Theses and Dissertations

2016

Check & Connect: An Investigation of Variables Influencing Social Validity and the Relation between Social Validity and Treatment Integrity

Shu-Chen Tsai
Lehigh University

Check & Connect:

An Investigation of Variables Influencing Social Validity and

the Relation between Social Validity and Treatment Integrity

By

Shu-Chen Tsai

Presented to the Graduate and Research Committee

of

Lehigh University

In Candidacy for the Degree of

Doctor of Philosophy

In

Special Education

Lehigh University

April 2016

Shu-Chen Tsai

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements of Doctor of Philosophy

_____

Date

_____

Chair
Lee Kern, Ph.D.
Professor of Special Education

_____

Accepted Date

Committee Members

_____

Grace Caskie, Ph.D.
Associate Professor of Counseling Psychology

_____

George DuPaul, Ph.D.
Professor of School Psychology

_____

Michael George, Ed.D.
Director of Centennial School

iii

謹獻於我的父母

特別是我的母親

因她的愛與勇氣

讓我跳脫台灣父母對女兒的傳統框架

支持我在海外圓博士夢

且讓我成為我

&

緬懷我的神師

宣國榮神父

深信他必為我感到驕傲

To my parents,

especially my mother,

for her love and courage

to free me from traditional Taiwanese parents' expectations on their daughters,

support me to pursue my doctorate overseas, and

let me be me.

&

In memory of my spiritual director,

late Fr. Fernado Zello, S.J.,

who, I am sure,

proudly smiles at me up in the heaven.

Acknowledgements

I would like to express my appreciation to my committee members, Drs. DuPaul, George, and Caskie for their guidance and patience with me. Special thanks to Dr. Caskie, who helped me stretch my knowledge boundary to learn about advanced statistical analysis —multiple imputation— and guided me to contemplate the art of statistical thinking and decision-making process. I also would like to thank Dr. Joan Fu, who helped me grasp difficult statistical concepts and debug SAS code through my mother tong when I needed.

I am especially in debt to my advisor, Dr. Kern, who has been so patient with me throughout the journey of my doctoral study, who has shaped me as a researcher to think critically, and who helped me embrace another identity of a scholar— a writer and to confidently become a good one.

My gratitude is also extended to several good friends: Fr. Ignatius Hung, my godmother, Raul, and Louis, who have carried me in their prayers, especially in my tough times. Special thanks to my dear friend, Miao-Ling (妙鈴), who had companied me to walk out the darkest moment during my doctoral study. My deep appreciation also goes to Ms. Share Warden, our Special Ed. doctoral students' guardian angel, who always blesses students before they walk into to the room to take the comprehensive exam. The gentle sign of the cross and warm hug she gave eased my nerve and remains vivid to me.

Above all, I am grateful for all the blessings the Lord has bestowed on me: to be nurtured by great minds in a prestigious university and to be companied by friends from all over the world. Though challenging, this intellectual growth is the best gift I've ever received— and I thank you, Lord!

Table of Contents

Abstract

Check & Connect, a mentoring intervention, has shown promise in promoting school engagement of students with disabilities; however, its social validity has not been examined in the literature. The absence of Check & Connect social validity data reflects the limitations of broader intervention literature and the area of social validity research. Therefore, the purpose of this study was to investigate social validity of Check & Connect and the influencing variables using existing data from the Center of Adolescent Research in Schools (CARS), a longitudinal, randomized controlled trial. A series of multiple regressions with multiple imputation was conducted. Results revealed that (a) students and mentors perceived Check & Connect as acceptable; (b) students' social validity ratings in Year 1, dosage across two-years, and change in mentor as a whole significantly explained 15% of the variance in students' Year 2 ratings, but only Year 1 ratings significantly predicted Year 2 ratings; (c) student and mentor characteristics (i.e., student behavior severity and special education status, and mentor years of teaching experience) were not significant in predicting students' and mentors' social validity ratings in Year 2; and (d) mentors' social validity ratings in Year 2 significantly predicted their treatment integrity in the same year despite the magnitude being small (10% of the variance). Implications for practice and future research pertaining to replication, assessment, methodology, and utility are discussed.

1

Chapter I

Statement of the Problem

Providing effective interventions for high school students with disabilities to address their school engagement and connectedness is imperative. Compared to general education students, students with disabilities, particularly those with emotional/behavioral disorders (EBD), learning disabilities (LD), or other health impairment (OHI) (e.g., attention deficits/hyperactivity disorders [ADHD]), are more likely to disengage from school (Dun, Chambers, & Rabren, 2004; Reschly & Christenson, 2006; Zablocki & Krezmien, 2013). Their disengagement manifests in performing poorly in academics (e.g., low grade point, low rates of homework completion) (Kent et al., 2011), frequently receiving office disciplinary referrals (ODRs)/exclusions (Karpinski, Neubert, & Graham, 1992), and engaging in risk behaviors (e.g., substance abuse) (Hollar, 2005; McNamara, Vervaeke, & Willoughby, 2008; McMamara & Willoughby, 2010). Consequently, they are prone to dropping out of school (Chapman, Laird, Ifill, & KewalRamani, 2011). High school incompletion further prohibits them from becoming productive and healthy citizens in adulthood (Bureau of Labor Statistics [BLS], 2012).

Mentoring is a promising intervention to address the difficulties that high school students experience. Drawing from resiliency literature, a caring adult serves a protective factor in adolescent development (Woolley & Bowen, 2007; Zimmerman, Bingeheimer, & Notaro, 2002). The connection/bonding between a supportive adult and a protégé provides adolescents with social capital to allocate resources and persevere in challenging circumstances. Empirical studies have shown that adolescents benefit from mentoring programs in several areas, including school engagement, mental health, and social/emotional development (Britner & Kraimer-Rickaby, 2009; Grossman & Bulle, 2006). For example, results from a well-established

mentoring program, Big Brother/Big Sister (Herrera, Grossman, Kauh, & McMaken, 2011; Jekielek, Moore, & Hair, 2002) indicate that compared to non-mentored students, mentored students had more positive attitudes toward school and better class attendance. Mentored students were also more likely to complete high school and attend college. Additionally, mentored students were less likely to engage in risk behaviors.

Check & Connect

Check & Connect is one of the few mentoring programs that has been evaluated with students with disabilities. Check & Connect research was originally funded by the Office of Special Education Programs (OSEP) as an initiative to evaluate its effects on school engagement for high school students with emotional/behavioral disorders or learning disabilities (Sinclair, Christenson, Lehr, & Anderson, 2003). In the past decade, Check & Connect research has been extended to elementary students at risk for dropout (e.g., Lehr, Sinclair, & Christenson, 2004) and to preschoolers with early literacy concerns (e.g., Miltich Lyst, Gabriel, O'Shaughnessy, Meyers, & Meyers, 2005).

Check & Connect promotes student engagement through establishing a trusting relationship between the student and a mentor. In Check & Connect, an adult who is willing to commit to work with the student for at least 2 years, from school to school if needed, and believes that the student has the ability to succeed, serves as a mentor (Sinclair, Christenson, Lehr, & Anderson, 2003). A mentor regularly meets with the student, building a connection and open communication, to ensure the student engages in school behaviorally, academically, cognitively and psychologically (Kortering & Christenson, 2009).

Check & Connect is also a data-driven and comprehensive intervention (Christenson, Stout, & Pohl, 2012). It is comprised of seven components: monitoring student engagement,

long-term commitment, relationship-building with the student, teaching problem-solving, promoting persistence, collaborating with schools and families, and providing timely and individualized intervention (Sinclair, Christenson, & Thurlow, 2005). In the Check part, the mentor uses monitoring sheets to regularly check and monitor the student's behavioral and academic engagement. In the Connect part, the mentor works with the student and the family for at least 2 years, and also may follow the student from school to school. The mentor meets with the student on a weekly to bi-weekly basis and establishes mutual trust. During the meeting, the mentor discusses the student's performance in school and any concerns he/she may have to enhance the student's school participation. Meanwhile, the mentor models the five steps of problem solving for the student. Furthermore, the mentor consistently conveys the message to the student that school completion is crucial and the mentor will not give up on the student. The mentor also frequently communicates with teachers and families with regard to the student's engagement. Based on the data collected from regular checking, the mentor determines if the student needs a more intensive individualized intervention and follows through to see that it is implemented.

Importance of Treatment Fidelity

Treatment fidelity is the extent to which an intervention is implemented as intended (Gresham, 2009). Four aspects of treatment fidelity are crucial in determining the validity of interventions and evidence-based practices (Gresham, 2014). First, pertaining to internal validity, without treatment fidelity, it is difficult to discern whether changes in the behavior of interest truly result from the intervention. Second, related to external validity, without consistent implementation, it is unclear if the intervention can be generalized to different settings. Third, with regard to construct validity, inconsistent implementation compromises what constitutes the

4

intervention. Finally, in terms of statistical conclusion validity, statistical inference based upon

the outcome data yielded from an intervention with low fidelity may mistakenly lead one to

conclude an intervention is effective while it is not (Type I error) or vice versa (Type II error).

Importance of Social Validity

Social validity is the extent to which consumers deem the goal(s) of an intervention to be

socially significant, the procedures of an intervention to be appropriate, and the outcomes to be

positive (Kazdin, 1977; Wolf, 1978). In general, social validity is obtained through asking

various consumers their perceptions of an intervention using questionnaires or interviews. These

include direct consumers, indirect consumers, and community members (Schwartz & Baer,

1991). Direct consumers are intervention recipients, such as the student and the teacher, while

indirect consumers are non-recipients for whom the intervention has a direct impact, such as

school personnel. Community members are those who may or may not have contact with the

aforementioned two types of consumers but for whom the intervention has an indirect impact.

For example, taxpayers may be concerned with the high cost of an intervention and appeal for a

suspension.

Given that implementing interventions in applied settings involves typical personnel,

assessing consumers' perceptions is particularly important for bridging the research-practice gap

and promoting evidence-based practices. As in the cases illustrated by Strain, Barton and

Dunlap (2012), what the researcher deemed as an important goal or an appropriate behavior for

intervention may not always align with a parent's or a teacher's judgments. Knowing

consumers' perceptions allows the researcher to refine the intervention and training so that it has

a good contextual fit. Strain et al. also documented that when practitioners perceived the

intervention as acceptable, they tended to implement the procedures with fidelity and speculated that they may continue to implement the intervention after the termination of a research project.

Research Gaps in Social Validity

Although accumulating studies have shown the promising results of Check & Connect (e.g., Sinclair, Christenson, Elevo, & Hurley, 1998; Sinclair, Christenson, & Thurlow, 2005), social validity has rarely been reported. In general, there is a dearth of social validity in the intervention research. Literature reviews on social validity (e.g., Carr, Austin, Britton, Kellum, & Bailey, 1999; Kennedy, 1992) reveal that reporting social validity slightly increased in major journals on behavior change (e.g., *Journal of Applied Behavior Analysis*, *Journal of Behavior Modification*) during the 70s-80s after the concept was introduced to the field; yet, a decreasing trend was observed during the 80s-90s. Further, comparing intervention studies published in three peer-reviewed journals from 1999 to 2005, Clark and Dunlap (2008) found that intervention studies published in the *Journal of Positive Behavior Intervention (JPBI)* had the highest rate of reporting social validity; yet, only one-third of those studies reported social validity.

In the research that has examined social validity, the scope is limited as well. For instance, examining the validity issues related to School-Wide Positive Behavior Support (SWPBS), Kern and Manz (2004) pointed out that the social validity of SWPBS is limited in several aspects. First, social validity lacks a wide range of consumers' perceptions, particularly students' perspectives. Second, social validity is typically assessed at the end of intervention, rather than over an extended period of time such as yearly assessments. Thus, it is unclear if social validity sustains over time. Third, within the tiered framework SWPBS employs, social

validity is predominately assessed for the primary tier. Little attention has been paid to the social validity of Tier 2 and 3 interventions.

Further tracking back to early research in the 80s-90s, a number of empirical investigations indicated that consumers' judgments of treatment acceptability are influenced by several factors, including consumers' knowledge of an intervention (e.g., Vereb & DiPerna, 2004), years of teaching experience (e.g., Witt & Robbins, 1985), time involved (e.g., Witt, Martens, & Elliot, 1984), the type of intervention (e.g., Harris, Preller & Graham, 1990), and the child's behavior severity (e.g., Cowan & Sheridan, 2003) and diagnostic labels (e.g., Stinnett, Crawford, Gillespie, Cruce & Langford, 2001). However, the majority of these investigations were conducted in analogue settings. To date, few studies with this research agenda have been carried out in natural settings (e.g., Cowan & Sheridan, 2003). Additionally, given that Check & Connect is a relationship-based intervention, frequent contact between mentors and students (i.e., dosage) and consistency in mentors are vital for relationship building (Grossman & Rhodes, 2002). Yet, the potential impact of dosage and change in implementer/mentor on students' acceptability has not been explored in the social validity literature. Furthermore, teachers' years of teaching experience, students' behavior severity, and students' diagnostic labels have been found to influence behavior-based interventions and may hold true in mentoring-based interventions. Hence, it is imperative to continue this research inquiry in natural settings, expanding identified variables beyond behavior-based interventions and examining other potential influencing variables on social validity.

Despite the fact that several theoretical models (e.g., Witt & Elliot, 1985; Lane, Beebe-Frankenberger & Lambros, 2001) postulate the relation between social validity and treatment fidelity, empirical studies that systematically test these models are scant and the findings are

7

inconsistent. Currently, some case studies (e.g., Ehrhardt, Barnett, Lentz Jr., Stollar, & Reifin, 1996; Strain, Barton, & Dunlap, 2012) and preliminary empirical data (e.g., Miramontes, Marchant, Heath, & Fischer, 2011) suggest that social validity and treatment fidelity may be correlated, while other studies (e.g., Sterling-Turner & Watson, 2002) show no significant correlation. Therefore, there is a need for more empirical studies to elucidate the relation between social validity and treatment fidelity.

Statement of Purpose

Given the promising outcomes of Check & Connect in preventing high school students with high-incidence disabilities from dropping out of school and the dearth of social validity in Check & Connect and the intervention literature, the purpose of the current study was to examine the social validity of Check & Connect, using existing data from a longitudinal, randomized-control trial study. Specifically, the current study aimed to extend the literature by (a) examining the students' report of social validity for an extended period of time and investigating the impact of different dosages on the student report of social validity; (b) investigating to what extent that the students' behavior severity, diagnostic status (i.e., no label vs. with label) and dosage predict students' social validity scores for Check & Connect; (c) investigating to what extent student-related characteristics (i.e., behavior severity, diagnostic status) and the mentor-related characteristics (i.e., years of teaching experience) predict the mentors' report of social validity, and (d) examining the relation between the mentors' report of social validity and their treatment integrity.

Research Questions

The specific research questions related to students' report of social validity and mentors' report of social validity were:

8

Research Question 1. Do students' social validity ratings at Year 1, the percentage of Check & Connect sessions participated, and change in mentors predict student social validity ratings at Year 2?

Hypothesis 1. Based on the Check & Connect and the social validity literature, it is hypothesized that students' ratings at Year 1 and the percentage of Check & Connect sessions participated and no change in mentors will significantly predict their ratings at Year 2.

Research Question 2. Does students' behavior severity, measured by the teacher-rated BASC-2 (i.e., TRS), diagnostic status, mentors' years of teaching experience and the percentage of Check & Connect sessions participated predict students' social validity ratings at Year 2?

