

A Dialogue between Treatment Integrity Research and Design-Based Implementation Research

Hao-Jan Luh

A thesis

submitted in partial fulfillment of the  
requirements for the degree of

Master of Education

University of Washington  
2016

Committee:

Leslie Herrenkohl

Todd Herrenkohl

Program Authorized to Offer Degree:

College of Education

©Copyright 2016

Hao-Jan Luh

University of Washington

**Abstract**

A Dialogue between Treatment Integrity Research and Design-Based Implementation Research

Hao-Jan Luh

Chair of the Supervisory Committee:

Dr. Leslie Herrenkohl

Educational Psychology

A growing body of research has focused on implementing innovative practices and designs in educational settings. One field regarding this issue is treatment integrity research under implementation science, in which researchers explore issues around implementing educational practices with fidelity. Another field is design-based implementation research (DBIR). In DBIR, designs and theories are engineered into educational contexts, which are tested and refined through iterative process. Each field has its own advancement as well as challenges. While researchers in both fields endeavor to bridge the gap between research and practice, there are only few discussions between treatment integrity research and DBIR. In this project, overviews about aspects, underlying assumptions, and challenges in these two fields are supplied. Based on the overviews, a comparison is made between treatment integrity research and DBIR. The discussions about the similarities and differences between these research fields further create a platform for a dialogue, which suggests more conversations to foster development in these research fields.

## A Dialogue between Treatment Integrity Research and Design-Based Implementation Research

Education is not preparation for life. Education is life itself.

- John Dewey

Over the past two decades, more researchers have attempted to relate research and practice (Durlak & DuPre, 2008; Mitchell, 2011). One of the research lines addressing this issue is implementation science, which is commonly seen in the field of school psychology as well as other fields related to psychotherapy (Bond, Drake, McHugo, Rapp, & Whitley, 2009; Forman et al., 2013; Prowse & Nagel, 2015). Implementation science focuses on issues relevant to the implementation of evidence-based practices (Durlak & DuPre, 2008; Forman et al., 2013). It is defined as “the scientific study of methods to promote the systemic uptake of research findings and evidence-based practices into professional practice and public policy” (Forman et al., 2013). Among the increasing emphasis on implementation science, researchers have started to pay more attention to issues related to treatment integrity. Treatment integrity has various dimensions in different definitions in research literature (Cochrane & Laux, John, 2008; Durlak & DuPre, 2008; Perepletchikova, Treat, & Kazdin, 2007). Whereas it is an important field as stronger treatment integrity tends to relate to better treatment outcomes, in a review study, it is stated that the relationship between treatment integrity and treatment outcomes is in fact complicated (Sanetti & Kratochwill, 2009). While higher treatment integrity could lead to better treatment outcome, lower treatment integrity might not necessary lead to negative outcomes. To understand how treatment integrity plays a role in implementation science, researchers have produced more research findings via case studies, experimental designs, as well as research reviews in

intervention studies (Bruhn, Hirsch, & Lloyd, 2015; Noell et al., 2014; Perepletchikova et al., 2007; Sanetti & Kratochwill, 2009; Sanetti, Kratochwill, & Long, 2013). With more promising findings, researchers also specify unsolved challenges regarding the complexity of conceptualization, measurement and promotion about treatment integrity (Gresham, 2009; Mowbray, Holter, Gregory, & Bybee, 2003; Noell et al., 2014). These challenges may be addressed with other research approaches related to implementation.

Another line of research that seeks to relate research and practice is design-based implementation research (DBIR) (Collins, Joseph, & Bielaczyc, 2004; Penuel, Fishman, Haugan Cheng, & Sabelli, 2011; Roschelle, Knudsen, & Hegedus, 2009). With design-based implementation research as an educational inquiry, researchers explore the development and refinement of designs and theories about teaching and learning in educational contexts (Collins et al., 2004). Generally, researchers conducting DBIR aim to understand how innovations and related theories operate within different contexts. Like implementation science and treatment integrity research in the field of school psychology, DBIR has received increasing attention in the field of learning sciences as well as other related educational fields over the past two decades (Penuel et al., 2011). To date, one can see the application of DBIR in fields such as developmental psychology, cognitive science, cultural psychology and many other areas (Bell, 2004; Cobb, 2001; Reiser, Tabak, Sandoval, Smith, Steinmuller & Leone, 2011). However, compared to traditional research approaches, design-based implementation science is still quite new. Whereas studies have shown the benefits of DBIR, like treatment integrity research, some challenges remain unresolved in this field (The Design-Based Research Collective, 2003). Similar to treatment integrity research, more input from different research fields may assist DBIR researchers to cope with these challenges.

Both treatment integrity research and design-based implementation research address issues pertinent to the implementation of educational practices, yet there are few bidirectional discussions between DBIR and implementation scientists. In this thesis project, a dialogue between treatment integrity research and DBIR is proposed. It is reasoned that a discussion about the similarities and differences between these two fields may increase more understanding among them. Additionally, such a comparison may provide a platform for a dialogue that could advance improvement in the research and practice of implementing innovations in educational settings. In the first two sections below, overviews about treatment integrity research and design-based implementation research will be provided. In the third section that follows, a comparison between these two fields will be drawn. In the fourth section, potential insights from these fields to each other will be discussed.

### **Treatment Integrity Research**

Treatment integrity is generally defined as to which degree a practice is implemented as expected (Bruhn et al., 2015; Schulte, Easton, & Parker, 2009). Along with the research development in the field, the concept of treatment integrity has broadened from adherence to multiple dimensions, such as competence and program differentiation (Sanetti & Kratochwill, 2009; Schulte et al., 2009). Meanwhile, more measurements and promotion strategies about treatment integrity have been developed along with the increasing research input in this field (McKenna, Flower, & Ciullo, 2014; Sanetti & Kratochwill, 2008). In this section about treatment integrity research, an overview will be offered regarding three focal aspects in the field: the conceptualization, the measurement, and the promotion of integrity. This overview will also list several underlying assumptions in this field as well as challenges treatment integrity researchers face in the literature. It is expected that this section will function as a platform for a dialogue

between treatment integrity research and design-based implementation research in this thesis project.

## **Conceptualization of Treatment Integrity in the Literature**

### **Conceptual Models**

As of today, there are various models for the conceptualization about treatment integrity (Sanetti & Kratochwill, 2009; Schulte et al., 2009). In a review article, Sanetti and Kratochwill (2009) listed dimensions that are included in current treatment integrity research, which indicates the complexity of treatment integrity in various domains. The authors showed that there are at least six conceptual models of treatment integrity in the current literature (Dane & Schneider, 1998; Fixsen, Naoom, Blase´, Friedman, & Wallace, 2005; Jones, Clarke, & Power, 2008; Noell, 2008; Power et al., 2005; Sanetti & Kratochwill, 2009; Waltz, Addis, Koerner, & Jacobson, 1993). While some models overlap on certain dimensions like adherence and competence, other models contain entirely different dimensions, such as feedback and sources (Sanetti & Kratochwill, 2009). In this subsection, several models of treatment integrity conceptualization will be presented to provide a general idea about current conceptualization of treatment integrity. These models are also displayed to show the complexity of treatment integrity conceptualization.

