410371985
- Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Metzger, M. W., & Solomon, B. (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology, 77*, 302-316. doi:10.1037/a0015302
- Sanetti, L. M., & DiGennaro Reed, F. D. (2012). Barriers to implementing treatment integrity procedures in school psychology research: Survey of treatment outcome researchers. *Assessment for Effective Intervention, 38*, 1-8. doi:10.1177/1534508411432466
- Sanetti, L. M., Dobey, L. M., & Gritter, K. L. (2012). Treatment integrity of interventions with children in the Journal of Positive Behavior Interventions from 1999 to 2009. *Journal of Positive Behavior Interventions, 14*, 29-46. doi:10.1177/1098300711405853
- Sanetti, L. M., Gritter, K. L., & Dobey, L. M. (2011). Treatment integrity of interventions with children in the school psychology literature from 1995 to 2008. *School Psychology Review, 40*, 72-84.
- Sanetti, L. M., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38*, 445-459.
- Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research, 38*, 32-43. doi:10.1007/s10488-010-0321-0

- Schoenwald, S. K., & Hoagwood, K. (2001). Effectiveness, transportability, and dissemination of interventions: What matters when? *Psychiatric Services*, *52*, 1190-1197.
- Southam-Gerow, M. A., & McLeod, B. D. (2013). Advances in applying treatment integrity research for dissemination and implementation science: Introduction to special issue. *Clinical Psychology: Science and Practice*, *20*, 1-13. doi:10.1111/cpsp.12019
- Southam-Gerow, M. A., Weisz, J. R., Chu, B. C., McLeod, B. D., Gordis, E. B., & Connor-Smith, J. K. (2010). Does cognitive behavioral therapy for youth anxiety outperform usual care in community clinics? An initial effectiveness test. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*, 1043-1052. doi:10.1016/j.jaac.2010.06.009
- Sutherland, K. S., Lewis-Palmer, T., Stichter, J., & Morgan, P. (2008). Examining the influence of teacher behavior and classroom context on the behavioral and academic outcomes for students with emotional or behavioral disorders. *Journal of Special Education*, *41*, 223-233. doi:10.1177/0022466907310372
- Sutherland, K. S., McLeod, B. D., Conroy, M. A., Abrams, L., & Smith, M. M. (2013). Preliminary psychometric properties of the BEST in CLASS Adherence and Competence Scale. *Journal of Emotional and Behavioral Disorders*. Advance online publication. doi:10.1177/1063426613493669
- Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2011). Intervention fidelity in special and general education research journals. *The Journal of Special Education*, *45*, 1-11. doi:10.1177/0022466911419516
- Vo, A. K., Sutherland, K. S., & Conroy, M. A. (2012). Best in class: A classroom-based model for ameliorating problem behavior in early childhood settings. *Psychology in the Schools*, *49*, 402-415. doi:10.1002/pits.21609
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. E. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, *61*, 620-630.
- Webster-Stratton, C., Reid, M. J., & Hammond, M. (2001). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start. *Journal of Clinical Child Psychology*, *30*, 283-302.
- Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology*, *33*, 105-124. doi:10.1207/S15374424JCCP3301_11
- Webster-Stratton, C., Reid, M. J., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: Evaluation of the Incredible Years Teacher and Child Training Programs in high-risk schools. *Journal of Child Psychology and Psychiatry*, *49*, 471-488. doi:10.1111/j.1469-7610.2007.01861.x
- Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly*, *28*, 199-209. doi:http://dx.doi.org.proxy.library.vcu.edu/10.1016/j.ecresq.2012.12.002
- Wolery, M. (2011). Intervention research: The importance of fidelity measurement. *Topics in Early Childhood Education*, *31*, 155-157. doi:10.1177/0271121411408621
- Yoder, P., & Symons, F. (2010). *Observational measurement of behavior*. New York, NY: Springer.