Hypothesis 2. Based on the Check & Connect and the social validity literature, it is hypothesized that students' behavior severity, diagnostic status, mentors' years of teaching experience and the percentage of Check & Connect sessions participated will significantly predict students' social validity ratings.

Research Question 3. Does students' behavior severity, measured by the teacher-rated BASC-2 (i.e., TRS), students' diagnostic status, mentors' years of teaching experience, and the percentage of Check & Connect sessions participated predict mentors' social validity ratings at Year 2?

Hypothesis 3. Based on the Check & Connect and the social validity literature, it is hypothesized that students' behavior severity, diagnostic status, mentors' years of teaching experience and the percentage of Check & Connect sessions participated will significantly predict the mentors' social validity ratings at Year 2.

Research Question 4. Do mentors' social validity ratings at Year 2 predict their treatment fidelity at Year 2?

Hypothesis 4. Given the Witt and Elliot (1985) and the Lane et al. (2001) theoretical models postulating that the higher social validity results in higher treatment fidelity, it is hypothesized that mentors' social validity ratings significantly predict their treatment fidelity.

Chapter II

Review of the Literature

The Importance of School Engagement

School engagement is critical for learning and school completion. Regardless of different terms used in the literature (e.g., student/school/academic engagement, connectedness), school engagement comprises at least three dimensions: behavioral engagement, emotional/affective/ psychological engagement, and cognitive engagement (Appleton, Christenson, & Furlong, 2008; Fredricks, Blumenfeld, & Paris, 2004; Furlong et al., 2003). Behavioral engagement refers to a student's participation in school and learning activities (e.g., attending classes, completing homework assignments) and display of positive conduct (e.g., following the rules). Emotional/affective/psychological engagement is the psychological state where a student feels connected with teachers and peers, has a sense of belonging to the school community, values the importance of education, and shows positive attitudes toward learning (e.g., liking school). Cognitive engagement occurs when a student invests in learning, self-regulates, and is motivated to take on challenges. Accumulating research has shown that higher school engagement is associated with better academic achievement across students of different ages and behavioral engagement. In particular, school engagement prevents students from dropping out of school (see Fredricks, Blumenfeld, & Paris, 2004 for a review). For example, Finn and Rock (1997) analyzed a sample of 1,803 8th-12th graders from the National Educational Longitudinal Study of 1988 (NELS-88) and found that after controlling for demographic variables, students who demonstrated academic success and completed school (classified as "resilient") significantly differed in engagement from students who completed school with poor academic performance and those who dropped out. Resilient students were characterized by high self-esteem and a

good sense of control over their lives. They also showed up in class on time, were prepared, avoided distractions, and completed homework.

Unfortunately, many high school students with high-incidence disabilities are more likely to drop out of school than their counterparts because of school disengagement. According to data from the National Center for Education Statistics (NCES) in 2013, although overall dropout rate has shown a decreasing trend from 12% in 1990 to 7% in 2012, high school dropout rates for students with disabilities are two times higher. Specifically, the dropout rate for students with disabilities was 21.1% during 2009-2010 and 20.1% during 2010-2011. Further National Longitudinal Transitional Study-2 (NLTS-2) showed differences among disability categories with students having learning disabilities (LD), emotional and behavioral disorders (EBD), and other health impairments (OHI) the most vulnerable groups for dropping out of school (Zablocki & Krezmien, 2013).

Tracing back the pathways of school dropout, researchers found that with respect to behavior disengagement, high school students with high-incidence disabilities exhibit more frequent indicators of dropout, including higher rates of problem behavior and poorer academic performance than their general education peers. For example, McNamara and colleagues (2008; 2010) reported that compared to non-disabled peers, students with LD or LD/ADHD were more likely to engage in risk-taking behaviors such as smoking and substance use. Reschly and Christenson (2006) also found that students with EBD or LD displayed poorer attendance and a lack of class preparation compared with their typical peers. Because of poor attendance, many high school students with high-incidence disabilities receive school disciplinary procedures and suspensions (Kaufman et al., 2010; Krezmien, Leone, & Achilles, 2006). Coupled with excessive absences and office disciplinary referrals (ODRs), high school students with high-

incidence disabilities often experience course failures and, as a result, many of them are retained (Bowman, 2005; Wagner, 1991; Zablocki & Krezmien, 2012). In the area of psychological disengagement, reflecting on their experiences, high school dropouts with high-incidence disabilities disclosed that they did not have a helpful person in school and perceived that they were socially alienated from teachers and peers (Dun, Chambers, & Rabren, 2004; Seidel & Vaughn, 1991). Regarding cognitive disengagement, research indicates that high school dropouts with high-incidence disabilities experienced boredom in school and viewed high school as having no utility post-graduation. Retrospectively, many dropouts with high-incidence disabilities indicated that they wished they had tried harder to persevere (Kortering & Brazeil, 1999).

Dropping out of school not only can lead to poorer quality of life, but also results in costs to society (Bridgeland, Dilulio, & Morison, 2006; Sum, Khatiwada, & McLaughlin, 2009). At the individual level, according to the report from the National Longitudinal Transition Study-2 (NLTS2), compared to school completers with disabilities, high school dropouts with disabilities had significantly lower employment rates and were less likely to pursue postsecondary education (Newman et al., 2011). Additionally, high school dropouts with disabilities were more likely to live in poverty and only 65% of dropouts with disabilities reported that they were in good health compared to 72% of high school graduates and 82% of youth with an associate or bachelor's degree. Further, high school dropouts with disabilities were more likely to be incarcerated than school graduates; in particular, dropouts with EBD had three times higher arrest rates than high school graduates with disabilities. At the society level, taxpayers have to bear the cost of incarceration and correction systems, which are largely made up with high school dropouts with high-incidence disabilities (Quinn, Rutherford, Leone, Osher, & Poirrier, 2005). Given the

13

profound impact of school disengagement on high school students with high-incidence disabilities and on society, it is critical to improve their school engagement with effective interventions.

Mentoring and School Engagement

Resilience refers to the state whereby youth overcome adversity (i.e., risk factors such as poverty, violence, dysfunctional family) and achieve expected developmental outcomes (Masten, 2001). Researchers attribute a youth's resilience to three protective factors: (a) the individual's disposition (e.g., cognitive abilities, temperament, self-regulation), (b) the family qualities (e.g., warm parenting, structure, expectations), and (c) the community resources (e.g., social network with competent adults, good schools and neighborhood) (Masten & Powell, 2003). Among the protective factors in the area of community resources, support of youth's resilience outside the family has not been extensively researched (Rhode & Lowe, 2008). However, accumulating studies show that one good relationship with a significant non-parental adult can ameliorate threats to a youth's development and promote positive wellbeing (Beam, Chen, & Greenberger, 2002; Zand et al., 2009). As evidenced in the analysis of the National Longitudinal Study of Adolescent Health (Add health), DuBois and Silverhorn (2005) found that adolescents who reported having a significant non-parental adult were more likely to complete high school, attend college, and have longer work hours. Additionally, having a significant non-parental adult was associated with a higher level of self-esteem, greater life satisfaction, higher engagement in physical activities, and lower levels of problem behavior (i.e., becoming a gang member, fighting).

Such a non-parental bonding also benefits adolescents with high-incidence disabilities. A study by Ahrens, Dubois, Lozano and Richardson (2010) showed that adolescents with LD who

14

had a significant non-parental adult perceived the adult as a role model or substitute parental figure who provided support and guidance. In addition, they were more likely to graduate from high school and have higher levels of self-esteem and overall positive outcomes than those who did not have a connection with a non-parental adult. Therefore, relationship-based interventions such as mentoring programs hold potential to address youth's school disengagement.

Mentoring is a sustained relationship between a non-parental adult and a youth, in which the adult (the mentor) provides ongoing assistance and guidance to support the development and competence of the youth (the protégé) (Rhodes, 2002). Mentoring relationships can be established naturally or formally. Natural mentoring occurs in the youth's social networks through daily interactions with adults, such as teachers, coaches, or pastors (Zimmerman, Bingenheimer, & Behrendt, 2005). By contrast, formal mentoring is characterized in several ways (DuBois, Holloway, Valentine, & Cooper, 2002). First, adults (e.g., teachers) that are well suited to the mentoring program are recruited to be mentors. Second, the mentor and the mentee are usually matched based on certain demographic characteristics (e.g., gender, race/ethnicity) and mutual interests. Third, the formal mentoring relationship is structured and goal-oriented. With training and supervision, the mentor is expected to meet the mentee at a certain frequency and duration to address the mentee's needs (e.g., academics). Based on the setting and the focus, formal mentoring programs can be further classified into community-based or school-based programs (Herrea, Sipe, McClannahan, Arbreton, & Pepper, 2000). In the formal community-based mentoring (CBM), mentors and mentees encounter one another in the community settings and the focus is usually on mentees' after-school social activities and communication with parents. Formal school-based mentoring (SBM) is typically confined to the school context and oriented around academic activities. Mentors also have frequent contact with teachers. Big

Brothers/Big Sisters of America (BB/BSA) is a renowned example of a formal mentoring program that has been expanded from community-based models to school-based models (Herrea, 1999).

Researchers have proposed several theoretical models to explain the underlying mechanisms that make mentoring programs work to foster positive youth outcomes. Rhode (2005) posited that a mentor and a youth are likely to meet frequently when a mentoring relationship is built on mutual trust and empathy. In such a strong connection, the mentor models effective communication, thinking, and appropriate behaviors/values. Hence, the youth learns how to self-regulate, becomes receptive to adult perspectives, and constructs the idea of self. The mentoring relationship enhances youth's social-emotional, cognitive, and identity development, and development in these three domains interact with each other. As a result, this growth further promotes the youth's positive outcomes, such as emotional wellbeing, academic success, and behavior regulation. In particular, the impact of the mentoring relationship on the youth's social-emotional development may further help the youth's parental and peer relationships, which may mediate and predict later positive outcomes. In the model, Rhode also indicated that the quality of the mentoring relationship and the impact of mentoring on youth development are moderated by various individual variables (e.g., interpersonal history, social competence, developmental stage), the program (e.g., the duration of mentoring relationship, program practices), the family, and the community.

Despite that mentoring research is still in a developing stage and more rigor is needed (DuBois, Doolittle, Yates, Silverthorn, & Tebes, 2006), findings of formal SBM programs have shed promise on promoting at-risk adolescents' school engagement. For instance, aggregating three recent randomized control studies of SBM, Wheeler, Keller and DuBois (2010) indicated

that SBM programs showed modest effects on mentored students. Specifically, in behavior engagement, mentored students reduced their truancy, school-related misconduct, and absenteeism. In the area of psychological engagement, mentored students perceived connection and support from the mentor and peers. With respect to cognitive engagement, mentored students had a better sense of scholastic efficacy.

Check & Connect

Unlike most SBM programs, Check & Connect (Christenson, Stout, & Phol, 2012) is specifically designed to address the school engagement of secondary students with or at risk for EBD or LD. Check & Connect derives its conceptual framework from resiliency, systems theory, cognitive behavioral theory, and intrinsic motivation. First, from resilience theory, Check & Connect underscores the importance of a caring adult relationship in promoting adolescent engagement and focuses on alterable risk factors through skill building. Second, from systems theory, Check & Connect addresses engagement from a comprehensive perspective, including the student, the school and the family. Third, from cognitive behavioral theory, Check & Connect emphasizes the use of problem solving skills to address disengagement. Fourth, from intrinsic motivation, Check & Connect stresses internalizing motivation and self-efficacy. Namely, drawing upon multiple theories, Check & Connect is devised to promote school engagement via the mentor, who builds a relationship with the mentee, collaborates with the school and family, teaches the mentee problem solving skills, and encourages the mentee to persist. In addition, based on Finn's work (1989, 1993), Check & Connect expands the construct of engagement to a taxonomy that includes academic engagement (e.g., time on task, credits earned), behavioral engagement (e.g., attendance, school disciplines), cognitive engagement (e.g., self-regulation), and affective engagement (e.g., relationships with teachers and peers)

17

(Landis & Reschly, 2013). Therefore, the mentor frequently monitors the mentee's signs of engagement (the Check part) and communicates with the mentee, school, and parents (the Connect part).

In addition to the original program with secondary students, Check & Connect has been extended to address the engagement of elementary students with or at risk for problem behavior (e.g., Lehr, Sinclair, & Christenson, 2004). Also, the mentoring component of Check & Connect has been used to coach parents of at-risk preschoolers to promote early literacy skills in the home setting (e.g., Miltich Lyst, Gabriel, O'Shaughnessy, Meyers, & Meyers, 2005). Further, under the SWPBS framework, Check & Connect has been employed as a Tier 2 or Tier 3 intervention (Christenson, Stout, & Pohl, 2012; Horner, Sugai, & Anderson, 2010). Particularly, with the pressing need for Tier 2 interventions, Check & Connect and Check, Connect and Expect interventions have been successfully integrated into a tiered system (Cheney et al., 2010).

Although only three studies published in peer-reviewed journals employed a randomized control trial design, the results of Check & Connect show promise. The first efficacy trial was conducted by Sinclair, Christenson, Evelo and Hurley in 1998. A total of 94 9th grade students with LD and EBD (47 in treatment, 47 in control) participated in the study. The participants' school engagement was measured by their school participation (i.e., behavioral engagement, including enrollment status, attendance, and teacher-rated assignment completion), school performance (i.e., cognitive engagement including credits accrued, academic competence, and problem behavior on the teacher-rated SSRS), and identification with school (i.e., affect/psychological engagement, survey on relevance of school and expectation to graduate). *T*-tests and the Chi-square analyses indicated that despite no significant differences in identification with school, the treatment group students participated significantly more in school, demonstrated

significantly fewer problem behaviors and earned significantly more credits than the control group students after one year of receiving Check & Connect.

Similarly, Sinclair, Christenson and Thurlow (2005) examined the long-term effects of Check & Connect on another group of 144 9th graders with EBD (71 in treatment, 73 in control). In addition to the behavioral engagement measured by attendance, the outcome measures included the frequency of school mobility, the cohort school completion rate, and the IEPs for transition (e.g., numbers of student's attending the meeting, the transition assessment). The Chi-square analysis revealed that with 5 years of intervention, although there was no significant impact on school completion rates, the treatment group students were more likely to persistently attend school, remain in one school within a year, and participate in their IEP transition meetings than the control group students.

Unlike the studies by Sinclair and colleagues, a recent study by Maynard, Kjellstrand and Thompson (2013) focused on at-risk Hispanic students from 6th to 12th grade. At pretest, a total of 260 students enrolled (134 in treatment, 126 in control); however, due to high attrition, only 189 students remained in posttest (89 in treatment, 100 in control). Stepwise hierarchical linear modeling (HLM) was used to analyze student outcomes in academic performance (i.e., grades in English, math, science, and social studies), student disciplinary problems (i.e., total number of ODRs), and student attendance (i.e., number of absence). Compared to the control group students, after 6 months of Check & Connect intervention, although there were no significant differences in attendance, the treatment group students showed significant better grades and fewer ODRs.

Methodologically, in addition to high attrition, small sample size, and lack of experimental rigor (What Works Clearing House, 2011), another limitation across the board is

the omission of treatment fidelity and social validity. Specifically, only two studies (i.e., Lehr,

Sinclair, & Christenson, 2004; Miltich Lyst, Gabriel, O'Shaughnessy, Meyers, & Meyers, 2005)

examined the social validity of Check & Connect (see further discussion in the Social Validity

section).