Earlier conceptualization of treatment integrity mainly addresses the concept of adherence, which refers to “the extent to which a therapist used interventions and approaches prescribed by the treatment manual and avoided the use of intervention procedures proscribed by the manual” (Gearing et al., 2011; Schulte et al., 2009; Waltz et al., 1993). In a model proposed in 1993, Waltz and colleagues examined the practice of treatment integrity in psychotherapy research. The authors argued that while measuring adherence in a given treatment, researchers

and practitioners need to discriminate four types of behaviors: 1) behaviors that are both unique and important to a treatment, 2) behaviors that are important but not unique to the treatment, 3) behaviors that are acceptable but are not necessary for the treatment, and 4) behaviors that should be avoided. In addition, the authors also differentiated competence from adherence in the article (Waltz et al., 1993). According to the authors, competence refers to “the level of skill shown by the therapist in delivering the treatment.” The authors emphasized that competence in treatment integrity research does not mean competence for therapies in general but refers to the competence for a certain type of treatment. Additionally, as the authors suggested, competence is not easily measured with measurement about treatment components (Gearing et al., 2011; Schulte et al., 2009; Waltz et al., 1993). Overall, this model suggests that adherence is not a sufficient domain to represent treatment integrity.

Based on the model of Waltz et al in 1993, Dane and Schneider (1998) proposed another model of treatment integrity. In their model, treatment integrity is discussed in three aspects: delivery of a treatment, receipt of a treatment, and enactment of a treatment (Dane & Schneider, 1998; Gresham, 2009; Schulte et al., 2009). In *the delivery of a treatment*, fidelity is composed of adherence, exposure, competence, and program differentiation. In *the receipt of a treatment*, fidelity includes participant’s exposure, understanding, and responsiveness about the treatment. In *the enactment of treatment*, fidelity entails participant mastery and use of the skills learned in the treatment. In this effort, Dane and Schneider not only differentiated treatment integrity for both practitioners and recipients but also added several domains in the conceptualization of treatment integrity. The authors discussed five core domains in their framework, including adherence, quality, exposure, participant responsiveness, and program differentiation (Forman et al., 2013; Power et al., 2005; Sanetti & Kratochwill, 2009). In addition to adherence, in their

model, the authors quantitatively defined exposure as the number or length of implemented sessions. They defined “quality” as the practitioners’ skills and enthusiasm in implementing a practice. “Participant responsiveness” was defined as how well participants are involved in the practice and how they view the practice. “Program differentiation” refers to different components that are implemented in different programs. In other words, it is examined whether components specific to one practice are implemented and whether components that are not related to the practice are eliminated.

The above two models represent several common domains in current treatment integrity research, but not all of them, as it involves only direct service between a practitioner and a recipient, e.g. a teacher and a student. Looking at treatment integrity from another perspective, Noell (2008) focused on the relationship between a consultant and consultee in indirect services, e.g. the relationship between a school psychologist and a teacher. Instead of looking into the above dimensions such as adherence or competence, the researcher examined two concepts, consultation procedural integrity and treatment plan integrity (Forman et al., 2013; Schulte et al., 2009). Noell stated that when a mediator-based intervention implementation is involved, consultation procedural integrity as well as treatment plan integrity should also be incorporated into the conceptualization of treatment integrity. The questions about treatment integrity then were expanded from practitioner-recipient(s) to consultant-consultee-recipient(s) in current literature.

After reviewing the literature of treatment integrity, Sanetti and Kratochwill (2009) listed treatment integrity dimensions within the current conceptualization models, including, exposure, dosage received, interventionist adherence, interventionist competence, participant adherence, participant responsiveness, program differentiation, quality of implementation, unique and

essential behaviors, essential but not unique behaviors, acceptable but not necessary behaviors, proscribed behaviors, source, destination, communication link, feedback, influence, consultation procedural implementation, and treatment plan implementation. Among these multiple dimensions, the authors listed six dimensions that are usually shared in some models, including adherence, exposure, quality, participant responsiveness, program differentiation, and competence. Instead of separating quality and competence, these two dimensions are seen interchangeable in the review by Schulte et al. (2009), indicating the need of more clarification of these dimensions.

### **Relations among Dimensions in Treatment Integrity Research**

As the frameworks above provide basic conceptualization about treatment integrity, more recent studies have looked into the relationship among the shared dimensions in the frameworks, especially adherence and other dimensions (Perepletchikova, Treat, & Kazdin, 2007). One type of discussions focuses on the relations between adherence and competence. These studies point out that, conceptually, it is more of a unidirectional relationship as competence could assume adherence but no vice versa (Perepletchikova, Treat, & Kazdin, 2007). However, studies have presented mixed results about the relationship between adherence and competence, showing that the correlation between these two domains in past studies span from not significant to highly correlated (Perepletchikova et al., 2007).

Another type of discussions focuses on the relationship between adherence and program differentiation (Perepletchikova et al., 2007). As Dane and Schneider (1998) suggested, when they separated adherence and program differentiation based on the model by Waltz et al. (1993), these two dimensions can be seen different yet related. In other words, Perepletchikova et al.

(2007) argued that while a practitioner adheres to the instruction of a manual or checklist about a certain practice, program differentiation will be ensured, as behaviors that are required by one practice not the other will be seen.

Yet another discussion focuses on issues related to adaptation and flexibility of a treatment or practice (Harn, Parisi, & Stoolmiller, 2013). According to several studies, there is rarely 100% of fidelity in a treatment, which also suggests a ceiling effect regarding fidelity across treatments or programs (Gresham, 2009; Sanetti & Kratochwill, 2009; Schulte et al., 2009). Treatment integrity that is higher than eighty percent is generally considered as high treatment integrity in current literature (Keller-Margulis, 2012; Perepletchikova, 2005). However, a question remains: To what degree should the components be implemented in a treatment (Fryling, Wallace, & Yassine, 2012; Noell, Gresham, & Gansle, 2002; St. Peter Pipkin, Vollmer, & Sloman, 2010)? It is also noteworthy that the discussions above mainly address adherence and adaptation, not other dimensions in the literature. All these discussions open up discussion about flexibility and adaptations (Durlak & DuPre, 2008; Forman et al., 2013; Sanetti & Kratochwill, 2009). Questions such as how high the treatment integrity should be as well as under what circumstances will treatment integrity and adaptation collide at the expense of treatment effects remain unsolved (Elias, Zins, Graczyk, & Weissberg, 2003).

The above discussions present frameworks about treatment integrity in the current literature. These frameworks not only showcase the complexity of treatment integrity, but also establish a foundation for further research on the measurement and promotion of treatment integrity. In the sections that follow, current measurement and promotion strategies of treatment integrity will be respectively discussed.

## Measurement of Treatment Integrity

Researchers have proposed several ways to measure treatment integrity, including direct observation, self-assessment, permanent products, hybrid methods and other methods such as video-taping. In direct observation, treatment integrity is measured as behaviors pertaining to critical components are observed (McKenna et al., 2014). It is one of the most common methods to measure the components during the implementation of an evidence-based program (Jeffrey, McCurdy, Ewing, & Polis, 2009; Sheridan, Swanger-gagne, Welch, Kwon, & Garbacz, 2009). However, researchers have also pointed out issues related with this method (McKenna et al., 2014; Sanetti, Chafouleas, Christ, & Gritter, 2009; Sheridan, Swanger-gagne, Welch, Kwon, & Garbacz, 2009). One of the drawbacks of this method is that implementers may act differently when they are observed, and observers may not fully capture how a program is really implemented by the implementers. Another drawback is that it is a resource intensive method. To use direct observation for treatment integrity, it requires more funding and labor to measure whether the implementers are implementing a program with fidelity.

In addition to direct observation, researchers have also proposed other methods for treatment integrity measurement (Keller-Margulis, 2012; McKenna et al., 2014; Sanetti et al., 2009). Through self-report, treatment integrity can be measured in multiple ways, including daily patient logs/diaries, retrospective patient reports, and 24-hr recall interviews (Riekert, 2006; Sanetti et al., 2009). Compared with direct observation, self-assessment provides more flexibility for treatment integrity measurement (Sanetti et al., 2009). However, these methods also have drawbacks. For example, Sanetti et al. (2009) points out that implementers using self-assessment for treatment integrity may tend to overestimate adherence in implementation, which may undermine the data collected in the practice. Some programs include permanent products that can

be used to document treatment integrity. These permanent products are usually created for the intervention activities, such as tokens, charts, and self-monitoring sheets. that are parts of the treatments (Mckenna, Flower, & Ciullo, 2014; Sheridan et al., 2009). However, not all practices generate permanent products to measure treatment integrity dimensions. (Perepletchikova et al., 2007; Sanetti et al., 2009). Given the drawbacks of above methods, researchers propose the use of hybrid methods, methods that incorporate the above methods for documenting treatment integrity. Overall, it is supposed that these methods may provide rich information about whether a practice is implemented with fidelity. However, as argued by Sheridan et al. (2009), it remains challenging to understand and interpret data obtained with different methods.