Treatment Integrity

   Construct of treatment integrity. Treatment integrity, in general, refers to the degree to

which a treatment (or an intervention) is accurately and consistently implemented as intended

(Yeaton & Sechrest, 1981) and different terms are used across intervention research fields, such

as "treatment compliance/adherence" in medicine, "treatment fidelity" in clinical psychology,

"procedural reliability" in applied behavior analysis, and "program implementation" in

rehabilitation (Gresham, 2014). Regardless of the term used, there is consensus among

researchers across fields that treatment integrity is a multidimensional construct that

encompasses at least three broad aspects of treatment implementation, treatment receipt, and

treatment enactment (Schlte, Easton, & Parker, 2009).

   Treatment implementation. How well a treatment is implemented includes four

dimensions: treatment adherence, dosage/exposure, therapist/interventionist competence, and

treatment differentiation (Gresham, 2014; Schlte, Easton, & Parker, 2009). Treatment adherence

refers to the degree to which the components of a treatment are implemented as planned.

Dosage/exposure is the frequency or the duration that a treatment is implemented.

Therapist/interventionist competence refers to how skillful a therapist implements a treatment.

Treatment differentiation is the extent to which only a planned treatment is delivered.   In other

words, treatment implementation elucidates how competently a therapist implements a planned

treatment without deviating from the treatment manual, such that the treatment can be distinguished from other treatments.

Treatment receipt. In contrast to treatment implementation, treatment receipt focuses on the participants' acceptance of the treatment, including comprehension, dosage/exposure received and responsiveness (Schlte, Easton, & Parker, 2009). Participant comprehension refers to how well a participant understands the content of the treatment. Participant dose/exposure is the amount of a treatment a participant receives. Participant responsiveness is the degree to which a participant deems a treatment relevant or engages in a treatment.

Treatment enactment. Treatment enactment refers to the extent to which a participant acquires the skills in the training sessions and generalizes them in intended settings (Schlte, Easton, & Parker, 2009).

Importance of treatment integrity. Treatment integrity is crucial for intervention research and practice. From a research aspect, first, clearly defining the components and the procedures of a treatment allows researchers to provide implementation training. Implementation training can further facilitate interventionist competence and treatment adherence (Perepletchikova, 2014). Second, documenting treatment integrity provides formative data to monitor any implementation drift and determine any additional support that is needed for interventionists (Ledford & Wolery, 2013). Third, assessing treatment integrity permits researchers to draw valid inferences about the treatment (Shadish, Cook, & Campbell, 2002). When a treatment is not implemented as intended, internal validity suffers because researchers cannot draw a causal relation between dependent variables and independent variables. Research may be subject to Type I error when researchers conclude a treatment is effective when an expected outcome is found, but in fact the effect is caused by extraneous variables (King &

Bosworth, 2014). Alternatively, researchers may conclude the treatment is ineffective when an expected outcome is not found due to poor treatment integrity, namely, Type II error (King & Bosworth, 2014). Additionally, inconsistent implementation introduces unsystematic error into the data and hinders statistical testing to detect a treatment effect due to the increased within-group variability. As a result, researchers may reach an invalid statistical conclusion. Furthermore, lack of treatment integrity leads to ambiguity when explaining what the treatment was and why it worked. Hence, the construct validity of the treatment is compromised (Gresham, 2014). Lack of treatment integrity also limits the external validity because a treatment study cannot be replicated across different participants and settings without a clear description of what the components were and how they were delivered (Gresham, 2014). Fourth, treatment integrity advances the intervention research in understanding the underlying mechanisms of change (Perepletchikova, 2014). When a treatment is shown effective with clear operational definitions and good implementation, particularly a treatment package, it allows researchers to conduct further research by dismantling or adding components in order to identify the active or sufficient ingredients of the treatment.

From a practice aspect, assessing treatment integrity bridges the research-to-practice gap (Bumbarger, 2014). When a treatment study proceeds from efficacy trials to effectiveness investigations, treatment fidelity data allow the researcher to evaluate whether a treatment is applicable in natural settings. Specifically, when parents or teachers are the implementers, documenting treatment integrity permits researchers to track if adaptations are made. Further, monitoring treatment integrity helps researchers determine if implementers need additional supports.

Assessment of treatment integrity. Assessing treatment integrity consists of two methods: direct and indirect assessment (Gresham, 2014). Direct assessment is conducted through direct observation. To gather representative data, researchers should first determine the content, frequency and the length of an observation. Then, during the observation process, a trained observer follows an observation protocol to indicate the presence or absence of a procedure. By contrast, indirect assessment utilizes self-reports, permanent products, or behavioral interviews (Lane, Bocian, MacMillan, & Gresham, 2004). The self-report strategy involves asking implementers to indicate whether they implemented a procedure/component (e.g., Implemented vs. Not Implemented), or to rate the degree to which they implement a procedure/component on a Likert scale (e.g., "0 = none of the time" to "5 = most of the time"). The permanent products strategy utilizes the products generated by a treatment such as completed worksheets or self-monitoring forms to monitor treatment integrity. Behavioral interviews occur when a consultant conducts a semi-structured interview to verbally gather the consultee's report regarding treatment implementation (Wilkinson, 2007). During a behavioral interview, the consultant first facilitates the consultee's identification of success and obstacles of treatment implementation. Then, the consultant evaluates the self-reported integrity with the consultee and provides performance feedback.

Both direct and indirect assessments of treatment integrity have their merits and limitations (Gresham, 2014; Smith, Daunic, & Taylor, 2007). In terms of accuracy, direct assessment yields more accurate data than indirect assessment because indirect assessment has the tendency of inflated estimation. However, little research is available to guide the amount of direct observation needed to make sure data are representative. Additionally, both direct assessment and indirect assessment may subject to reactivity. For direct assessment, the effects

of observer reactivity may occur when an implementer is aware of being observed. Likewise, for indirect assessment, self-reports may cue the implementer to correspond in his/her do-say behavior. Pertaining to the ease of assessment, direct assessment is labor intensive whereas indirect assessment is less intensive. However, the ease of assessment may trade off important information. For example, permanent products only reveal the integrity regarding the content but not the process.

To accurately evaluate the relation between treatment integrity and outcome, reporting treatment integrity should encompass overall integrity, component integrity and session integrity (Gresham, 1997; Schloesser, 2002). Overall integrity refers to the extent to which all components were implemented across sessions. Component integrity reflects the integrity of each component implemented across sessions. The session integrity represents the integrity of all components within one session. Perepletchikoa (2014) explained that the rationale for all three estimates of treatment integrity is that first, high in overall integrity may not indicate high integrity in the other two areas. Ignorance of component integrity and session integrity may lead to inconsistent treatment delivery and failure to produce a positive outcome, especially when overall integrity was high. Second, closely monitoring all three areas of treatment integrity boosts the credibility and replicability of the treatment. Third, such monitoring enables fine-grained analysis to identify the mechanisms of change.

Variables related to treatment integrity. Variables that may affect treatment integrity include treatment characteristics, implementer characteristics, client characteristics, and organization/environment characteristics (Perepletchikova, & Kazdin, 2005; Sanetti & Kratochwill, 2009). Treatment characteristics such as complexity, time, materials/resources required and the number of implementers (e.g., teachers, parents) may impact treatment integrity

(Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000). In general, the higher the degree of complexity in these variables is, the lower treatment integrity will be. The type of treatment also influences treatment integrity. For example, skill-based psychotherapies show greater treatment integrity than non-skill-based (Perepletchikova, Treat, & Kazdin, 2007). Further, effectiveness of a treatment may impact treatment integrity. An implementer is more likely to have greater treatment integrity when a treatment results in rapid changes than when a treatment shows slow improvements.

Among implementer characteristics, experience, motivation, and perception about the treatment are critical variables that affect treatment integrity (Perepletchikova, & Kazdin, 2005). An experienced implementer is more likely to deviate from the original treatment than an inexperienced one because of a tendency to integrate different techniques. Factors associated with the treatment or the client may also reduce the implementer's motivation to maintain high treatment integrity, such as negative outcomes or client resistance. Moreover, when the implementer perceives a treatment is acceptable (i.e., social validity; see Social Validity section for a detailed discussion), he/she is more willing to deliver the treatment with integrity.

With regard to client characteristics, a client's attitudes, problem behavior, and treatment history may impact treatment integrity (Perepletchikova, & Kazdin, 2005). A client's resistance, anger or hostility in a session can impede relationship building with the implementer and the implementer's performance. Treatment integrity may decline when the client has severe problem behavior and a long duration of treatment needs.

Organizations/environments play a vital role in enhancing treatment integrity. In a review of 81 empirical studies on prevention and health programs for children and adolescence, Durlak and DuPre (2008) found that the organizations with characteristics such as positive work

climate, shared decision-making, effective leadership, and managerial support facilitated program implementation.

Research on treatment integrity and its barriers. Despite its importance, the dearth of treatment integrity in Check & Connect reflects the general lack of attention to treatment integrity in intervention research across fields. Several reviews revealed a "curious double standard" phenomenon in which researchers make careful efforts to operationally define and measure dependent variables but do not apply the same efforts to independent variables (Peterson, Homer, & Wonderlich, 1982). For example, in psychotherapy research, Perepletchikova, Treat, and Kazdin (2007) examined psychosocial interventions published in six top journals in the field of psychiatry or clinical psychology from 2000-2004. Rating results from the Implementation of Treatment Integrity Procedures Scale (ITIPS) indicated that only 3.5% of the studies provided adequate treatment integrity. In education, Sanetti and colleagues (2011, 2012, and 2014) conducted a series of treatment integrity reviews of school-based interventions across six journals in school psychology (i.e., *Journal of School Psychology* [JSP], *Psychology in the Schools* [PITS], *School Psychology Quarterly* [SPQ], *School Psychology Review* [SPR], *School Psychology International* [SPI]) and special education (i.e., *Journal of Positive Behavior Interventions* [JPBI]) from 1995-2010. Results indicated that only one third of the studies in the five school psychology journals operationally defined the interventions (which is essential for treatment integrity) whereas 59% of the studies did in JPBI. Further, of 321 studies met the inclusion criteria, half did not report treatment integrity (61.5% for SPI).

In their review, Ledford and Wolery (2013) included 14 journals targeting interventions for students with disabilities using single-subject designs and expanded the time frame from 1980 to 2010. The researcher found that only 44% of the studies (N= 1215) reported treatment

fidelity, although a variable and ascending trend was seen over the years. In addition, the researchers indicated that a majority of the studies failed to stipulate when treatment fidelity was conducted, for what behavior, and for which participants.

Some common barriers perceived by researchers may explain the low prevalence of treatment integrity in the intervention research. Perepletchikova, Hilt, Chereji, and Kazdin (2009), and Sanetti and Reed (2012) conducted an online survey of the corresponding authors of the research articles identified in their previous reviews (i.e., Perepletchikova, Treat, Kazdin, 2007; Sanetti, Gritter, & Dobey, 2011). The respondents from both surveys rated the lack of theory and guidelines as well as the demands of time, cost, and labor as the two strongest perceived barriers to assessing treatment integrity, even though the respondents indicated that they appreciated the importance of treatment integrity. Additionally, Perepletchikova and associates found that the more barriers a researcher perceived the lower treatment integrity and that the lack of editorial/publication requirements was a contributing barrier to treatment integrity as well.

Social Validity

Historically, the concept of social validity originated from the field of applied behavior analysis (ABA), which underscores the examination of socially important behaviors (Baer, Wolf, & Riseley, 1968). The concept of social validity was first introduced to the field by Wolf's (1976, 1978) and Kazdin's (1977) seminal papers. Van Houten further developed the concept (1979; see Foster & Mash, 1999 for a brief review).

Construct of social validity. Social validity is a multi-dimensional construct (Foster & Mash, 1999). Wolf (1978) first defined social validity as consumers' subjective judgments of an intervention at three levels: (a) whether the goals of an intervention are socially significant, (b)

whether the procedures of an intervention are socially appropriate and acceptable, and (c) whether an intervention results in socially important and satisfactory outcomes. Later, Fawcett (1991) further elaborated on Wolf's definition. Fawcett asserted that socially valid goals should comprise broad social goals, that behavioral categories should be related to the broad social goals, and that the specific response should be linked to the behavioral categories. Likewise, the validation of procedures should specify the comprehensive intervention used, whether an intervention package(s) and/or specific procedures as part of an intervention package. Further, the validation of outcomes should include proximal effects, intermediate effects related to main effects, and distal effects. Contrary to Fawcett, researchers Schwartz and Baer (1991) broadened the construct of social validity from the standpoint of consumers. In addition to direct recipients of an intervention as consumers, they asserted that consumers should broadly include those who are not direct recipients but have direct or indirect impact on the viability of an intervention. Therefore, they expanded the traditional scope to four categories: (a) direct consumers, (b) indirect consumers, (c) members of the immediate community, and (d) members of the extended community. Direct consumers are those who receive an intervention and directly affect the viability of an intervention through continuous participation or refusal to participate (e.g., students participating in Check & Connect). Indirect consumers are those who refer or purchase an intervention for direct consumers, or those who may be affected by the change targeted in an intervention (e.g., teachers or school personnel of students participating in Check & Connect). Despite non-recipients of an intervention, indirect consumers' satisfaction or refusal also affects the viability of an intervention. Members of the immediate community, such as peers or bus drivers, are those who have regular contact with both direct and indirect consumers and affect the viability of an intervention indirectly through interaction or lack of

28

interaction with the direct and indirect consumers. Members of the extended community are those who live in the same community but may not have contact with both direct and indirect consumers.

Although many terms have been used interchangeably to refer to the concept of social validity, such as "consumer satisfaction," "treatment acceptability," "educational relevance," and "clinical significance," (Foster & Marsh, 1999), Hawkins (1991) argued that the term, "consumer satisfaction" was more accurate in capturing the process of gathering consumers' opinions than "social validity." Hence, Hawkins suggested using "habilitative validity" to truly reflect validating the goals, the procedures, and the outcomes of an intervention. He further elucidated the concept of habilitative validity by defining habilitation or adjustment as the extent to which an intervention brings forth the maximum benefits and minimizes the costs to an individual and to his/her significant others. In short, social validity involves judgments from a variety of consumers (i.e., direct and indirect consumers and immediate and distal community members) pertaining to whether the goals, the procedures, and the outcomes of an intervention result in maximizing the benefits and minimizing the cost for the client and his/her significant others.

Importance of social validity. Although neither a primary measure nor required for all studies, particularly for "bridge research" (Kennedy, 2005), social validity is important for intervention research. As Baer, Wolf and Risely (1987) noted, "…social validity is not sufficient for effectiveness but is necessary to effectiveness" (p. 323). There are a number of reasons social validity is important (Hawkins, 1979).

From the perspective of promoting an intervention, a socially invalid intervention can result in consumers' complaints or termination regardless of its effectiveness (Wolf, 1978). Assessing social validity allows researchers or service providers to identify the undesired

components. Furthermore, by knowing consumers' likes and dislikes about the intervention, particularly those with well-established effects, researchers or service providers can provide procedural adjustments. Meanwhile, dislike for effective interventions may reflect the misconceptions of direct consumers or the general public; therefore, it suggests a need for further clarification and education. Since consumers know best about their environment, assessing social validity provides researchers or service providers insight about resources available for enhancing habilitation. For accountability and comprehensiveness of intervention effects, social validity, particularly subjective validation, is a feasible and economical way to assess across clients and different points of time. Meanwhile, through social validity, research can monitor if consumers remain satisfied with the intervention program. In addition, on-going social validation offers researchers or service providers a panorama of intervention information, particularly when complaints remain after previous issues have been resolved.

From the perspective of ethics, although behavioral interventions do not require social validity to meet the dimensions or standards of ethics, it may guide the selection of target behaviors and procedures that seek the best interests of the client, rather than the benefits of authority figures (e.g., parents, teachers) (Fuqua & Schwade, 1986). Further, social validity increases the likelihood that the practices of behavior analysis align with societal wants and values (Adkins, 1997). That is, effective interventions may not be accepted by society or consumers (e.g., certain types of punishment). Therefore, interventionists should select effective procedures that will not infringe the client's rights yet have high acceptability (Kazdin, 1981).

Assessment of social validity. Assessing social validity is best viewed as a process (Houten, 1979), rather than a static point assessment. The assessment could occur prior to, during or after intervention depending on the purpose (Winett, Moore, & Anderson, 1991). In

30

this process, researchers and consumers play an interactive role. Researchers design and implement an intervention, while consumers evaluate the use of the intervention. With consumers' feedback and input, researchers may further revise an intervention for the future. Meanwhile, different types of consumers are informed or educated about the importance of an intervention for a problem behavior by giving an opportunity to evaluate an intervention (Finney, 1991).

Social validation of goals, procedures and outcomes can be established through three different methods: subjective evaluations, objective evaluations, and experimental/functional analysis (Fuqua & Schwade, 1986). However, subjective evaluations for procedures referred to as treatment acceptability are predominantly used by researchers (Finn & Sladeczek, 2001).

Subjective evaluations. Subjective evaluations involve using a Likert rating scale with various points (e.g., seven) indicating the degree of agreement or conducting interviews to gather consumers' options about the goals, procedures and outcomes of an intervention.

For goal validation, qualified judges, such as experts, professionals, competent peers or significant others in an individual's life, are solicited to evaluate whether the selection of a target behavior is appropriate, important, or based on social criteria (Fuqua & Schwade, 1986). For procedure validation, relevant consumers indicate to what extent the procedures of an intervention are acceptable on a rating scale and then their responses are tallied to scores representing the overall acceptability (Finn & Sladeczek, 2001). In general, the higher total scores indicate the higher acceptability.

Owing to Kazdin's pioneering research in treatment acceptability in the 1980s, several rating scales with adequate psychometric properties had been developed (Finn & Sladeczek, 2001), such as the Treatment Evaluation Inventory (TEI; Kazdin, 1981), the Intervention Rating

Profile-20 (IRP-20; Witt & Martens, 1983), the Children's Intervention Rating Profile (CIRP; Witt & Elliott, 1985), and the Behavior Intervention Rating Scale (BIRS; Von Brock & Elliott, 1987). The TEI is a 15-item questionnaire with items rated on a 7-point Likert scale. The questions address the acceptability, fairness, effectiveness, side effects of the treatment, consumers' willingness to carry out the treatment and consumers' overall reactions to the treatment. The IRP-20 is a 20-item questionnaire with questions rated on a 6-point Likert scale. In addition to the shared commonalities with the TEI, the IRP-20 underscores the feasibility of implementing an intervention in a large size class (i.e., over 30 students) and considers relevant resources required such as training, time and technical skills. Different from other treatment acceptability scales, the CIRP is specifically designed to assess children's perceptions of a treatment. The CIRP contains seven questions rated on a 7-point Likert scale. The questions were written at the fifth grade level and also cover issues pertaining to fairness, effectiveness and over acceptability (e.g., "The method used to deal with the behavior problem was fair."). The BIRS contains 24 items questions rated on a 6-point Likert scale that represents a three-factor structure: acceptability, effectiveness, and time to effectiveness (Elliot & Treuting, 1991).

For outcome validation, after intervention, representative consumers or experts appraise any meaningful behavior change in an individual on the rating scale. An alternative method is to ask consumers to review an individual's video samples before and after intervention, evaluate any improvement in the individual, and indicate change on a rating scale.

Objective evaluations. Unlike the use of self-report or report by another in subjective evaluation, objective evaluation involves using objective indices, such as social comparison to selected goals and performance criteria and observable behaviors associated with satisfaction for procedural validation. Social comparison is to compare an individual's performance with his/her

32

typical peers' before and after intervention (Kazdin, 1977). Standards or norms can be established by randomly selecting representative samples or by using convenience samples in an individual's immediate environment (i.e., local norms; Foster & Mash, 1999). By comparison, an individual's deviation may warrant the need for an intervention and the importance of an intervention goal can be justified. After intervention, norms also serve as a yardstick for evaluating whether an intervention results in acceptable outcomes.

Observable behaviors associated with consumer satisfaction such as participation, attendance, affect, office disciplinary referrals and advocacy can be used as adjunctive measures with ongoing data collection for procedure validation (Fuqua & Schwade, 1986; Schwartz & Baer, 1991). When satisfied with an intervention, a consumer is more likely to continue participating in the program and attend the sessions regularly and promptly. During sessions, a satisfied consumer is likely to display positive affect and enthusiasm and is even eager to earn incentives that are set up for the intervention program. Furthermore, outside of sessions a satisfied consumer is more likely to advocate for the intervention and recommend it to others. On the contrary, when dissatisfied with an intervention, a consumer is more likely to have poor attendance, tardiness, non-compliance or even early withdrawal from the program. During sessions, a dissatisfied consumer is more likely to manifest negative affect, non-compliance, poor performance in earning incentives or requests for personnel change. Outside the sessions, a dissatisfied consumer is more likely to complain to others or the media. Using observable behaviors associated with satisfaction is beneficial for consumers who have limited verbal skills to express their dissatisfaction, particularly for people with developmental disabilities or those who are reluctant to disclose dissatisfaction frankly due to courtesy.

Experimental/functional analysis. This approach involves experimentally manipulating

the target behavior of interest/alternative performances or testing alternative interventions. There

are two methods of experimentally assessing goals and outcomes (Houten, 1979). The first

method is to test the degree to which the target behavior of interest will result in the optimal

performance/outcome at a low cost. For example, to identify the appropriate goal for teaching

preschoolers to share materials with their peers, Rogers-Warren and Baer (1976) examined the

frequency of preschoolers' offerings at low, median and high levels on their peers' acceptance.

The authors found that only when preschoolers' offerings were presented at a median level did

they result in the highest percentage of peers' acceptance. This type of experimental validation,

however, is rarely employed by researchers (Houten, 1979). Furthermore, identifying the

optimal level of a target behavior is more frequently researched in the literature than identifying

a target behavior (Fuqua & Schwade, 1986).

The second method is to scrutinize different performance standards on the acquisition of

more complex or subsequent skills, particularly when a target behavior is prerequisite skill.

Houten (1979) illustrated the use of accuracy vs. rate in determining the termination of single-

digit multiplication drills needed to improve long multiplication and division problems. In the

experiment, the students sometimes received the single digit multiplication drill prior to taking a

test on the long multiplications. Results showed that the students made very few errors during

the drills and their performances on the test improved in the drill conditions, whereas their

performances remained the same in no drill conditions. When solely examining the accuracy of

basic multiplication facts, one would conclude that the students did not require further training.

However, when examining the rate of solving the long multiplication/division problems, the

students' performances drastically increased with the drills. With respect to social validation of

procedures, effective intervention procedures are more socially valid than ineffective procedures, especially when experimentally comparing effective and sophisticated intervention procedures with one another (Hawkins, 1991).

Conceptual models of social validity. Several conceptual models have been proposed to explain the relations among the key variables (Eckert & Hintze, 2000). The first model was developed by Witt and Elliot (1985) and illustrated a sequential and reciprocal relationship among four components: (a) treatment acceptability, (b) treatment use, (c) treatment integrity, and (d) treatment effectiveness (see Figure 1). In this model, if a treatment is deemed as acceptable then it would likely be used, and if the integrity of implementation is high it is more likely to result in behavior changes. With the experience of treatment effectiveness, the consumer's acceptability is more likely to increase accordingly.

Reimers, Wacker, and Koepple (1987) later extended Witt and Elliot's model to a decision-making model (see Figure 2). The model posited a directional relationship among six elements (i.e., knowledge/understanding, acceptability, integrity, effectiveness, maintenance and environmental disruption) and provided recommendations for each situation (i.e., treatment education, treatment modification or a new treatment). First, the model assumed that understanding of a treatment is a premise before assessing acceptability. Lack of treatment knowledge would lead to poor treatment implementation and treatment outcomes and consequently, treatment education should be provided. Second, the model depicted that good understanding of a treatment would lead to high acceptability, consistent implementation and positive effects, and the behavioral change would be maintained or varied depending on the degree of the environmental disruption. Also, the behavioral change may be limited when the treatment effectiveness is low even if the treatment is perceived as acceptable and implemented

35

with integrity. As varied or low behavior maintenance occurs, the treatment would be modified or a new treatment would be proposed. In contrast, even when a treatment is well understood, it may be deemed as unacceptable and then implemented poorly. Hence, the treatment effect would be comprised. In such a case, treatment modification would be made or an alternative would be recommended.

Unlike the previous two models that limited social validation to treatment acceptability, Lane, Beebe-Frankeberger and Lambros (2001) proposed a self-reinforcing model that integrated goal validation, generalization and habilitation (see Figure 3). The self-reinforcing model hypothesizes a directional and recursive relationship among the significance of a treatment goal, treatment acceptability, treatment integrity, generalization and maintenance. Namely, if a treatment goal is deemed significant, the probability of a treatment being perceived as acceptable and implemented as designed would be high. When treatment integrity is high, the treatment is more likely to be used across settings, which facilitates the behavior change over time (i.e., maintenance) and across various people/settings (i.e., generalization). Consequently, the consumer's habilitation increases, which further reinforces the significance of a treatment goal.

Evaluation of Social Validity in Research

Trends. Social validity research was prosperous during the 1980s and the 1990s; however, this endeavor did not continue to develop and expand in scope even with the increasing shift from ABA to SWPBS in the field of special education. From the perspective of measuring social validity, Kennedy (1992) examined the trends in two renowned behavioral journals. In the *Journal of Applied Behavior Analysis (JABA)*, a total of 125 studies published in 1968-1990 reported social validity data. The trend of reporting social validity increased since the mid-1970s and after a high peak in 1983, the percentage of studies with social validity measures declined.

36

In the *Journal of Behavior Modification*, a total of 53 studies published in 1977-1990 measured social validity. The percentage of studies reporting social validity increased in the late 1970s and decreased in 1980. In 1985-1990, approximately 20% of the studies published in both journals reported social validity. Subsequently, Carr, Austin, Britton, Kellum, and Bailey (1999) re-examined the publication trend of social validity in JABA and expanded Kennedy's work by extending the publication year from 1968 to 1998, measuring the frequency of social validity in treatment outcome and treatment acceptability. The review revealed a similar trend as Kennedy's, showing an increase in reporting treatment outcome measures, treatment acceptability measures and the combination of the two measures in the 1970s to mid-1980s. Afterward, reporting either treatment outcome or acceptability declined to approximately 25% of research articles. Additionally, in the 1990s, the percentage of studies reporting social validity conducted in analogue settings decreased, whereas the studies conducted in natural settings reported social validity twice as frequently as those conducted in analogue settings.

Although studies conducted in natural settings are more likely to measure social validity, the proportion reporting social validity has remained small over the years. Along with the two reviews discussed, Clarke and Dunlap (2008) further compared the differences of reporting social validity among three peer-reviewed journals (i.e., JABA, the *Journal of Positive Behavior Interventions* [JPBI], *Education and Training on Mental Retardation and Developmental Disabilities* [ETMRDD]) that focus on intervention research for children and youth with disabilities from 1999-2005. Of the three journals, even though JPBI had the highest average percentage of research articles with social validity measures, it was only 31% compared with 20% for ETMRDD and 3% for JABA. Over the years, the percentage of research articles with social validity measures in JPBI remained small and steady as evidenced in the review by O'Dell

et al. (2011). O'Dell et al. analyzed the trends of studies published in JPBI in 1999-2008 and

found that of 118 studies reviewed only 39 studies (i.e., 33%) assessed social validity. In

contrast to behavioral based interventions, measuring social validity is completely absent in the

mentoring literature.

        Variables that influence social validity. Current literature shows that the majority of

studies employed factorial designs to investigate variables that affect teachers' and students'

acceptability ratings. In general, researchers manipulated variables of interest and presented

them in written cases (or read out loud) for teachers and students to evaluate on a rating scale

(e.g., the BIRS, the IRP, the TEI and the CIRP). The variables investigated included

intervention-related variables, target child-related variables, and the rater-related variables.

These influencing variables will be discussed from the teacher aspect and the student aspect.

        Role of treatment variables and the target child's characteristics on teachers'

acceptability ratings. Teachers favor certain interventions over others depending upon the target

student's behavior severity. For example, Harris, Preller, and Graham (1990) had two groups of

teachers rate acceptability on the IRP after reading a vignette that described two behavioral

interventions (i.e., praise and token economy) and two cognitive-behavioral interventions (i.e.,

self-monitoring and self-instruction) applied to a student either with a mild or a severe problem

behavior. One hundred fifteen teachers rated the case with a mild problem behavior (e.g.,

assignment incompletion and off-task) and 87 teachers rated on the case with a severe problem

behavior (e.g., fighting and talking out loud). A 2x4 factorial analysis of variance (ANOVA)

showed a significant main effect for type of interventions and a significant interaction between

type of interventions and behavior severity. Specifically, teachers viewed all four interventions

as acceptable for both cases. However, teachers deemed that using self-instruction was more

acceptable for a severe problem behavior than for a mild one, while using praise was more acceptable for a mild problem behavior than a severe one.

When dealing with severe problem behaviors, teachers prefer interventions described in a jargon free manner. In their study, Witt, Moe, Gutkin, and Andrews (1984) presented an intervention for two hypothetical cases: one for a mild problem behavior (i.e., assignment incompletion) and one for a severe problem behavior (i.e., fighting at recess) to 112 elementary teachers. The intervention was described in three ways: the behavioral description (i.e., emphasis of contingent applications), the humanistic description (i.e., emphasis of understanding feelings), and the pragmatic description (i.e., emphasis of natural consequence of behavior). After reading the written descriptions for the two cases, the teachers rated their acceptability on the IRP. A 2x3 factorial ANOVA revealed that the pragmatic description was significantly more acceptable than the other two descriptions and the interventions were more acceptable when applied to the severe case than to the mild one.

Additionally, in addressing problem behaviors, teachers prefer to collaborate with a professional to develop an intervention that focuses on reinforcing behaviors, but implement it by themselves. In the study by Kutsick, Gutkin, and Witt (1991), 240 elementary teachers rated their acceptability of an intervention on the Intervention-Process Rating Scale (IPRS) based on a written hypothetical case. The case description contained: (a) an intervention, either positive (i.e., reinforcement of incompatible behavior) or reductive (i.e., response cost); (b) an intervention development process through teacher-psychologist collaboration, psychologist's development, or teacher's development; and (c) the target student's disruptive behavior, being either mild (i.e., 20 min instructional time lost per day) or severe (i.e., 90 min instructional time lost per day). Results from a 3x2x2 factorial ANOVA showed that despite no significant effect

39

for behavior severity, teachers significantly favored a collaborative process and a positive intervention. Further, follow-up analysis revealed that teachers judged a reductive intervention developed through collaboration as significantly more acceptable than developed individually.