### **Promotion of Treatment Integrity**

In addition to measurements for treatment integrity, several strategies have been developed to increase treatment fidelity in implementing educational practices (Sanetti & Kratochwill, 2009). One of the current promotion methods that has been proven effective in multiple studies is performance feedback, in which a consultant provides feedback about students' learning performance and treatment fidelity to the teacher on a regular basis about the implementation (Forman et al., 2013; Keller-Margulis, 2012; Noell et al., 2005; Sanetti & Kratochwill, 2009). In this type of consultation, the consultant provides direct feedback about the teacher's program implementation. The consultants can also use goal setting and graphs to promote the teacher's understanding about his or her progress and can thereby promote treatment fidelity (Noell et al., 2005).

Directed rehearsal along with negative reinforcement and directed rehearsal is also used to increase treatment integrity. In this method, teachers are asked to practice the implementation

skills if treatment integrity is not as high as expected on a more regular basis (DiGennaro, Martens, & Kleinmann, 2007; Gross, Duhon, & Doerksen-Klopp, 2014; Ward, Johnson, & Konukman, 1998). For example, in one study, Ward, Johnson, and Konukman (1998) focused on certain behaviors that were incorrectly implemented by four teacher participants in a physical education course. The researchers established a negative reinforcement by telling the participants that if an error was made in implementing certain behavior in the lesson, the participant would need to rehearse the behavior 10 times before leaving. The researchers found that teachers exhibited 100% correct implementation of the steps after directed rehearsal combined with negative reinforcement was used. In other words, it is indicated that such a strategy can increase treatment fidelity as well. Sometimes, negative reinforcement contingency is also used with performance feedback to increase treatment fidelity along with this consultation strategy (DiGennaro, Martens, & Kleinmann, 2007; Sanetti & Kratochwill, 2009).

Power et al. (2005) also proposed a partnership model to increase treatment fidelity in the implementation of evidence-based practices. In this model, Power and colleagues focus on the collaboration between researchers and stakeholders. The researcher chooses the essential components of an evidence-based program and the practitioner chooses strategies to achieve these component in implementation. The goal is not only to address acceptability of implementation but also to engage teachers in the implementation.

During implementation, other strategies are also used to address treatment fidelity. One of the methods is implementation planning (Keller-Margulis, 2012; Sanetti & Collier-Meek, 2014; Sanetti, Collier-Meek, Long, Kim, & Kratochwill, 2014; Sanetti & Kratochwill, 2009). In this method, a consultant works with a teacher to develop a strategy for monitoring fidelity in a structured manner. Using implementation planning protocol guides, the consultant could help the

teacher to identify potential barriers and facilitators in implementing an evidence-based program. The steps of implementing the program are listed. The teacher is encouraged by the consultant to discuss possible modification for contextual fit. Questions about the implementation plan will be used to document potential strategies to make plans to address the barriers and facilitators together. In concluding the consultation, the consultant offers the teacher a summary of the session.

In sum, all these strategies have been used to promote treatment integrity and have shown effectiveness. Future research is encouraged to replicate the strategies and examine their effects across different practices and settings.

### **Examples of Treatment Integrity Research**

Researchers mainly use two research methods to explore treatment integrity: They use literature reviews to document how treatment integrity is approached and its relation with outcomes (Bruhn et al., 2015; Cochrane & Laux, John, 2008; Noell et al., 2014; Perepletchikova et al., 2007; Sanetti, Dobey, & Gritter, 2012). They also use experimental designs to test how treatment integrity can be promoted in various studies. To provide a clearer understanding about this field, two examples are provided about treatment integrity research.

One example of treatment integrity research is a literature review conducted by Gresham and Gansle (1993). In this review, the authors looked into 181 experimental studies about behaviorally based interventions between 1980 and 1990. They examined aspects such as whether treatment integrity was measured and reported, to which extent the practices were implemented with fidelity, how the practices were operationally defined, and the effect sizes in the studies. The authors also examined the correlations between treatment integrity and treatment

outcomes, which was moderate positive. At the end of the review article, the authors reiterated the significance of treatment integrity measurement and promotion as well operational definition of treatments. As one of earlier literature review articles relevant to treatment integrity, this review was conducted to explore the importance of treatment integrity in research and practice.

Another example of treatment integrity research is provided by Noell and colleagues (2005). They examined the effectiveness of briefly weekly interviews, weekly interviews with a commitment component, and performance feedback on treatment integrity in behavioral consultation. A total of 45 dyads of elementary school teachers and students participated in the study. Students were referred for interventions because of academic or/and behavioral issues. In an experimental analysis with factors of time and condition, the research showed that performance feedback led to stronger treatment integrity and better treatment outcomes compared with the other two groups. Researchers also found that the effects between weekly interview group and the commitment group were statistically different. In addition, they also found positive correlations between treatment integrity and students' behavioral outcomes.

As the multiple forms of treatment integrity research exceed the scope of this project, these two example are provided to briefly illustrate how treatment integrity is approached with literature reviews and experimental designs in this field. Along with the above literature review, these examples create a window for researchers to look into some underlying assumptions and challenges in the field.

### **Underlying Assumptions in Treatment Integrity Research**

In this subsection, two underlying assumptions in current treatment integrity research will be discussed.

In treatment integrity research, the implementation of practices in educational settings is seen as a unidirectional communication process. As students receive educational practices from teachers, teachers also receive instructions from intervention materials and consultants. Compared to consultants and program designers, teachers and students have less weight in on what and how to implement. This underlying assumption is reflected in the conceptualization models. Most domains in the models, such as adherence, competence, and program differentiation, are unilaterally framed by program designers and researchers who evaluate treatment integrity. Program designers and researchers act as active speakers, while teachers and students play the role of passive listeners. Such an assumption can also be embodied by the promotion approaches above. When either performance feedback or directed rehearsal is adopted to foster treatment integrity (Fallon, Collier-Meek, Maggin, Sanetti, & Johnson, 2015), the role of teachers is passive, as they are regarded as the executors not designers of the implementation. Even in a partnership model (Power et al., 2005), teachers might not be able to contribute to the content of practices that are implemented as the components of the practices are mostly predetermined. This underlying assumption may be associated with the fact that in fields such as special education and school psychology that value treatment integrity, interventions are greatly informed by behaviorism and social learning theories (Forman et al., 2013). While the theories address how the context can induce behavioral changes, teachers' and students' subjectivity is not the main focus. Therefore, treatment integrity is mostly conceptualized in a unidirectional manner.

In addition, it is suggested in treatment integrity research that quantitative data generated from experimental designs are useful to understand how treatment integrity plays a role in implementation. For example, viewing quantitative adherence as the main indicator of treatment

integrity, researchers may choose experimental designs to test the influence of either adherence on treatments or promotion methods on adherence. Despite more discussions about qualitative characteristics in the field of treatment integrity research, such as competence and enthusiasm, experimental designs are assumed to represent the nature of treatment integrity in the implementation of educational practices.

### **Challenges in Treatment Integrity Research**

Although treatment integrity research has greatly developed over the past decades, more questions and limitations also have surfaced along with the development, leading to challenges to treatment integrity researchers (Sanetti & Kratochwill, 2009). In this subsection, these challenges will be discussed.