In the study by Witt and Robbins (1985), 98 teachers read a case of a student with mild or severe problem behavior (i.e., assignment incompletion or fighting at recess) in which either a teacher had the student stay in at recess or a principal sent the student to his office at recess. Teachers then rated their acceptability on the IRP. Results indicated that teachers judged that the two interventions were significantly more acceptable for the severe problem (i.e., fighting) than the mild. Moreover, teachers viewed the teacher-implemented intervention as significantly more acceptable than the principal-implemented.

In general, teachers favor a positive intervention that requires low teacher time and skill and meanwhile, poses no risk or negative impact on the target student and other students. However, with the increase of the target student's behavior severity, teachers tend to prefer an intervention with moderate teacher time involvement and view interventions for severe problems as more acceptable. In their hypothetical case, Witt, Martens, and Elliot (1984) manipulated two types of interventions that required varied teacher time to address a target student's problem behavior with varied severity. The types of interventions consisted of positive (i.e., praise, reinforcement and token economy) and reductive (i.e., ignoring, response cost and seclusion time out) interventions. The teacher-time involvement included three levels: low amount (i.e., less than 30 min per day), medium amount (i.e., 1-2 hr per day for preparation and 30 min to 1 hr per day for maintaining the intervention), and high amount (i.e., more than 2 hr for preparation and more than 1 hr per day for maintaining the program). The target student's behavior severity included low level (i.e., daydreaming behavior), moderate level (i.e., using obscene language)

and high level (i.e., property destruction). After reading the written case, 180 teachers rated their acceptability on the IRP. A three-way fixed-factorial ANOVA indicated a significant main effect for teacher-time involvement and significant interaction effects for teacher-time involvement by behavior severity and intervention type by time involvement. Namely, teachers judged interventions with a higher level of time involvement as less acceptable. Further, compared to mild and moderate behaviors, teachers perceived using interventions with low time involvement to address severe behaviors as significantly less acceptable. Teachers also perceived low level of time involvement as most acceptable for positive interventions, while medium level of time involvement as most acceptable for reductive interventions.

Witt, Elliot, and Martens (1984) conducted a similar study with 180 pre-service and student teachers, using the same design as the Witt, Martens, and Elliot (1984) study except that teachers' acceptability on the five subscales (i.e., general acceptability, risks to the target student, risks to other students, teacher time required, and teacher skill required) of the IRP was analyzed. A three-way multivariate analysis of variance (MANOVA) and further univariate analysis showed that teachers perceived positive interventions as significantly more acceptable, less risky, and requiring less teacher time than reductive interventions. In addition, in the subscale of risks for other students, teachers viewed spending more time to implement an intervention for a mild behavior problem as detrimental for other students.

In a two-experiment study, Elliott, Witt, Galvin, and Peterson (1984) examined the effects of intervention complexity and behavior severity on experienced teachers' acceptability. In Experiment 1, 71 teachers read a total of nine written cases that illustrated an intervention (i.e., praise, reinforcement or token economy) applied to a student with a problem behavior (i.e., daydreaming, obscene language or property destruction). The complexity of the intervention and

behavior severity varied in the same manner as in the Witt, Martens, and Elliot study.

Afterward, teachers rated their acceptability on the IRP. A two-way fixed factorial ANOVA indicated that praise and reinforcement (mild to medium complexity) were significantly more favorable than token economy (high complexity). Despite no significant effect for behavior severity, teachers rated using praise, the least complex intervention to address a mild behavior problem like daydreaming, as the most acceptable, while using token economy, the most complex intervention for the same mild problem as the least acceptable. In Experiment 2, the same method was employed with 70 teachers except that the interventions were altered to ignoring, response cost and seclusion time-out. A two-way ANOVA indicated that response cost, with a medium level of complexity, was rated significantly more favorably than ignoring and time out, with low and high levels of complexity. In addition, response cost was the most acceptable intervention for the most severe problem behavior (i.e., property destruction).

Different from the aforementioned studies, Cowan and Sheridan (2003) examined the impact of behavior severity and intervention complexity on teachers' post-treatment acceptability for conjoint behavioral consultation. In the study, 67 teachers received conjoint behavioral consultation (CBC) to address problem behaviors of their students at risk for academic failure or students with BD, LD or ADHD. Prior to consultation, parents and teachers indicated students' behavior severity on a 7-point research-based rating scale (1 = mild; 7 = very severe). Intervention complexity was defined by the number of interventions used in the consultation cases, which included home-school communication notes, self-monitoring, training (e.g., social skill training, parent training or modeling) and reductive consequences (e.g., time out, ignoring or loss of privileges). At the last consultation meeting, 62 teachers completed acceptability rating on the BIRS. Pearson correlation showed a significantly positive association between

teachers' acceptability and students' behavior severity. Further, a hierarchical regression analysis revealed that behavior severity and intervention complexity significantly explained the variance in teachers' acceptability at a modest level (R^2 = .16). However, only teacher-rated behavior severity showed a significant effect.

In addition to time involvement, teachers favor interventions with strong effectiveness. Clark and Elliott (1988) investigated information about intervention effectiveness (i.e., weak vs. strong outcomes) on 133 elementary teachers' acceptability ratings of modeling and overcorrection applied for a student who displayed withdrawal or aggressive behaviors. Based on written vignettes, teachers rated their acceptability on the BIRS. A 2 x 2 x 2 MANOVA revealed significant main effects for intervention and effectiveness. The teachers' ratings were higher when modeling or overcorrection was presented with strong effects than when presented with weak effects.

When considering the target student's behavior severity, teachers perceive that providing effectiveness information for an intervention is more acceptable in dealing with a mild behavior than a severe behavior. In their study, Von Brock and Elliott (1987) further examined the effects of intervention effectiveness information, behavior severity and intervention type on teacher ratings of acceptability. A total of 216 teachers also read written vignettes that described interventions (i.e., token economy, response cost, and time out) used for problem behaviors (i.e., low vs. high severity) and effectiveness information (i.e., no information, consumer satisfaction information, and research-based outcome information). Teachers rated their acceptability on the BIRS, which was analyzed on the three subscales representing three factors: acceptability, effectiveness (e.g., level of change, generalization) and time of effectiveness (i.e., how quickly a behavior will improve). A 3x3x2 MANOVA showed that token economy and response cost

were significantly more acceptable, effective, and time efficient than time out. Further, for a mild behavior, teachers deemed an intervention with research-based effectiveness information as more acceptable, effective, and time efficient than an intervention without any information, whereas for a severe behavior, the effectiveness information did not affect teachers' ratings.

Other than behavior severity, the target student's intelligence, popularity, duration of problem behavior, and gender all influence teachers' acceptability. As shown in an experiment by Martens and Meller (1989), 163 teachers read vignettes that illustrated an intervention (i.e., home-based reinforcement vs. response cost) used to address a student with varied intelligence (i.e., average vs. below average), sociability (i.e., popular vs. unpopular) and length of problem behavior (i.e., two weeks vs. two years) and judged their acceptability on the IRP. A 2x2x2x2 ANOVA of the IRP's three subscale scores (i.e., general acceptability, practicality and skill required for implementation) indicated that in the acceptability factor, home-based reinforcement was significantly more favorable than response cost. Moreover, teachers perceived interventions were more acceptable for the student with average intelligence than with below average intelligence. However, the meaningfulness of the main effect in intelligence was reduced because a significant interaction effect for intelligence by sociability was found, revealing that teachers perceived interventions as less acceptable when applied to popular students with below average intelligence. For the practicality factor, teachers judged home-based reinforcement as more practical than response cost. With regard to the skill required for implementation, teachers deemed that interventions for chronic problem behaviors required significantly more skills to implement than interventions for behaviors with short duration. Further, teachers perceived less skill requirement when applying interventions to students with average intelligence and short

duration of problem behavior, whereas, regardless of students' intelligence level, teachers viewed more skill requirements for interventions applied to chronic problem behaviors.

Likewise, Piscecco, Huzinee, and Curtis (2001) examined intervention type, ADHD subtype, and the target student's gender on 159 elementary teachers' acceptability on the three factors of the BIRS (i.e., acceptability, effectiveness and time efficiency). The three independent variables were also manipulated and presented in six written vignettes for teachers to read and evaluate. Each vignette contained an intervention (i.e., daily report card, response cost, medication, or classroom lottery) for a boy or a girl with varied ADHD symptoms (i.e., hyperactive-impulsive, inattentive, or combined type). A 3x2 MANOVA showed a significant interaction effect for intervention type by gender. Further univariate ANOVAs indicated that teachers considered using daily report card as more acceptable and time efficient with girls than boys. Teachers also viewed response cost as more acceptable, effective and time efficient for girls than for boys. In contrast, teachers perceived medication as more acceptable and time efficient for boys than for girls.

Interaction of treatment types and teacher characteristics on teachers' acceptability ratings. Teachers' knowledge of a specific disability affects their acceptability for medication and more experienced teachers tend to rate interventions less acceptable than inexperienced teachers. Power, Hess, and Bennett (1995) examined teachers' knowledge and years of teaching experience on teachers' acceptability for behavioral and pharmacological interventions for students with ADHD. Seventy-six elementary teachers and 71 middle school teachers who had experience in teaching students with ADHD completed the ADHD Knowledge Scale, a 17-item measure with a true-false option. Teachers also rated their acceptability for three interventions used in the written vignettes for a student diagnosed as ADHD. The three interventions were a

daily report card procedure, response cost, and medication. A one-way repeated ANOVA indicated that both teacher groups perceived daily report cards as more favorable than the other two interventions. Chi square (x^2) analysis showed that both teacher groups perceived daily report card with medication as more acceptable than response cost with medication, and medication alone was rated the least favorable. Further simultaneous regression analyses revealed a partial correlation between ADHD knowledge and acceptability. Specifically, only middle school teachers' knowledge about ADHD negatively correlated with their acceptability for medication; however, the researchers deemed the significance might be due to a chance. Despite a significant difference in years of teaching experience for both teacher groups, no significant correlation was found between teachers' years of teaching experience and their acceptability for either behavioral interventions or medication.

Vereb and DiPerna (2004) investigated the relations among teachers' knowledge, treatment acceptability, teaching experience, and training in ADHD. Forty-seven elementary teachers completed two surveys: the Knowledge of ADHD Rating Evaluation (KARE) and a professional experience questionnaire. The KARE is comprised of four sub-scales: Knowledge of ADHD, Knowledge of Treatment, Medication Acceptability, and Behavior Management Acceptability. The first two sub-scales required responses of true, false or don't know options, while the last two sub-scales employed a 4-point Likert scale format (i.e., 1 = not at all likely; 4 = very likely). The professional experience questionnaire contained questions regarding teachers' teaching and teaching experiences regarding ADHD. Results of the correlations among the sub-scales of the KARE showed two significant correlations. Teachers' knowledge of ADHD and their acceptability of medication were positively related (.37), while teachers' knowledge of treatment and their acceptability were negatively associated (-.32). The

46

correlations between scores in the KARE and the professional experience questionnaire indicated that teachers' teaching experience with ADHD significantly correlated with their acceptability for medication (.40). In addition, teachers' training in ADHD was positively associated with their knowledge of ADHD (.49) and their acceptability for both medication (.31) and behavior management (.34).

Witt and Robbins (1985) conducted two experiments to examine the impact of teachers' years of teaching experience with interventions and child behavior severity on their acceptability ratings. In Experiment 1, 196 teachers were classified in two groups based on their years of teaching: low experience (below 8 years) and high experience (above 8 years). Both teacher groups read written scenarios and rated their acceptability on the IRP. The written scenarios depicted six interventions (i.e., differential of other behaviors [DRO], differential reinforcement of low rates of responding [DRL], reprimands, time out, staying after school, and corporal punishment) used to address a problem behavior with varied severity (i.e., mild, moderate, and severe). In Experiment 2, written cases with same interventions were used as well as grouping based on teaching experience. Ninety-eight teachers rated the IRP, except that in the written cases, the behavior severity was altered to two levels (mild vs. severe) and a description of an interventionist (either a teacher or a principal) was added. Factorial ANOVAs showed that in both experiments, teachers with low teaching experience perceived the interventions were more acceptable than those with high teaching experience. Moreover, in Experiment 1, teachers with high teaching experience had higher acceptability ratings than teachers with low teaching experience when interventions (except staying after school and reprimand) were applied to severe behavior problem.

47

With respect to other teacher variables, teachers' group status as special educators or general educator seems to have no impact on teachers' acceptability ratings. However, when compared with other professional groups, teachers' acceptability ratings differ. In two experiments, Epstein, Matson, Repp, and Helsel (1986) first examined the impact of group status on teachers' treatment acceptability. A total of 89 teachers who enrolled in a special education course (27 special educators and 62 regular educators) reviewed a written case that described five common interventions used for a student with ADHD, including medication, behavior modification, counseling, special education, and affective education. After reading the case, teachers evaluated each intervention and rated their acceptability on the Treatment Evaluation Inventory (TEI). A 2x5 ANOVA showed special education program was rated as most acceptable and medication as the least acceptable. The ratings between special educators and regular educators showed no difference and there was no interaction effect for teacher group by intervention. In the second experiment, Epstein and colleagues investigated the target student's diagnosis on the treatment acceptability of 77 college students enrolled in special education courses. The college students also read a written case describing a student with either MR or LD receiving the same five interventions used in Experiment 1 and rating acceptability on the TEI. As in Experiment 1, special education program was rated most favorably while medication was the least acceptable. Additionally, the label of the target student did not affect the college students' ratings.

Fairbanks and Stinnett (1997) further compared the effects of group status and the target student's label on intervention acceptability among 31 teachers, 33 school psychologists and 33 social workers. All participants rated their acceptability on the IRP after reviewing written scenarios describing verbal praise or token economy applied to address the problem behavior of

a student with LD, BD or ADD. A 2x3x2 ANOVA revealed significant main effects for interventions and group. Namely, verbal praise was rated significantly more favorably than token economy, and teachers had higher ratings than school psychologists and social workers. Furthermore, a significant interaction effect for intervention by group indicated that the three groups perceived verbal praise similarly. However, teachers judged token economy as significantly more acceptable than school psychologists and social workers, and school psychologists deemed token economy significantly more acceptable than social workers.

Interaction of treatment variables, target child variables, and teachers' characteristics on teachers' acceptability ratings.   Two studies examined all three types of variables on teachers' acceptability. Stinnett, Crawford, Gillespie, Cruce and Langford (2001) examined the impact of high school location, target student's label and intervention types on the treatment acceptability. A total of 114 college students who majored in elementary teacher education (72 graduated from rural high schools, 71 graduated from urban high schools) evaluated medication and special education placement applied to a student with or without the label of ADHD and then rated their acceptability on the IRP. Additionally, college students completed the Teacher Rating Scale, a 5-point Likert scale (0 = never to 4 = excessively) to assess participants' attribution of severity of the target student's problem behavior (i.e., attention vs. social problems). Three (i.e., the IRP scores and two sub-scale scores on the TRS) 3-way ANOVAs (location x label x intervention) showed that on the IRP, college students who graduated from urban high schools perceived interventions as significantly less acceptable than those graduated from rural high schools. In addition, on the attention problem scores, college students perceived that the target student without the ADHD label had significantly fewer attention problems when in special education placement than in medication condition. When the intervention was held constant in

49

special education placement, college students judged that the target student with the ADHD label exhibited more attention problems than the student without the label. Further, on the social problem scores, college students perceived that the target student with the ADHD label had significantly fewer social problems than one with the label. College students who graduated from urban high schools perceived that the target student displayed significantly more social problems in the medication condition than in the special education placement. Compared to those who graduated from rural high schools, participants graduating from urban high schools perceived the target student as having more social problems when receiving medication.

Schneider, Kerridge, and Katz (1992) investigated the effects of teacher gender and group status on their acceptability for the intervention used for two cases. A total of 53 teachers participated in the study. Twenty-five teachers (9 general educators, 6 special educators in a school setting, and 10 special educators in a hospital setting) evaluated a hypothetical case illustrating seven interventions applied to a withdrawn boy. The interventions included modeling, coaching, problem solving, token economy, medication, play therapy and family therapy. Twenty-eight teachers (13 general educators, 5 special educators in a school setting, and 10 special educators in a hospital setting) evaluated another hypothetical case describing eight interventions (the aforementioned interventions and time out) used for an aggressive girl. After listening to audiotapes, teachers rated their acceptability on the TEI. Three repeated ANOVAs were conducted: separate analysis for each case and combined case analysis after excluding the TEI scores for time out. Results showed that with the exception of coaching, family therapy was more acceptable than the other interventions and medication was rated as the least acceptable. In addition, coaching was more acceptable for the aggressive girl than the withdrawn boy. Compared to the two groups of special educators, general educators perceived

50

coaching and token economy as less acceptable. Regarding teacher gender difference, male teachers viewed coaching and token economy as more acceptable than female teachers. Male teachers also perceived coaching and medication as more acceptable when applied to the withdrawn boy than female teachers.

   Treatment type and the target child's behavior severity on students' acceptability ratings. Compared to studies on teachers' acceptability, the research base on students' acceptability is relatively small. However, research suggests that students do have unique perspectives and preferences that are different from adults. Elliott, Turco, and Gresham (1987) compared the acceptability for group contingencies among 217 fifth graders, 140 school psychologists, and 45 experienced teachers. Three case scenarios described three types of group contingency (i.e., dependent, independent, and interdependent) applied to three classroom behavior conditions (i.e., only two students, half of the class, and the whole class who misbehaved). After reading the scenarios, students rated their acceptability on the CIRP, whereas both school psychologists and teachers rated theirs on the CIRP and IRP. Four-way repeated MANOVAs revealed that fifth graders perceived all three types of group contingencies as slightly acceptable (i.e., all the means slightly higher than the cut-off scores for acceptable). In contrast, adults showed a distinct preference. Specifically, adults deemed the independent group contingency as the most acceptable while the dependent group contingency as the least acceptable. No rater difference was found among the adult groups.

   Elliott, Witt, Galvin, and Moe (1986) conducted two studies to understand sixth graders' perspectives regarding the impact of behavior severity and interventions on their ratings of acceptability. In Study 1, Elliott and associates surveyed a sample of 23 sixth graders. The questionnaire contained 20 questions about the degree of behavior severity (18 on behavior

problems and 2 on prosocial behaviors) and 12 interventions that teachers used to address talking

out of turn or property destruction. The sixth graders evaluated the behavior severity on a 4-

point rating scale (1 = not a problem to 4 = a very big problem) and the interventions on a 5-

point rating scale (1 = I dislike it very much to 5 = I like it very much) along with a free

response. Results of a dependent *t*-tests and a Q-sort analysis showed that the sixth graders

judged work incompletion, swearing, cheating on a test, and playing with a knife as very

problematic. The sixth graders also viewed property destruction significantly more severe than

talking out of turn. Pertaining the interventions, the sixth graders significantly preferred using

traditional, negative contingencies (e.g., losing recess, going to the principal's office) applied

directly toward the misbehaved individual rather than involving peers and using positive

interventions. Additionally, the sixth graders' free responses suggested interventions of relying

on an outside authority or excluding the misbehaved student from current activities. In Study 2,

Elliott and associates created three written cases to manipulate the degree of behavior severity

and types of interventions on the acceptability of 79 sixth graders. Each case consisted of a

problem behavior (property destruction or talking out of turns) with one of four intervention

types in each category applied (i.e., four verbal, four reinforcement and four traditional). Upon

finishing reading the cases, the sixth graders rated their acceptability on the CIRP. Results from

a 2x4 factorial ANOVA showed a significant main effect only for intervention. Further follow-

up contrast analyses indicated that the sixth graders preferred interventions directly applied to the

misbehaved individual using negative verbal (e.g., correcting the misbehavior privately) or

negative reinforcement (e.g., losing recess when misbehaving). In addition, the traditional,

negative interventions (e.g., staying in from recess, going to the principal's office, or going to a

quiet room for few minutes) were more acceptable for an individual displaying property

destruction than talking out of turn. In contrast, when an intervention was applied to the class, the sixth graders preferred positive group reinforcement (e.g., the class got extra time for recess when the individual behaved). In another words, the sixth graders preferred interventions that punished the individual but reinforced the group.

Turco and Elliott (1986a) examined intervention type and the target student's behavior severity on the treatment acceptability of a total of 144 general education students (46 at fifth grade, 48 at seventh grade and 50 at ninth grade). The experimenter read aloud a description of a hypothetical student who either disturbed others or destroyed others' property and received one of the eight interventions: public or private reprimand, public or private praise, self-monitoring for frequency of being reprimanded or praised, and home-based reprimand or praise. Afterward, the students rated their acceptability on the CIRP. A 2x8 factorial ANOVA revealed a significant main effect for intervention. Namely, the students considered praise for appropriate behaviors at home as the most acceptable followed by reprimand for misbehaviors at home, private reprimand and public praise, while public reprimand was the least acceptable.

Interaction among treatment types, the target child characteristic and the rater's characteristics on students' acceptability ratings. Turco and Elliott (1986b) further explored other variables that might affect student raters' acceptability, including the target student's gender, the student rater's own behavior and the student rater's grade level. Procedures included reading hypothetical cases aloud to 151 general education students (45 at fifth grade, 52 at seventh grade, and 54 at the ninth grade). The cases described either a boy or a girl who received one of the four interventions (i.e., public praise or reprimand, home-based praise or reprimand) because of disturbing others or destroying others' property. After the students rated their acceptability on the CIRP, their teachers rated each participating student's classroom

53

behavior as above average or below average. Results from a 4 (intervention) x 2 (target student's gender) x 2 (student rater's gender) x 3 (student rater's grade) ANOVA indicated that similar to the Turco and Elliott (1986a) study, public reprimand was rated the least favorable while home-based praise was rated the most favorable. Specifically, both the fifth and seventh graders viewed home-based praise as significantly more acceptable than other interventions. However, the ninth graders considered both home-based praise and reprimand as significantly more acceptable than the other two interventions. Furthermore, only the ninth graders who were rated by their teachers as below average in the classroom behaviors perceived all the interventions were significantly more acceptable for a misbehaved girl than a misbehaved boy.

Taken together, the findings indicate that characteristics related to the intervention, the target child, and the teacher/student rater affect teachers' and students' acceptability ratings. In general, for teachers, positive interventions (e.g., praise, modeling, home-based reinforcement, daily report card, token economy) were more favorable than reductive interventions (e.g., response cost, time out, overcorrection, medication). Teachers also favor interventions characterized as effective, less complex, and less time consuming. However, when taking the target child's characteristics into consideration, particularly, behavior severity, teachers would be willing to spend more time and use more complex interventions or collaborate with others to develop reductive interventions. With respect to other target child variables, teachers' acceptability varied depending on the child's IQ, popularity, duration of problem behavior and gender. However, the effect of labels is inconsistent across studies (i.e., Fairbank & Stinnett, 1997; Stinnett et al., 2001). Pertaining to teacher-related variables, despite that the correlational direction is inconsistent across studies (i.e., Power, Hess, & Benett, 1995; Vereb & DiPerna, 2004), teachers' knowledge of ADHD is associated with their acceptability for medications. In

addition, novice teachers tended to have higher acceptability ratings than seasoned teachers, while seasoned teachers were more acceptable of interventions applied to severe behavior problem than novice teachers. Demographic variables such as gender and high school graduation location affected teachers' acceptability ratings. Male teachers had higher acceptability ratings for coaching and token economy than female teachers. Teachers who graduated from rural high schools had higher acceptability than teachers who graduated from urban high schools. Additionally, compared with other professionals (e.g., school psychologists and social workers), teachers had higher acceptability ratings and particularly showed favor in token economy.

For students, negative or home-based interventions were more favorable when applied directly to the misbehaved individual, whereas positive group reinforcement was more acceptable when applied at school to the whole class. Although the target child's behavior severity did not show a significant effect on students' ratings, in general, students found interventions more acceptable when applied to more severe behavior (i.e., property destruction) than mild behavior (i.e., talking out of turn). Regarding student rater related factors, only older students (i.e., the ninth graders) perceived reprimand at home as acceptable and took the target child's gender into consideration when evaluating interventions.

Despite that many variables have been identified, several limitations deserve further research. First, the majority of the studies were pre-treatment and analogue in nature, with the exception of the study by Cowan and Sheridan (2003). Although analogue studies have the merit of easily manipulating potential variables of interest, without further investigation it is unclear whether the variables identified from analogue studies would remain influential for direct consumers after they experience the intervention. Second, compared to adults, studies on

55

students' acceptability are relatively few. In addition, the majority of participating students in the studies were typically developing. As a result, the perceptions of typical students may not represent the perspectives of students with problem behavior. Third, even though Cowan and Sheridan (2003) involved authentic direct consumers (teachers and students) in their study, the researchers did not examine the impact of teacher-related factors on the teachers' acceptability. As the results showed that only students' behavior severity was significant and explained 16% of the variance in teachers' acceptability, whether including teacher-related variables as additional predictors would increase the explanation of the variance in teachers' acceptability requires further investigation. Moreover, despite positive student acceptability ratings, the lack of students' self-reported behavior severity restricted the researchers' ability further to examine the extent to which students' behavior severity accounted for students' own acceptability ratings. Along the lines of students' acceptability, the role of teacher related factors on students' ratings have not yet been examined in the social validity literature. Understanding the possible impact of teacher-related factors, such as years of teaching experience, can be particularly critical for the mentoring-based interventions. Fourth, although several variables have been investigated, the potential influence of treatment dosage on acceptability ratings has not yet been examined. This aspect is particularly critical for social validity research in natural settings and the mentoring literature given that maintaining consistent treatment dosage is challenging in those settings and is vital for the mentor-mentee relationship. Fifth, regardless of analogue or natural studies on social validity, current literature has not yet systematically examined whether social validity ratings are consistent over different time points.

Relation between social validity and treatment integrity. Despite the conceptual models of social validity, currently only five studies have examined the relation between

treatment integrity and social validity. Of five studies, three include behavioral interventions

(i.e., Ehrhardt, Barnett, Lentz, Stollar, & Reifin, 1996; Miramontes, Marchant, Heath, & Fischr,

2011; Sterling-Turner & Watson, 2002) and two include academic interventions (i.e., Allinder &

Oats, 1997; Mautone, DuPaul, Jitendra, Tresco, Junod, & Volpe, 2009). In general, the

preliminary data indicate that treatment integrity and social validity may be aligned. However,

whether treatment integrity and social validity are significantly correlated and, if so, the strength

of the correlation remains unclear.

Unlike the other four studies utilizing group design, Ehrhardt et al. (1996) conducted four

consultation case studies on the use of scripts to promote preschool teachers' or parents'

treatment integrity and acceptability. The consultants, two certified school psychologists, used

the structured problem solving procedures to assist the consultees to identify the target behaviors

and create the implementation scripts. After baseline observation, the consultees followed the

scripts to implement the interventions. The consultants also followed the consultees'

implementation scripts to measure their treatment integrity (i.e., percentage of steps

implemented). Additionally, the consultants used the Script Acceptability Questionnaire (SAQ),

a 5-point Likert scale (1= strongly disagree, 5 = strongly agree) with eight items, to assess the

consultees' acceptability weekly or bi-weekly. Each consultee's treatment integrity data and one

item score on the SAQ (i.e., "The script deals effectively with the problem.") were co-plotted

along with each child's target behavior. The results showed that the five children's target

behaviors improved. The consultees agreed or strongly agreed that the scripts were effective and

maintained 40-100% treatment integrity.

Along with the majority of social validity studies, Sterling-Turner and Watson (2002)

adapted an analogue approach to investigate the relation between treatment integrity and social

validity. A total of 64 undergraduate students enrolled in educational psychology participated in the study. The participants first rated their pre-treatment acceptability of a treatment protocol prescribed for a client with tic disorder on the IRP-15. Then, the first researcher trained the participants to implement the treatment protocol described for the pre-treatment acceptability rating. The treatment protocol included antecedent procedures (e.g., reminding), consequence procedures (i.e., DRO and punishment), and data collection. After training, each participant implemented the treatment protocol to a confederate client (participants were unaware of deception) in a treatment room for four 5-min sessions. The participants' implementation was videotaped and treatment integrity was evaluated on a checklist. After completing all treatment sessions, the participants rated their acceptability on the modified IRP-15 again. On average, the participants implemented 69% of the procedures correctly and showed an increase in their post-acceptability ratings (i.e., $M = 66.41$ for posttest; $M = 59.88$ for pretest). Contrary to researchers' hypothesis of a moderate to high correlation, Spearman rank-order analysis showed no significant correlation between treatment integrity and pre/post treatment acceptability (i.e., $r_s = .001$ for pretreatment; $r_s = .13$ for post-treatment acceptability).

Miramontes, Marchant, Heath, and Fischer (2011) conducted a survey to understand the social validity of a statewide PBS initiative and the relation between school implementation integrity and respondents' acceptability. A total of 270 participants, including teachers, administrators, and service providers recruited from a convenience sample (i.e., during a state conference) rated their acceptability on the researcher-generated questionnaire, which contained 18 items anchored on a 5-point Likert scale. Overall, the survey results indicated that the majority of teachers, administrators, and service providers perceived that the PBS initiative had a positive impact in their schools and would recommend the program to others. Miramontes et al.

further analyzed the correlation between the school annual treatment integrity measured by the School-Evaluation Tool (SET; Sugai, Lewis, Palmer, Todd, & Horner, 2001) version 2.0 and the participants' responses to each questionnaire item. Spearman correlation showed that treatment integrity was significantly correlated with six social validity items. Namely, the SET scores were positively associated with respondents' (a) use of strategies/interventions, (b) perceived positive impact, (c) staff consensus/buy-in, (d) satisfaction with school's universal procedures, (e) satisfaction with Tier 1 procedures, and (f) satisfaction with Tier 3 goals and procedures. Regardless, the correlation was small (range .118 to .195).

Expanding the social validity literature from behavioral to academic interventions, Allinder and Oats (1997) examined whether teachers who perceived curriculum-based measurement (CBM) as acceptable would increase their use of CBM. Twenty-one elementary and special education teachers rated their acceptability on the CBM Acceptability Scale (CBM-AS), a 20-item, 6-point Likert scale with good psychometric properties. The researchers rated the teachers' implementation integrity on a 5-point Likert scale (i.e., 0 = none of changes timed correctly to 4 = each adjustment timed in accordance with the decision rule) at the end of the study based on the students' CBM graphs. Based on the mean scores on the CBM-AS, the researchers divided the teachers into high acceptability group (mean score above 5) and low acceptability group (mean score below 5). Then, the researchers compared the differences between the two groups on each item of the CBM-AS and in treatment integrity. A series of *t*-tests revealed that teachers in the high acceptability group had a significantly higher degree of viewing CBM as suitable/beneficial for students and practical to implement. Teachers in the high acceptability group were also more willing to use CBM and recommend it to others. Further, MANOVA analysis indicated that teachers in the high acceptability group conducted

more CBM probes and set more ambitious goals for their students than teachers in low acceptability group.