One of the biggest challenges researchers face is the multiple dimensions across various models (Nelson, Cordray, Hulleman, Darrow, & Sommer, 2012; Sanetti & Kratochwill, 2009). Given the complexity of treatment integrity conceptualization, Schulte and colleagues (2009) argued that multiple dimensions are important in treatment integrity. However, despite the importance of multiple dimensions in treatment integrity research, problems also emerge as there is a lack of understanding regarding what comprise the dimensions in treatment integrity research. As stated by Sanetti and Kratochwill (2009), it is important to develop an agreement about dimensions in a model for treatment integrity conceptualization. In other words, it is stated that without a consistent model of treatment integrity, researchers may indicate various domains in research which causes difficulties in literature conversations. For example, adopting Waltz et al.'s model, one study may measure only adherence and competence as treatment integrity; another study may use Dane and Schneider's model and also include participant responsiveness

and program differentiation. Moreover, the emerging measures and promotion strategies may not fully capture important treatment integrity concepts either, leading to inaccurate data collection and undermined conclusions. Researchers in the field use various dimensions to capture treatment integrity but it also causes problems regarding research communication and practical measurement and promotion.

Another challenge treatment integrity researchers face is the issue about adaptation and flexibility (Durlak & DuPre, 2008; Sanetti & Kratochwill, 2009). To date, it is still under debate to what extent a treatment can be adapted and remain efficacious. Researchers have argued that adaptation might be inevitable and even beneficial for treatments (Durlak & DuPre, 2008; Sanetti & Kratochwill, 2009), while results have been reported regarding the complexity of adaptation issues (Huey & Polo, 2008; Lau, 2006; Weisz, Doss, & Hawley, 2005). As pointed out in the above section, in treatment integrity research, teachers and students are generally assumed as recipients of educational practices. With researchers and designers being the speakers and teachers and students as the listeners, the adaptation of practices may be more fixed because practices are usually decided by researchers and designers. Such a fixed orientation may lead to the difficulties of exploring adaptation, as teachers and students may contribute to the reasons and process of adaptation. Also, as shown, current treatment integrity research mainly advances on experimental designs and literature review, which may limit the findings about adaptation issues as well. To address the challenge of adaptation and flexibility, it might be helpful to explore how both teachers and students play a role in adaptation with multiple research methods.

In addition to the inconsistency of treatment integrity conceptualization, the measurement of treatment integrity is also challenging for both researchers and practitioners (McLeod, Southam-Gerow, & Weisz, 2009). While there are various methods to measure treatment

integrity, each method has its pros and cons. Self-report method may be the least costly for treatment integrity measurement, but the credibility of self-report can be problematic. Other methods also have their own advantages and disadvantages. Hybrid methods provide abundant information, but it may require more resources to collect data with such methods. When resources and time are taken into considerations, data collection about treatment integrity may also be thorny issue (Schulte et al., 2009). Future research exploring measurement issues is thus valuable.

Last but not least, there is also a challenge in maintaining and promoting treatment integrity in both research and practice (Gross et al., 2014). First, since the dimensions vary across models, the question is not only what dimension researchers should measure but also what dimensions researchers should maintain and promote. Most promotion strategies focus only on one dimensions such as adherence. Domains such as competence and program differentiation are sometimes overlooked. Second, in spite of a growing body of research in promoting treatment integrity (Fallon, Collier-Meek, Maggin, Sanetti, & Johnson, 2015; Sanetti, Kratochwill, & Long, 2013), the methods need to be examined across more contexts. Third, while some promotion methods have proven to be effective in promoting treatment integrity, it is not examined whether the effectiveness is sustainable (DiGennaro et al., 2007). In sum, promotion strategies are required to be examined across contexts and over time to ensure scalability and sustainability (Sanetti, Kratochwill, & Long, 2013).

This subsection delineated current challenges in treatment integrity research. Future studies are encouraged to address these challenges. In the below section, design-based implementation research will be introduced. It is suggested that insight from design-based

implementation research can offer some input to address the challenges in treatment integrity research, and vice versa.

### **Design-Based Implementation Research**

Design-Based Implementation Research (DBIR) is an approach of systematic inquiry in education (Bell, 2004). Researchers use this approach to systematically investigate learning and teaching within educational context (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003). Through this approach, theories and innovative practices are developed and refined to improve teaching and learning that are contextualized (Sannino, Sutter, & Engeström, 2011). Over the past two decades, the research literature has seen the development of this approach in learning sciences and other educational research fields (Collins et al., 2004; Penuel et al., 2011). As discussed by Cobb and colleagues (2003), design-based implementation research has been used in various settings, including one-on-one teacher-student experiments, classroom experiments, preservice teacher development studies, in-service teacher development studies, and school restructuring experiments. In the following subsections, the goals and characteristics of DBIR will be discussed. An example study will also be provided to describe how this approach is used in educational research. In the last subsection, challenges facing this approach will be discussed. As the overview about treatment integrity provided in the first section, this section will serve as a part of platform for a conversation between treatment integrity research and design-based implementation research.

#### **Characteristics of Design-Based Implementation Research**

The development of design-based implementation research traces back to the 1990s, when Brown (1992) and Collins (1992) first initiated the idea, “design experiment,” in

educational research. They intended to address several needs related to the study of learning, including exploring learning in context with theories, analyzing learning in real-life, investigating learning with broader measures, and generating research findings from formative research (Collins et al., 2004). To address these needs, researchers have developed several features within DBIR (Collins et al., 2004; Russell, Jackson, & Frank, 2013).

As specified in the literature (Collins et al., 2004; Penuel et al., 2011), design-based implementation research has been proposed with rationales and corresponding characteristics in contrast to traditional research in education. First, this approach is a design science instead of an analytic science (Collins et al., 2004). Collins (1992) defines analytic sciences as seeking to explore the explanation about phenomenon and design sciences as attempting to explicate how different designs impact learning in different situations. Such a distinction is important because testing innovations with an analytic science lens may not provide information on how innovations work in a certain context. As a design science, design-based implementation science is practical in that researchers “engineer” designs in learning contexts and test these designs in a certain context (Cobb et al., 2003).

Second, design-based implementation research aims at sustainable and scalable development of a design (Cobb et al., 2003). To achieve this aim, researchers seek to understand and elucidate the design through the lens of theories (Cobb et al., 2003; Shavelson, Phillips, Towne, & Feuer, 2003). That is to say, DBIR is itself also theory-oriented (Cobb et al., 2003; Shavelson et al., 2003). In DBIR, researchers develop and refine theories in different contexts through systematic inquiry (Penuel et al., 2011). In addition to test the effectiveness of practices, design research is also conducted to examine theories behind these practices. The theories examined in DBIR should be “domain-specific” (Cobb et al., 2003). For example, researchers

using DBIR to investigate a social-emotional learning program in educational settings should employ suitable social emotional learning theories and refine it during the research. In doing so, the researcher will gain understanding about the design as well social emotional learning per se. This characteristics of design research enables the communicability and scalability of an innovation across studies and contexts.

Third, DBIR has been developed to evaluate environmental support in educational research. Collins et al.(2004) argued that researchers and practitioners should not narrow their focus on the knowledge and skills students learn. Instead, education should also motivate students to be “expert learners”. It is important to understand the kind of support an environment can provide to students. Correspondingly, in contrast to traditional research methodology, measures in design-based research are used to evaluate whether learning environments can prepare students to be creative, courageous, and productive in new thoughts and products (Collins et al., 2004). In brief, DBIR also measures environmental support to students beyond student performance.