Mautone et al. (2009) investigated the relation between treatment integrity and social validity of two reading consultation models (i.e., intensive vs. traditional data-based intervention) for children with ADHD. During intervention, consultants used treatment integrity checklists to observe teachers' implementation. At the end of the intervention phase, a total of 73 teachers from the two intervention groups rated their acceptability on the modified BIRS. Pearson product moment correlation showed that although the BIRS total score and mean treatment integrity was positively correlated for the two models, the correlation was not significant (i.e., r = .30 for the intensive model; r =. 35 for the traditional model; r = .32 for the entire sample). Further MANOVA analysis indicated that teachers in the intensive model had significantly higher treatment integrity than teachers in the traditional model. However, no significant differences in either group's acceptability were found even though the score in the intensive group was slightly higher.

The literature indicates limitations in examining the relation between treatment integrity and social validity. First, a dearth of empirical studies is available to elucidate the relation and the strength of treatment integrity and social validity. Second, even among the existing studies, methodological shortcomings limit interpretation of the results. For example, it is difficult to discern if treatment integrity and acceptability were positively associated in the Ehrhardt et al. (1996) study given the nature of case study, the report of only one acceptability item score, the lack of psychometric properties or the acceptability questionnaire, and the lack of statistical analysis. In addition, the lack of significant correlation in the Sterling-Turner and Waston (2002)

study and the Mautone et al. (2009) study may be due to small sample size. Therefore, further empirical support for the relation between treatment integrity and social validity is needed.

Summary

In sum, many high school students with high-incidence disabilities encounter difficulties in school engagement. Research shows that Check & Connect, a mentoring-based intervention is a promising approach to promote the resiliency and school engagement of secondary students with disabilities. However, the reports of its treatment integrity and social validity are absent in the literature. Further review of the literature shows an overall low prevalence rate in reporting treatment integrity and social validity. The literature also reveals the needs for further advancing social validity research in the following areas: (a) continuing to investigate the influencing variables in natural settings, (b) expanding social validity research to Tier 2 interventions, (c) assessing social validity for an extended period of time, (d) involving students in social validity rating, and (e) examining the relation between treatment integrity and social validity. Therefore, this study aims to investigate the variables that affect mentors' and students' social validity ratings and the relation between mentors' social validity ratings and their treatment integrity.

Chapter III

Method

Background

This study utilized data from the Center for Adolescents Research in Schools (CARS; Kern, Evans, & Lewis, 2011) to investigate the relation between the social validity of Check & Connect and its influencing variables. CARS was a 5-year Center Grant funded by the Institute of Education Sciences (IES), Department of Education to address the poor outcomes of secondary students with emotional/behavioral disorders (EBD). During the first 3 years, CARS developed a comprehensive intervention package through a multidisciplinary approach and five stages of iterative development. The first stage was the initial intervention development, focusing on identifying and piloting the best available interventions in education and mental health that addressed three conceptual areas in (a) school and teacher capacity, (b) youth competence, and (c) family and community supports. The second stage was the preparation for implementation, focusing on further piloting the core interventions and preliminary assessment strategies identified in the first stage. The second stage of piloting also further reduced the three conceptual areas in the first stage to two foci for intervention: classroom teachers and school mental health professionals (e.g., school psychologists, counselors and social workers). The third stage was to further evaluate the interventions in a larger and diverse sample and gather feedback for refinement. Specifically, the team provided one-on-one coaching and performance feedback to train school personnel and meanwhile, the school personnel assisted the team to identify the key components of interventions for treatment integrity assessment. The fourth stage focused on further refinement based on the data from the interventions and assessments. The

final stage focused on refining and testing a fully manualized intervention package with diverse sample.

During the last two years, the comprehensive package was tested in a randomized control trial (RCT) with 54 high schools across Kansas, Missouri, Ohio, Pennsylvania and South Carolina. Schools were randomly assigned to a treatment or wellness condition. Students in the treatment condition received a comprehensive intervention package. The comprehensive package was comprised of interventions that addressed the academic and social functioning support, individualized classroom support, and individualized mental health support. First, for the academic and social functioning support, all students received Check & Connect and a weekly Interpersonal Skills Group. Second, for individualized classroom support, when a student manifested disengagement in school based on the Check & Connect data (e.g., frequent absences, office referrals; see Check & Connect section below in detail), the grant staff conducted a classroom assessment, including teacher interview and direct observations and then recommended specific interventions from which teachers selected one or more that they deemed as feasible and acceptable for improving student behavior. The specific interventions included (a) clearly defining classroom expectations and routines; (b) providing students specific praise and performance feedback to increase student-teacher interactions; (c) using response cards, guided notes and peer tutoring to increase opportunities to respond; (d) providing students appropriate testing and assignment accommodations; and (e) teaching students to self-identify precursors to problem behavior and initiate de-escalation strategies to prevent behavioral escalations. Third, for individualized mental health support, grant staff assisted school mental health professionals to conduct assessments (e.g., student and parent interview and rating scales) to identify interventions for students who displayed distress or impairment related to depression,

63

anxiety, ADHD, and conduct and oppositional defiant disorders. Specific mental health interventions included cognitive behavioral therapy (CBT), self-management, and parent education. For students in the wellness condition, monthly newsletters were sent to their parents/guardian, teachers, and mental health providers with helpful tips about problems that adolescents frequently encounter (e.g., drug use, gangs).

Measures used during the RCT included distal outcome measures, proximal outcome measures, intervention acceptability, and student and family characteristics. Distal outcome measures included the use of: (a) the Woodcock Johnson-III (WJ-III; Woodcock, McGrew, & Mather, 2001) for academic skills, (b) the Impairment Rating Scale (IRS) for school functioning, (c) the Reynolds Adolescent Depression Scale (RADS-2; Reynolds, 1987), (d) the Multi-Dimensional Anxiety Scale for Children (MASC; March, Parker, Sullivan, Stallings & Conners, 1997), (e) the Youth Risk Behavior Survey (YRBS; Center for Disease Control, 1990) for behavioral, emotional and mental health symptoms, and (f) the Brief Multidimensional Student's Life Satisfaction Scale (BMSLSS; Seligson, Huebner, & Valois, 2003) for overall life satisfaction. Proximal outcomes measures consisted of Check & Connect data (i.e., grades, attendance, behavioral infractions), classroom direction observations, school permanent records, and the Classroom Performance Survey (CPS). Intervention acceptability measures included teachers' and students' social validity ratings, measured by the School Intervention Rating Form (SIRF; see below section for detail). Student and family characteristics assessed during an intake interviews to gather individual and school level demographics. In addition, the Services Assessment for Children and Adolescents (SACA; Hoagwood et al., 2000) and the Services for Children and Adolescents-Parent Interview (SCAPI; Jensen et al., 2004) were combined and administered to gather information regarding school/community services received.

64

This study focused on social validity ratings and treatment integrity for Check & Connect during the RCT (i.e., administered at the end of Years 1 and 2 of the RTC). Details of the participating students, mentors, Check & Connect implementation and the measures were described as below.

Participants

Students. To participate in the larger study, students had to meet the following criteria: (a) absence of Autism Spectrum Disorders (ASD) or intellectual disability (ID), (b) displayed impairments in social/emotional/behavioral and school functioning, and (c) had a parent/guardian who could speak English fluently. The rationale for excluding students with ASD and ID was because some interventions (e.g., cognitive behavior therapy) required cognitive abilities. The impairment in social/emotional/behavioral functioning was determined by the cutoff *T*-score on one of the three standardized measures: (a) 60 or above on either the internalizing or externalizing composites of the Behavior Assessment System for Children-Teacher or Parent Version (BASC), (b) 60 or above on the Multidimensional Anxiety Scale for Children (MASC), or (c) 50 or above on the Reynolds Adolescent Depression Scale 2 (RADS-2). Impairment in school functioning was defined as exhibiting any two of the following indicators: (a) *office referrals-* four or more times during the semester prior to enrollment, or five or more times during the current semester, (b) *poor school participation-* five or more non-illness related absences or tardiness to class during any month of the current or previous semester, (c) *school suspensions-* two or more in-school or out-of-school suspensions in the current academic year, or (d) *failing grades-* one or more Fs, or two or more Ds in any core academic subject in one of two most recent grading periods. A total of 647 students consented to participate in the study, 341 in the treatment group and 306 in the control group (Figure 5). Of the students, 66.5 % was male,

and the major ethnicity makeup was White/Caucasian (52.1 %) followed by African American (38.6 %), Hispanic/Latino (5.3 %), and other (4%). Further, 48.5 % of the students had a special education classification with, 21.6 % identified as LD, 9.9 % as ED, 8.5 % as OHI, and 8% as unknown.

Mentors. A total of 234 mentors participated in Check & Connect, 153 in Year 1, 156 in Year 2, and 75 in both years. Seventy-six percent of the mentors were teachers, and of the teachers, 20 % served as a case manager. Other makeup of mentors included counselors/school psychologists (4.26 %), social workers (2.65 %), administrators (2.12 %), and other professionals such as coaches and school security officers (10.58 %). Those mentors were recruited because they manifested willingness and the following the following characteristics: (a) willing to persist with students for one year; (b) believe in students' strengths and abilities; (c) willing to collaborate and cooperate with families and grant staff; and (d) good skills in advocacy, organization, and case management (CARS Classroom-Based Interventions Manual, 2011).

Components of Check & Connect

Check & Connect contained two major components: the check part and the connect part. For the check part, mentors utilized a Check & Connect monitoring sheet (see Figure 6) to document students' risk indicators (i.e., behavior fraction, grades, school participation) and provide feedback to students. For the connect part, based on the monitoring data, mentors assisted students to use the five-step strategy to problem-solve risks: (a) Stop! Think about the problem, (b) What are some choices? (c) Choose one, (d) Do it, and (e) How did it work?

Procedures of Check & Connect

Mentor training. Grant staff employed a coaching model to train mentors. During an initial 15-20 min meeting, grant staff first explained the rationale for and the format of Check &

Connect, and then provided examples and non-examples. Grant staff also modeled and role-played with mentors. In addition, grant staff showed a video of a partial mentoring session to mentors. Finally, grant staff answered questions that mentors raised at the end of the training.

Implementation. After training, mentors were responsible for using a Check & Connect monitor sheet (see Figure 6) to obtain students' risk indicators data (i.e., frequency of behavior fraction and missing assignments, and grades), and holding individual meetings with students for at least 10 min once per week.

Treatment fidelity checks. Initially, grant staff assessed mentors' treatment integrity each week for 3 weeks within the first month through direct observation or permanent product. Once mentors met integrity criteria, set at above 80% during the third check, grant staff faded integrity checks to once a month. However, when mentors failed to meet the criteria of 80% accurate implementation, grant staff problem-solved with mentors. The problem-solving process (see Figure 7) began by determining the need for re-training (e.g., didn't understand the procedures, needed reminders). Subsequently, additional training or supports were used that were matched to the reason for the fidelity lapse, such as providing additional explanations, in vivo reminders, modeling, or making intervention adjustments. After providing re-training, grant staff collected three additional integrity checks. When the integrity on any of the observations was below the set criteria of 80%, grant staff scheduled an additional problem-solving meeting. One additional booster session was conducted. If staff again failed to meet integrity, the intervention was discontinued.

Measures

Measures used in this study included the Teacher Rating Scale (TRS) of the Behavior Assessment System for Children, Second Edition (BASC-2) at Year 2 of the RCT, Check &

Connect treatment integrity at Year 2 of the RCT, mentors' School Intervention Rating Form (SIRF) at Year 2 of the RCT, students' SIRF at Year 1 and 2 of the RCT, Check & Connect sessions participated, mentors' years of experience, change in mentors and diagnostic status.

TRS, BASC-2. The BASC-2 is a norm-referenced, standardized behavior rating scale for children and youth aged from 2 to 25 (Reynolds & Kamphaus, 2004). The TRS contains 139 items that represent five composites: Externalizing Problems, Internalizing Problems, School Problems, Behavioral Symptoms Index and Adaptive Skills. A teacher who knows a student well but may not be a mentor rates a student's behavior on a four-point Likert scale (i.e., 1 = Never to 4 = Almost always). The BASC-2 has moderate to good validity and reliability. The manual reports that the construct, convergent, and discriminant validity of the TRS with other behavioral rating scales (e.g., Conners' Teacher Rating Scale-Revised) were high in the .70s and .80s. The coefficient alpha reliability was in .90s for the composite scales and in .80s for the individual scales. The correlation of test-retest reliability was in .80s for the composite scores and between .70s and .80s for individual scores. *T*-scores of Externalizing and Internalizing Problems were used as a measure of students' behavior severity.

Check & Connect treatment integrity. Treatment integrity was recorded on a checklist that delineated core implementation steps (e.g., meeting with the student, recording risk indicators; see Figure 4) for all sessions across each student. The integrity checklist was developed during the development phase, field-tested and refined until it measured the core components. The treatment integrity was measured by percentage. The weekly percentage was calculated by dividing the total number of steps implemented by the total steps possible during the week and multiplying by 100%. The monthly percentage was calculated by averaging the

weekly percentage. For the purpose of analysis, only the monthly grand mean (i.e., average of all monthly means) for each student were used.

Mentor SIRF.   Mentor SIRF is a 21-item survey on a 7-point Likert scale (e.g., 1 = Not at all acceptable; 4 = Somewhat acceptable; 7 = Very acceptable) coupled with four open-ended questions. The SIRF assesses mentors' treatment acceptability regarding understanding, effectiveness, cost, time, willingness, and side effects. The SIRF was adapted from the Treatment Acceptability Rating Form-Revised (TARF-R; Reimers & Wacker, 1988) to reflect application of Check & Connect while still maintaining the fundamental elements of questions. For example, "How clear is your understanding of the suggested procedures?" on the TARF-R was revised to "How clear is your understanding of the intervention after having used with your student/s?" on the SIRF. The TARF-R contains 20 questions on a 7-point anchor and measures the acceptability dimensions in effectiveness, side effect, cost, willingness and reasonability. The TARF-R has good psychometric properties. The overall internal consistency coefficient is .92 and the coefficients for composites range from .69 to .95 (Finn & Sladeczek, 2001). The internal consistency reliability of Mentor SIRF is at .75 and considered adequate. Harrison, State, Evans and Schamberg (2015) further assessed the psychometric properties of the SIRF and found that the construct of the SIRF contains three factors: Suitability, Perceived Benefit and Convenience. The internal consistency is .83 for Suitability, .82 for Perceived Benefit, and .68 for Convenience.

For analyzing mentors' overall acceptability of Check & Connect, only the total score of each Mentor SIRF was used. Possible scores range from 21 as the lowest, 84 as middlemost to 142 as the highest, and the higher the score the higher degree of acceptability. Because in some cases one adult served as a mentor for several students, to ensure independent measures for

further statistical analysis, a mentor-student pair was randomly selected when a mentor rated his/her SIRF based on multiple students (see Data Aggregation section).