Fourth, researchers stated that DBIR methods are generally iterative (Cobb et al., 2003; Russell et al., 2013; Shavelson et al., 2003). As Cobb et al. (2003) mentioned, design research is both prospective and reflective. It is prospective as a design is placed with conjectures in a context. It is reflective as design research involves “conjecture-driven tests.” Original conjectures about a design is tested during the implementation of the design. As an iterative process, DBIR incorporates an extended cycle of design-analysis-redesign. Theory-based designs are tested and refined in a progressive manner (Collins et al., 2004). A design is first placed and tested in a real-life educational setting. Based on the performance of the design, it is revised.

Such a process enables researchers and practitioners to improve the theory and practice involved in the implementation of innovative programs (Collins et al., 2004).

Fifth, to achieve the above goals, design-based implementation research should be collaborative (Cobb et al., 2003; Shavelson et al., 2003). That is to say, in this approach, multiple stakeholders participate in teams to address issues. The teams may involve teachers, researchers, students, community members as well as school and district leaders. In the collaboration among all stakeholders, different professionals and various knowledge are involved to address the problem in the educational context (Coburn & Penuel, 2013). As it requires negotiations among the team members to identify and address the issues, team members should be cognizant of the power and authority dynamic when they collaborate to identify and deal with the persistent problems.

These features and goals comprise the central ideology and methodology behind design-based implementation research. As mentioned, while DBIR takes different forms in various studies (Bell, 2004; Fishman, Penuel, Allen, Haugan Cheng, & Sabelli, 2013; Zheng, 2015), many features in DBIR are commonly seen in these studies to systematically test and refine both theories and practices in educational settings. To further exemplify how design-based implementation research is carried out in education research, a research study is delineated below.

### **Example of Design-Based Implementation Research**

Design-based implementation research takes various forms in educational research. In this section, an example is provided to showcase how researchers use DBIR to engineer a design and refine it in an educational context.

In one study, researchers investigated how to integrate student teachers' practice experience and theories they learned in teacher education (Tigchelaar & Korthagen, 2004). The authors depicted types of difficulties teacher educators face when they teach students who just had teaching practice. Teacher educators may attempt to link theories with various experience students have after practicum but find it challenging. The difficulties form a common and constant issue in teacher education. To address this issue, Tigchelaar and Korthagen first examined theory frameworks around teacher behavior, including conscious and unconscious behavior, Gestalts, and the behavior of student teachers and experienced teachers. In examining how these theories would work in teacher education, the researchers initiated three cycles of problem analysis, including design, evaluation, and revision. In phase one, methods based on the above theories were provided to students and experienced educators. While evaluation was partly conducted in phase one, in phase two the authors use more evaluation methods such as verbatim reports and logbook reflection to evaluate students' learning process. In phase three, the researchers revised the approaches given prior evaluation results. Based on the revision, the researchers then started a new cycle to test the revised approaches, which were followed by another cycle of revision and more detailed steps for integrating student teachers' experience and theories learned in class.

The above research study exemplified several characteristics in DBIR, such as iterative process and the guidance of theories. As DBIR takes many forms in educational research, it is important to note that characteristics adopted in studies may vary. Nevertheless, this example is informational for discussions about the underlying assumptions and challenges in design-based implementation research, which will be presented below.

### **Underlying Assumptions of DBIR**

In this subsection, two underlying assumptions in design-based implementation research will be discussed.

First, in design-based implementation research, it is assumed that multi-directional communication among multiple stakeholders is necessary for engineering designs. The characteristics in DBIR, such as iterative process, collaborative partnership, sustainability all assume there is good communication among researchers and practitioners. To begin with, as the design should be iteratively engineered and tested, constant input from teachers, principals, and other participants, multi-directional input is necessary, which both requires mutual understanding. Also, it is assumed that a design may not be scalable or sustainable as multiple information can be missing, since lack of trust may interfere the efforts to scale up and sustain the design (Klingner, Boardman, & McMaster, 2013). As such, DBIR suggests the importance of multi-directional communication to develop and refine a design.

Second, relevant to the above assumption is that in DBIR, everyone in the implementation of a design is a learner. It is assumed that the design may need to be refined in the context, which requires every stakeholder to constantly and actively learn to respond to the process. For example, as stated by Brown (1992), while a design is implemented, both teachers and students should be potential active learners. Teachers are seen as role models of learning and should respond to students' process based on students' needs. Under such a less predetermined and more flexible circumstances, it is assumed that every one may need to actively adjust their roles to make the design scalable and sustainable.

The third underlying assumption in design-based implementation research is that a design needs to be theory-based, not simply evidence-based. It is supposed that learning theories not

only make a design effective but also make it sustainable when implementation is contextualized. Such an assumption is associated with the emphasis of theories in the field of learning sciences. As Barab (2004), “Learning Sciences focuses on testing and advancing learning theory of which the resultant design work provides a fertile ground through which to embed and develop evidence-based claims about particular assumptions regarding learning.” For researchers who use DBIR, it is inevitable to examine and refine theories when a design is implemented. Overall, it is assumed in DBIR that to engineer a design, data generated from experiments are insufficient. The design has to be theory guided to be effective and sustainable.

### **Limitations and Challenges in Design-Based Implementation Research**

Current literature provides discussions about challenges that face design-based implementation research. First, as Collins et al. (2004) suggested, researchers conducting DBIR have to deal with extensive amount of data. Data collected from real-life learning settings can be overwhelming as many variables are involved. To capture the design in context, researchers need to use mixed methods to collect data, such as video records (Zheng, 2015). The analysis of these data usually takes more time and resources than researchers can afford (Collins et al., 2004). As Cobb et al. (2003) suggested, to conduct retrospective analyses, it is important to analyze longitudinal datasets in a systematic manner. As data collection needs to be organized with great effort, the difficulty to conduct DBIR can be discouraging. As a lot of data will remain unanalyzed (diSessa & Cobb, 2004), efficiency may also be an issue in this approach.

Second, DBIR is more time-consuming compared with traditional research. As a design science (Collins, 1992), it takes time for researchers to examine and refine designs and theories involved in the research. Also, to conduct iterative research, researchers and practitioners need to

meet regularly and discuss the development and refinement of innovations, which also requires more time to conduct the research. According to Penuel et al. (2011), such time-consuming process narrows the opportunities for researchers to secure funding compared to research that is short-term and provides high treatment integrity data.

Third, as collaboration is required in design-based implementation research, the coordination can be challenging for team members. The difference among team members is likely to not only generate new ideas but also induce tension among them (Coburn, 2003). Coburn (2003) argues that when reform is initiated in breadth and depth, tension might emerge among team members, thus impact team members' ability to bring out the innovation. Moreover, such a tension may increase when an innovation challenges team members' current views and behaviors or when the scale of an innovation is larger. It is suggested that a well-orchestrated coordination is necessary to deal with this issue (Coburn & Penuel, 2013).

### **Differences and Similarities between Treatment Integrity Research and Design-Based Implementation Research**

The above overviews delineate the features in both treatment integrity research and design-based implementation research. To provide further information about these two fields as well to initiate a dialogue between the two, a comparison will be made in the following sections.

One salient difference between these two fields concerns levels of specificity. Current treatment integrity research focuses more on specific domains, especially adherence (Sanetti & Kratochwill, 2009), whereas design-based implementation research entails more general aspects of an innovation, including its depth, sustainability, spread, and ownership (Coburn, 2003). In other words, whereas treatment integrity researchers investigate the specific impact of treatment

integrity on the outcomes of a practice, DBIR places emphasis on the development of an innovation and learning environment in a more general sense. Penuel, Phillips, & Harris (2014) present an analysis of teachers' implementation of a curriculum from both an integrity perspective and an actor-oriented perspective. The integrity perspective and actor-oriented perspective in the analysis reflect the difference of focus specificity in current treatment integrity research and design-based implementation research. The researchers point out that from an integrity perspective, transparent and clear guidance related to the practice should be provided to teachers for them to follow during the implementation of a practice. In contrast, from an actor-oriented perspective, how teachers make sense of the practice and environment is the focus of the research. Due to the different levels of focus, different methodologies are also involved (Penuel et al., 2014).

Along with the focus on different domains, another difference between treatment integrity research and design-based implementation research is the target in these two fields. Treatment integrity research focuses on observable key dimensions in an educational practice (Greenwood et al., 2009). In contrast, DBIR incorporates more contextual considerations as well as practitioners' attitudes and beliefs in the inquiry (Gutierrez & Vossoughi, 2010). For example, in current treatment integrity research, the teacher's behaviors related to the key treatment component are usually the focus. In DBIR, it is explored whether contextual support is provided to the teacher and how the teacher considers his or her role in implementing the practice (Penuel et al., 2014).

Another difference can be seen between treatment integrity research and design-based implementation research. These two fields differ on the requirement of time and resources. Whereas DBIR is more time-consuming (Penuel et al., 2011), treatment integrity research may

take less time to conduct. For example, in certain treatment integrity approaches, it is not necessary for researchers to meet with teachers to glean data about treatment integrity. Data might be collected with self-observation or permanent products (Sanetti & Kratochwill, 2009). In contrast, because collaboration and iteration are essential in design-based implementation research (Collins et al., 2004), researchers and practitioners need to meet regularly to address issues encountered in the implementation of innovations. However, as discussed in the aforementioned sections, such difference is not fundamental. Depending on the form of treatment integrity research, different levels of resources and time may be required. For example, in a partnership model along with hybrid methods to measure treatment integrity, researchers might need to spend more time to promote treatment integrity, collect and analyze data from the implementation.

As there exist different focuses between these two fields, some degrees of overlap can also be discovered. To begin with, while design-based implementation research values the collaboration among researchers and practitioners, there is also an increasing emphasis on such a collaborative relationship in treatment integrity research. As introduced by Power et al (2005) proposed a partnership model to measure and promote treatment integrity. They encouraged researchers to provide insight on treatment components and practitioners to provide feasible strategies to implement these components. Implementation planning as a strategy also includes more dialectic inputs between researchers and practitioners into educational practices (Sanetti et al., 2014). The rationale behind both a partnership model and implementation planning that involves both researchers and practitioners is to some degree shared by DBIR, as researchers underscore the importance of collaboration in this field. It is important to note that these two

fields both face a challenge as such a collaboration or partnership requires more resources and efforts.

In addition, researchers in both fields seek to conduct research with multiple methods, leading to a huge amount of data for analysis. Treatment integrity researchers encourage more use of hybrid methods in collecting data about treatment integrity (McLeod et al., 2009). Researchers using design-based implementation research also underscore the importance of mixed methods to obtain more information about a design in context. Both fields recognize the benefits of using more than direct observation and self-report. Moreover, as McLeod et al. (2009) suggested in a commentary article, treatment integrity researchers may capture more contextual factors, such as organizational climate in their studies, which seems to reflect the emphasis of design-based implementation research to some extent. However, it is noteworthy that while it has its merits to collect data from multiple resources, both fields face challenges in that it requires more resources to collect and analyze these amount of data.

In this section, the similarities and differences between treatment integrity research and design-based implementation research are discussed. As introduced in the beginning of this thesis project, such discussions may be utilized to address challenges in each field.

### **Mutual Improvement and Future Challenges**

After comparisons are made between both treatment integrity research and design-based implementation research, can these two fields shed some light on each other in the implementation of educational practices? Given the discussions in the above sections, the answer to this question is yes, at least partially in the form that limitations in each field can be addressed

by bringing in partial elements from the other field. Meanwhile, there are shared challenges that face both fields in the literature.

Some limitations in treatment integrity research may be addressed from the perspective of design-based implementation research. One of the main issues in treatment integrity is that there is an inconsistency of conceptualizations of treatment integrity (Sanetti & Kratochwill, 2009). This lack of agreement on treatment integrity conceptualization may result from a lack of overarching guiding theory about treatment integrity. While there are multiple conceptualization models, these models have rarely been systematically theorized in implementation science. In this sense, treatment integrity research may benefit from a theory guiding the models of treatment integrity. In other words, researchers can aggregate the current models into an overarching guiding theory. Since treatment integrity is part of the implementation process, researchers should not only develop the theory but also examine the theory with data evidence in practice. In this case, some features in design-based implementation research may be utilized by treatment integrity researchers. First of all, DBIR is theory-oriented and is conducted through an iterative process. In this process, researchers can test and modify theories in practice and continually refine these and alternative theories with ongoing data. Treatment integrity researchers seeking to examine a guiding treatment integrity theory can refine their theories through this iterative process. They can test whether the theory they propose fits well with outcomes and which models work well when put into practice. Unlike traditional research methodology, the iterative process provides opportunities for treatment integrity researchers to develop a guiding theory about treatment integrity through both research and practice.

Another limitation that can be addressed is that treatment integrity literature faces challenges when contexts are taken into considerations. As researchers have stated, the

importance of different dimensions may vary across different practices and even vary across stages in one practice (Schulte et al., 2009). In addition, the relation between fidelity and adaptation also differ in various practices (Schulte et al., 2009). Design-based implementation research can be employed to address this issue as the approach itself is not only innovation-specific but also highly contextualized. These two features in DBIR allow researchers to examine theories within different contexts and explore how theories interact with the contexts. Along with the iterative process, the interaction can be dynamically captured with this approach. Such an advantage of design-based implementation research can be employed to explore the issue of fidelity and adaptation. Treatment integrity researchers who seek to evaluate treatment integrity dimensions in different contexts can collaborate with researchers using design-based implementation research. Working together, treatment integrity researchers can examine to what extent dimensions such as adherence and competence in implementing a design within a specific context. Moreover, since DBIR involves collaborative planning and regular meeting, the importance of treatment integrity dimensions can be examined across stages. The iterative process of modifying a design also provides opportunities to investigate the relationship between fidelity and modification. Therefore, features such as innovation-specificity, contextualization, and iteration in design-based implementation can be used to address current issues in treatment integrity literature.

While design-based implementation research can be used to address several issues in treatment integrity research, the advancement in treatment integrity may also shed some light on DBIR. One benefit treatment integrity research can provide to design-based implementation research is that it points to dimensions that might be important for researchers to measure and consider. As discussed, one challenge DBIR faces is the amount of data for collection and

analysis. Due to the amount of data researchers usually have to discern what kind of data they need to analyze first and what kind of data they do not need to use at the moment. Treatment integrity research is beneficial in this case as it points out what to look at first when an innovation does not generate expected outcomes. Although there is not consensus among treatment integrity conceptualization, several common dimensions as well as dimensions that are less discussed may potentially indicate the reasons why an innovation is not working as planned. As Penuel et al. (2004) suggested, analysis through treatment integrity research can provide more specific information for researchers in design-based implementation research to work on. Also treatment integrity research may indicate what to examine when issues occur. For example, measurement developed in treatment integrity research may be used in DBIR. While a design is not working as expected, in addition to immediately revise it, researchers may look into treatment integrity data and decide whether treatment integrity is a reason behind the issues. Combined with other data gleaned from the research process, researchers could also use the treatment integrity data as further inquiry to support teaching and learning in the context.

Treatment integrity research can also provide instant information in design-based implementation research for funding and communication purposes. As stated by Penuel et al. (2011), the time-consuming feature in DBIR can be problematic when it comes to funding. The problem is that to conduct iterative research, researchers may not be able to provide instant information for funding purposes. As Penuel et al (2011) posited, implementation fidelity instead of mutual adaptation is usually more underscored by funding agencies. In this case, treatment integrity research may benefit researchers using DBIR. With the specificity of treatment integrity measures and promotion, researchers using DBIR can collect data about treatment integrity. Although data may not present the overall picture of the research, researchers can use the instant

information about treatment integrity for funding purposes. In other words, the direct, specific documentation can be used for both communication and funding purposes. This is important, as the complexity of DBIR may create barriers between researchers and other professionals who are not involved in the study.

Although both fields can benefit from each other to some degree, they both face challenges that remain unsolved. First of all, both fields place an increasing attention on the collaboration among multiple stakeholders. As discussed above, while collaboration or partnership in these two fields may be beneficial, challenges also come along. The issue is that it needs good orchestration to involve multiple stakeholders when a practice or design is placed into a context (Coburn & Penuel, 2013; Power et al., 2005). More resources and time are needed to establish such a collaboration among multiple stakeholders. Second, both fields use mix or hybrid methods to collect data, which also requires a lot of resource and time. These two limitations are pointed out in the literature but remain unsolved.

### **Conclusion**

This project seeks to initiate a dialogue between treatment integrity research and design-based implementation research. The gradual developments of both fields indicate the increasing efforts to bridge the gap between research and practice within these two fields. However, there remains scarce communication between these two fields. On the one hand, treatment integrity research has focused on specific dimensions. Researchers in the field look into how treatment integrity can be conceptualized, measured, and promoted. On the other hand, DBIR seeks to develop and refine a design as well as theories related to learning and teaching in contexts. With the accumulated discussions in each domain, this project first explores the similarities and

differences between the fields. It is shown that the two fields are different in the specificity and target of focus. It is also shown that while DBIR emphasizes the collaboration among multiple stakeholders and mixed research methods to collect data, treatment integrity research has started to recognize the benefits of collaboration among partners as well as the advantages of utilizing hybrid methods to collect treatment integrity data. The comparisons provide a platform for a dialogue for these two fields. The dialogue in the project suggests more mutual understanding as well as an exploration of complementarity between these fields. Challenges in treatment integrity research such as the complexity of conceptualization models and the conflict between fidelity and flexibility may be addressed with the emphasis of theories in design-based implementation research. Treatment integrity research may also benefit DBIR researchers to more efficiently spot challenges in implementing new innovations. Together, researchers in both fields may collaborate to cope with challenges they both face. It is encouraged to explore the possibility of addressing the challenges. Future research is also expected to investigate the value in synergizing these fields for the implementation of educational practices and designs.

## References

- Barab, S. (2004). Using design to advance learning theory, or using learning theory to advance design. *Educational Technology*, 44(3), 16-19.
- Bell, P. (2004). On the Theoretical Breadth of Design-Based Research in Education. *Educational Psychologist*, 39(4), 243–253.
- Bond, G. R., Drake, R. E., McHugo, G. J., Rapp, C. a., & Whitley, R. (2009). Strategies for Improving Fidelity in the National Evidence-Based Practices Project. *Research on Social Work Practice*, 19(5), 569–581.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2, 141–178.
- Bruhn, A. L., Hirsch, S. E., & Lloyd, J. W. (2015). Treatment Integrity in School-Wide Programs: A Review of the Literature (1993–2012). *The Journal of Primary Prevention*, 36(5), 335–349.
- Cobb, P. (2001). Supporting the improvement of learning and teaching in social and institutional contexts. Cognition and instruction: Twenty-five years of progress. S. M. Carver and D. Klahr. Mahwah, NJ, Erlbaum: 455-478.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, 32(1), 9.
- Coburn, C. E. (2003). Rethinking Scale: Moving Beyond Numbers to Deep and Lasting Change. *Educational Researcher*, 32(6), 3–12.
- Coburn, C. E., Penuel, W. R., & Geil, K. E. (2013). *Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts*. New York, NY.
- Cochrane, W. S., & Laux, John, M. (2008). A Survey Investigating School Psychologists' Measurement of Treatment Integrity in School-based Interventions and Their Beliefs about its Importance. *Psychology in the Schools*, 45(6), 499–507.
- Collins, A. (1992). Toward a design science of education. In E. Scanlon & Tim O'Shea (Eds.), *New directions in educational technology* (pp. 15-22). New York: Springer-Verlag.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design Research: Theoretical and Methodological Issues. *Journal of the Learning Sciences*, 13, 15–42.
- 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(1), 23–45.
- DiGennaro, F. D., Martens, B. K., & Kleinmann, A. E. (2007). A Comparison of Performance

- Feedback Procedures on Teachers' Treatment Implementation Integrity and Students' Inappropriate Behavior in Special Education Classrooms. *Journal of Applied Behavior Analysis*, 40(3), 447–461.
- diSessa, A. A., & Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *Journal of the Learning Sciences*, 13(1), 77–103.
- 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(3-4), 327–350.
- Elias, M. J., Zins, J. E., Graczyk, P. A., & Weissberg, R. P. (2003). Implementation, Sustainability, and Scaling Up of Social- Emotional and Academic Innovations in Public Schools. *School Psychology Review*, 32(3), 303–319.
- Fallon, L. M., Collier-Meek, M. a, Maggin, D. M., Sanetti, L. M. H., & Johnson, A. H. (2015). Is Performance Feedback for Educators an Evidence-Based Practice? A Systematic Review and Evaluation Based on Single-Case Research. *Exeptional Children*, 81(2), 227–246.
- Fixsen, D. L., Naoom, S. F., Blase´, 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 Net- work (FMHI Publication No. 231).
- Fishman, B. J., Penuel, W. R., Allen, A., Haugan Cheng, B., & Sabelli, N. (2013). Design-Based Implementation Research: An Emerging Model for Transforming the Relationship of Research and Practice. *National Society for the Study of Education*, 112(2), 136–156.
- Forman, S. G., Shapiro, E. S., Coddling, R. S., Gonzales, J. E., Reddy, L. A., & Rosenfield, S. A. (2013). Implementation science and school psychology. *School Psychology Quarterly*, 28, 77–100
- Fryling, M., Wallace, M. D., & Yassine, J. N. (2012). Impact of Treatment Integrity on Intervention Effectiveness. *Journal of Applied Behavior Analysis*, 45(2), 449–453.
- Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review*, 31(1), 79–88.
- Greenwood, C. R. (2009). Treatment integrity: Revisiting some big ideas. *School Psychology Review*, 38(4), 547–553.
- Gresham, F. M. (2009). Evolution of the treatment integrity concept: current status and future directions. *School Psychology Review*, 38(4), 533–540.
- Gresham, F. M., & Gansle, K. A. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychology Review*, 22(4), 254–272.

- Gross, T. J., Duhon, G. J., & Doerksen-Klopp, B. (2014). Enhancing Treatment Integrity Maintenance Through Fading with Indiscriminable Contingencies. *Journal of Behavioral Education, 23*(1), 108–131.
- Gutierrez, K. D., & Vossoughi, S. (2010). Lifting Off the Ground to Return Anew: Mediated Praxis, Transformative Learning, and Social Design Experiments. *Journal of Teacher Education, 61*, 100–117.
- 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*(2), 181–193.
- Huey, S. J., & Polo, A. J. (2008). Evidence-based psychosocial treatments for ethnic minority youth. *Journal of Clinical Child & Adolescent Psychology, 37*(1), 262–301.
- Jeffrey, J. L., McCurdy, B. L., Ewing, S., & Polis, D. (2009). Classwide PBIS for Students with EBD: Initial Evaluation of an Integrity Tool. *Education & Treatment of Children (ETC), 32*(4), 537–550.
- Jones, H. A., Clarke, A. T., & Power, T. J. (2008). Expanding the concept of intervention integrity: A multidimensional model of participant engagement. In *Balance, 23*, 4–5.
- Keller-Margulis, M. A. (2012). Fidelity of implementation framework: A critical need for response to intervention models. *Psychology in the Schools, 49*(4), 342–352.
- Klingner, J. K., Boardman, A. G., & McMaster, K. L. (2013). What does it take to scale up and sustain evidence-based practices? *Exceptional Children, 79*(2), 195–211.
- Lau, A. S. (2006). Making the case for selective and directed cultural adaptations of evidence-based treatments: Examples from parent training. *Clinical Psychology: Science and Practice, 13*(4), 295–310.
- McKenna, J. W., Flower, a., & Ciullo, S. (2014). Measuring Fidelity to Improve Intervention Effectiveness. *Intervention in School and Clinic, 50*(1), 15–21.
- Mckenna, J. W., Flower, A., & Ciullo, S. (2014). Measuring Fidelity to Improve Intervention Effectiveness, *50*(1), 15–21.
- McLeod, B. D., Southam-Gerow, M. a., & Weisz, J. R. (2009). Conceptual and methodological issues in treatment integrity measurement. *School Psychology Review; School Psychology Review, 38*(4), 541–546.
- Mitchell, P. F. (2011). Evidence-based practice in real-world services for young people with complex needs: New opportunities suggested by recent implementation science. *Children and Youth Services Review, 33*(2), 207–216.
- Mowbray, C. T., Holter, M. C., Gregory, B., & Bybee, D. (2003). Fidelity Criteria : Development , Measurement , and Validation, *24*(3), 315–340.

- Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *Journal of Behavioral Health Services and Research, 39*(4), 374–396.
- Noell, G. H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W. P. Erchul & S. M. Sheridan (Eds.) (Ed.), *Handbook of research in school consultation: Empirical foundations for the field* (pp. 315–334). Mahwah, NJ: Erlbaum.
- Noell, G. H., Gansle, K. A., Mevers, J. L., Knox, R. M., Mintz, J. C., & Dahir, A. (2014). Improving Treatment Plan Implementation in Schools: A Meta-analysis of Single Subject Design Studies. *Journal of Behavioral Education, 23*(1), 168–191.
- Noell, G. H., Gresham, F. M., & Gansle, K. a. (2002). Does treatment integrity matter? A preliminary investigation of instructional implementation and mathematics performance. *Journal of Behavioral Education, 11*(1), 51–67.
- Noell, G.H., Witt, J.C., Slider, N.J., Connell, J.E., Gatti, S.L., Williams, K.L., Koenig, J.L., Resetar, J.L., & Duhon, G.J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review, 34*(1), 87-106.
- Penuel, W. R., Fishman, B. J., Haugan Cheng, B., & Sabelli, N. (2011). Organizing Research and Development at the Intersection of Learning, Implementation, and Design. *Educational Researcher, 40*(7), 331–337.
- Penuel, W. R., Phillips, R. S., & Harris, C. J. (2014). Analysing teachers' curriculum implementation from integrity and actor-oriented perspectives. *Journal of Curriculum Studies, 46*(6), 751–777.
- Perepletchikova, F. (2005). Treatment Integrity and Therapeutic Change: Issues and Research Recommendations. *Clinical Psychology: Science and Practice, 12*(4), 365–383. <http://doi.org/10.1093/clipsy/bpi045>
- Perepletchikova, F., Treat, T. a, & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology, 75*(6), 829–841.
- Power, T. J., Blom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kelleher, C., & Manz, P. H. (2005). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools, 42*(5), 495–507.
- Prowse, P. T., & Nagel, T. (2015). A Meta-Evaluation: The Role of Treatment Fidelity within Psychosocial Interventions during the Last Decade. *Journal of Psychiatry, 18*(2).

- Reiser, B. J., Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., & Leone, A. J. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: 25 years of progress*. Mahwah, NJ: Lawrence Erlbaum.
- Riekert, K. A. (2006). Integrating regimen adherence assessment into clinical practice. In W. T. O'Donohue & E. R. Levensky (Eds.), *Promoting treatment adherence: A practical handbook for health care providers* (pp.17–34). Thousand Oaks, CA: Sage.
- Roschelle, J., Knudsen, J., & Hegedus, S. (2009). From new technological infrastructures to curricular activity systems: Advanced designs for teaching and learning. In M. J. Jacobson & P. Reimann (Eds.), *Designs for Learning Environments of the Future: International Perspectives from the Learning Sciences*. New York: Springer. ce Erlbaum.
- Russell, J. L. I. N., Jackson, K., & Frank, K. a. (2013). Theories and research methodologies for design-based implementation research: Examples from four cases. *The National Society for the Study of Education Yearbook*, 112(2), 157–191.
- Sanetti, L. M. H., Chafouleas, S. M., Christ, T. J., & Gritter, K. L. (2009). Extending Use of Direct Behavior Rating Beyond Student Assessment. *Assessment for Effective Intervention*, 34(4), 251–258.
- Sanetti, L. M. H., & Collier-Meek, M. a. (2014). Increasing the Rigor of Procedural Fidelity Assessment: An Empirical Comparison of Direct Observation and Permanent Product Review Methods. *Journal of Behavioral Education*, 23(1), 60–88.
- Sanetti, L. M. H., Collier-Meek, M. A., Long, A. C. J., Kim, J., & Kratochwill, T. R. (2014). Using Implementation Planning to Increase Teachers' Adherence and Quality to Behavior Support Plans. *Psychology in the Schools*, 51(8), 879–895.
- Sanetti, L. M. H., Dobe, 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(1), 29–46.
- Sanetti, L. M. H., & Kratochwill, T. (2008). Treatment Integrity in Behavioral Consultation: Measurement, Promotion, and Outcomes. *International Journal of Behavioral Consultation and Therapy*, 4(1), 95–114.
- Sanetti, L. M. H., & Kratochwill, T. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review*, 38(4), 445–459.
- Sanetti, L. M. H., & Kratochwill, T. R. (2009). Treatment integrity assessment in the schools: An evaluation of the Treatment Integrity Planning Protocol. *School Psychology Quarterly*, 24(1), 24–35.
- Sanetti, L. M. H., Kratochwill, T. R., & Long, A. C. J. (2013). Applying adult behavior change theory to support mediator-based intervention implementation. *School Psychology Quarterly*, 28(1), 47–62.

- Sannino, A., Sutter, B., & Engeström, Y. (2011). *From design experiments to formative interventions. Theory & Psychology, 21*(5), 598–628.
- 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, *38*(4), 460–475.
- Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003). On the science of education design studies. *Educational Researcher, 32*(1), 25–28.
- Sheridan, S. M., Swanger-gagne, M., Welch, G. W., Kwon, K., & Garbacz, S. A. (2009). Fidelity Measurement in Consultation : Psychometric Issues and Preliminary Examination. *School Psychology Review, 38*(4), 476 – 495.
- St. Peter Pipkin, C., Vollmer, T. R., & Sloman, K. N. (2010). Effects of Treatment Integrity Failures During Differential Reinforcement of Alternative Behavior: a Translational Model. *Journal of Applied Behavior Analysis, 43*(1), 47–70.
- The Design-Based Research Colletive. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5–8.
- Tigchelaar, A., & Korthagen, F. (2004). Deepening the exchange of student teaching experiences: Implications for the pedagogy of teacher education of recent insights into teacher behaviour. *Teaching and Teacher Education, 20*(7), 665–679.
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*(4), 620–630.
- Ward, P., Johnson, M., & Konukman, F. (1998). Directed Rehearsal and Preservice Teachers ' Performance of Instructional Behaviors, *8*(3), 369–380.
- Weisz, J. R., Doss, A. J., & Hawley, K. M. (2005). Youth psychotherapy outcome research: a review and critique of the evidence base. *Annual Review of Psychology, 56*, 337–63.
- Zheng, L. (2015). A Systematic Literature Review of Design-based Research from 2004 to 2013. *Journal of Computers in Education, 2*(4), 399–420.