Student SIRF. Student SIRF is a six-item survey on a 7-point Likert scale (e.g., 1 = Did not like; 4 = Like some; 7 = Like a lot) that, similar to the Mentor SIRF, assesses students' treatment acceptability for Check & Connect pertaining to understanding, like, helpfulness and side effect. The SIRF was also adapted from the TARF-R by retaining relevant questions to students and revising the wording accordingly. For example, "How much discomfort is your learner likely to experience as a result of these procedures?" was changed to "Did anything about the intervention make you feel uncomfortable?" The internal consistency coefficients of Student SIRF are .77 and .80 in Year 1 and 2 respectively.   For the purpose of analysis, only each student's total score was used. Possible scores range from four as the lowest (because no score is assigned to two questions when an N/A is indicated), 23 as middlemost to 42 as the highest. Similar to the interpretation of Mentor SIRF, the higher the score indicates the intervention as more acceptable.

Check & Connect sessions participated (Dosage). Because mentors were required to meet with students weekly, the dosage of Check & Connect that a student received was measured by total number of weekly sessions participated divided by the total available sessions (e.g., 36 weeks for an academic year). For example, if a student had 20 sessions with a mentor in Year 2; therefore, the dosage for the student would be 20 out of 36 weeks in the school year, that is, 56%.

Mentors' years of experience. The total years in the position that a mentor held was used to measure mentors' experience.

Change in mentors. Because students' change in mentors in Year 2 of the RCT was a binary variable, it was coded as "1" for change or "0" for no change by comparing if mentors' IDs in each year were identical.

Diagnostic status. Because students' diagnostic status is a binary variable, it was coded as "1" for with special education/psychiatric label(s) or "0" for no label(s).

Data Analyses

Data analyses were conducted on SPSS 23 and SAS 9.4. Analysis procedures included: (a) power analysis, (b) data aggregation, (c) data screening, (d) attrition analysis, (e) multiple imputation, and (f) multiple regression for each research question. Each procedure was described as below.

Power analysis. A priori power analysis was calculated on G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) assuming a medium effect size of f^2 level at .15, power of 0.8 and alpha level of .05. A minimal sample size required for Research Question 1 analysis was 77, for Research Questions 2 and 3 analyses was 92, and for Research Question 4 analysis was 55.

Data aggregation. Although the data set contained five waves (i.e., Wave 1-3 in Year 1 and Wave 4-5 in Year 2), students' and mentors' measures in SIRF, dosage and treatment integrity were not available consistently across each wave. Therefore, to obtain yearly scores, Year 1 data were defined as measures assessed within Wave 1-3 and Year 2 as Wave 4-5. Then, data were further aggregated based on three conditions. First, in a situation when participants had one measure within Wave 1-3 and another one within Wave 4-5, regardless of whether the student-mentor pair was or was not the same, the former was considered as Year 1 score and the latter as Year 2 score. Second, for participants who had repeated measures with the same

student-mentor pair, the scores within Wave 1-3 were averaged as the Year 1 scores, while the scores within Wave 4-5 were averaged as the Year 2 scores. Third, for participants who had repeated measures with different student-mentor pairs within Wave 1-3 or Wave 4-5 (e.g., Student A with Mentor A at Wave 2 and then with Mentor B at Wave 3), to ensure independent observations, an online random picker was used to randomly select a pair to represent the scores in Year 1 or 2. The specific steps involved were: (a) all different student-mentor ID pairs were listed on an Excel spreadsheet; (b) when the Student SIRF score was available, the mentor's ID on the Student SIRF determined the corresponding scores for dosage and treatment integrity; (c) when the Student SIRF was missing or when two pairs appeared in Year 2, the scores for dosage and treatment integrity were randomly selected.

Data screening. After data aggregation, missing data patterns and statistical assumptions were screened on SPSS 23. Percentage of missing data in each variable ranged from 1.2% for special education status to 62.2% for mentors' years of teaching. Little's missing completely at random test (MCAR; Little, 1988) was then conducted to examine the missing mechanism (Rubin, 1976): (a) missing complete at random (MCAR), which is that the missing data of a variable(s) do not depend on the missing or observed scores of the variables; (b) missing at random (MAR), which is that the missing data of a variable depend on other observed measures in the analysis model but not on the scores of the variable itself; or (c) missing not at random (MNAR), which is that the missing data of a variable are related to the scores of the variable itself controlling for other variables. The MCAR test revealed significance ($\chi^2(250) = 408.08$, $p < .001$), suggesting that the missing pattern was not MCAR.

Statistical assumptions were checked for each research question, including linearity, multicollinearity, outliers, normality of residuals, and homoscedastic residuals. Overall,

correlation matrices for all predictors showed greater than zero and less than .80, indicating

linear relations among predictors and no collinearity. Collinearity diagnosis statistics also

indicated no multicollinearity: the average variance inflation factor (VIF) was close to 1

(Bowerman & O'Connell, 1990), and tolerance was greater than .20 (Menard, 1995). Despite

that one to four outliers were detected in some variables, Cook's distance values were less than

1. Therefore, no outliers were removed from the dataset. For residuals, histograms showed that

the distribution was slightly skewed; however, standard residual plots showed random patterns,

except for the mentors' treatment integrity.

Attrition analyses. To understand the characteristics of students who stayed throughout

the intervention and those who dropped out in the first year of intervention, attrition analyses

were conducted using SPSS 23. Chi-square analyses were applied to examine the differences in

demographic variables (i.e., gender, ethnicity, special education diagnosis) between stayers and

dropouts, while independent *t*-tests were performed to scrutinize the differences in their behavior

severity on the BASC-2, dosage, and Year 1 SIRF.

Multiple imputation. Multiple imputation (MI) was conducted given that the MCAR

test was significant, and multiple imputation has been recommended to handle missing data

based on the assumption that the missingness was MAR (Schafer & Graham, 2002).

Specifically, the fully conditional specification (FCS) with the predictive mean matching (PMM)

was selected as the imputation algorithm to generate 40 datasets ($m = 40$) using SAS 9.4. The

rationale for the FCS was that the dataset contained categorical variables and may not have joint

multivariate normality (IDRA, 2015). Further, because of the high amount of missing

information (MI) for the mentors' years of teaching variable, more than five datasets were

imputed (Graham, Olchowski, & Gilreath, 2007). Despite slight skewness observed in some

variables, MI remained robust in the presence of non-normality (Dong & Chao-Ying, 2013).

Therefore, no transformation was made to the original dataset. The imputation model included

all predictors and dependent variables (Alison, 2001). Due to complete missing values for two

cases, each imputed dataset only contained 339 cases. Once datasets were imputed, data analysis

for each research question proceeded, and only pooled results were reported.

Research Question 1. **Do students' social validity ratings in Year 1, the percentage**

of Check & Connect sessions participated, and change in mentors predict student social

validity ratings in Year 2? A simultaneous multiple regression was conducted. The outcome

variable was the total scores of the Student SIRF in Year 2, while the predictor variables was the

total scores of the Student SIRF in Year 1, the percentage of Check & Connect sessions

participated over the course of intervention and the change of mentors at the end of Year 2. Due

to the natural of a categorical variable, the change of mentors was dummy coded (i.e., 0 = no

change in mentors as a reference group; 1 = change in mentor). The R^2 and the b weights were

examined.

Research Question 2. Does students' behavior severity, measured by the teacher-

rated BASC-2 (i.e., TRS), diagnostic status, mentors' years of teaching experience and the

percentage of Check & Connect sessions participated predict students' social validity

ratings in Year 2? A hierarchical multiple regression analysis was conducted. The outcome

variable was the total scores of the Student SIRF, and predictors were Externalizing and

Internalizing Problems Composite T scores of the TRS, diagnostic status (i.e., no label vs. with

label), mentors' years of teaching experience and percentage of Check & Connect sessions

participated at the end of Year 2. Diagnostic status was also dummy coded due to the nature of a

categorical variable (i.e., 0 = no label as a reference group; 1= with label). The predictor

variables were entered into the regression model in three steps: Step 1- percentage of Check & Connect sessions at Year 2, Step 2- Externalizing and Internalizing Problems Composite T scores of the TRS and diagnostic status, Step 3- mentors' years of teaching experience. The rationale for such a hierarchy was first to control for the variability of dosage that students received and then to examine the impact of student characteristics followed by teacher characteristics. The R^2 and the increment in the R^2 at each new step were examined.

Research Question 3. Does students' behavior severity, measured by the teacher-rated BASC-2 (i.e., TRS), students' diagnostic status, mentors' years of teaching experience, and the percentage of Check & Connect sessions participated predict mentors' social validity ratings in Year 2? A hierarchical multiple regression analysis was performed. The outcome variable was the total scores of the Mentor SIRF. The predictors were Externalizing and Internalizing Problems Composite T scores of the TRS, students' diagnostic status (i.e., label vs. no label), mentors' years of teaching experience, and the percentage of Check & Connect sessions participated. Students' diagnostic status was dummy coded. The predictor variables were examined in three steps: Step 1- the percentage of Check & Connect sessions participated, Step 2- Externalizing and Internalizing Problems Composite T scores of the TRS and students' diagnostic status (i.e., 0 = no label as a reference group; 1= with label), and Step 3- years of teaching experience. The rationale for such a hierarchy was first to control for the variability of dosage that students received and then to examine the impact of student characteristics followed by teacher characteristics. The R^2 and the increment in the R^2 at each new step were examined.

Research Question 4. Do mentors' social validity ratings in Year 2 predict their treatment fidelity in Year 2? A simple regression was conducted to examine the relation

between mentors' total scores on the SIRF and the grand mean of percentage of monthly

treatment integrity. The R^2 was also examined.

Chapter IV

Results

Attrition Analyses

Chi-square and independent *t*-tests were used to examine the differences between students who stayed for the course of two-years of intervention (stayers) and students who dropped out prior to the end of Year 2 (dropouts) in terms of demographic variables (i.e., gender, ethnicity, and special education diagnostic status), dosage, SIRF scores, and behavior severity. Results of Chi-square tests showed no differences between stayers and dropouts in ethnicity; however, stayers and dropouts significantly differed in gender ($\chi^2(1, N = 245) = 8.77, p < .01$) and special education diagnostic status ($\chi^2(1, N = 243) = 5.83, p < .05$). Specifically, based on the odds ratio, the odds for female dropouts were 2.39 times higher than male dropouts, and the odds for students without special education diagnoses were 2.04 times higher than those with diagnoses. Results of independent *t*-tests indicated that stayers and dropouts differed significantly in externalizing behavior ($t(65) = 2.30, p < .05$, two-tailed) and internalizing behavior ($t(73) = 3.00, p < .01$, two-tailed). In particular, on average, stayers displayed significantly fewer externalizing ($M = 62.70$) and internalizing ($M = 56.68$) behavior symptoms than dropouts did ($M = 71.81$, for externalizing behavior; $M = 63.25$, for internalizing behavior).

Descriptive Statistics

Table 1 displays pooled, descriptive statistics for student and mentor variables. For behavior characteristics, overall, students' Year 2 mean T score was 62.16 (range = 41-104, prior and after imputation) on the Externalizing Problem Composite (EPC) of the BASC-2 and 56.07 (range = 40-94, prior and after imputation) on the Internalizing Problem Composite (IPC). When examining behavior severity based on special education status and change in mentor, students

77

with special education status exhibited significantly more externalizing problem behavior ($M =$ 64.64) than those without special education status ($M = 59.36$), $t(13533) = -19.74$, $p < .001$, $r = .17$. Likewise, students with special education status exhibited significantly more internalizing problem behavior ($M = 58.85$) than those without special education status ($M = 52.91$), $t(13558) = -32.38$, $p < .001$, $r = .27$. Students who stayed with the same mentors ($M = 63.98$) showed significantly more externalizing problem behavior than those who had change in mentors ($M = 61.00$), $t(11176) = 10.70$, $p < .001$, $r = .10$. Students who stayed with the same mentors ($M = 57.00$) also showed significantly more internalizing problem behavior than those who had a change in mentors ($M = 55.47$), $t(10317) = 7.60$, $p < .001$, $r = .07$.

For dosage, overall, students met with mentors on average 70.65% (range $= 0$-100%, prior and after imputation) of Check & Connect sessions in Year 1, 65.27% (range $= 0$-100%, prior and after imputation) in Year 2, and 67.96% (range $= 0$-100%, prior and after imputation) over the course of two years. In general, students with special education status had significantly higher dosage ($M = 72.11$%, for Year 1; $M = 65.72$% for Year 2; $M = 68.92$%, for two years) than those without special education status ($M = 69.00$%, for Year 1; $M = 64.77$% for Year 2; $M = 66.89$%, for two years), $t(13068) = -6.43$, $p < .001$, $r = .06$ for Year 1, $t(13316) = -2.36$, $p < .05$, $r = .02$ for Year 2, $t(13126) = -5.87$, $p < .001$, $r = .05$ for two years. Similarly, students who stayed with the same mentors ($M = 77.29$%, for Year 1; $M = 70.07$% for Year 2; $M = 73.68$%, for two years) had significantly higher dosage than their counterparts ($M = 66.41$%, for Year 1; $M = 62.21$% for Year 2; $M = 64.31$%, for two years), $t(13003) = 23.72$, $p < .001$, $r = .20$ for Year 1, $t(12464) = 19.92$, $p < .001$, $r = .18$ for Year 2, $t(12574) = 28.32$, $p < .001$, $r = .24$ for two years.

For Student SIRF, overall, students rated Check & Connect above the total mid-point score of the scale (i.e., 4 as the lowest, 24 as the midmost, and 42 as the highest) over two years ($M = 33.94$, range = 4-42 in Year 1, prior and after imputation; $M = 34.86$, range =10-42 in Year 2, prior and after imputation). Furthermore, students with special education status had significantly lower SIRF scores ($M = 33.60$, for Year 1; $M = 34.51$ for Year 2) than those without special education status ($M = 34.32$, for Year 1; $M = 35.26$, for Year 2), $t(13427) = 6.37$, $p <$ .001, $r = .05$ for Year 1, $t(13354) = 7.9$, $p < .001$, $r = .07$ for Year 2. Students who stayed with the same mentors had significant higher SIRF scores ($M = 34.11$, for Year 1; $M = 35.19$, for Year 2) than those who experienced change in mentor ($M = 33.82$, for Year 1; $M = 34.65$, for Year 2), $t(12259) = 2.61$, $p < .001$, $r = .02$ for Year 1, $t(12225) = 5.73$, $p < .001$, $r = .05$ for Year 2.

In terms of mentors' variables, similar to students' SIRF, mentors also rated Check & Connect in Year 2 above the total mid-point score of the scale (i.e., 21 as the lowest, 84 as middlemost to 142 as the highest), indicating that mentors perceived the intervention as acceptable ($M = 112.12$, range = 57-143, prior and after imputation). Overall, mentors also implemented Check & Connect with high integrity ($M = 80.13\%$, range = 0-100%, prior and after imputation). Further comparisons showed that mentors who had students with special education status scored significantly higher in their SIRF ($M = 113.73$) and implemented Check & Connect with significantly higher integrity ($M = 82.01\%$) than those who had students without special education status ($M = 110.31$, for SIRF; $M = 77.99\%$, for integrity), $t(12786) = -16.78$, $p < .001$, $r = .15$ for SIRF, $t(13009) = -9.1$, $p < .001$, $r = .08$ for integrity. Mentors who had students that stayed with them rated significantly lower in their SIRF ($M = 110.92$) and implemented Check & Connect with significantly lower integrity ($M = 78.64\%$) than those who had students that experienced change in mentors ($M = 112.90$, for SIRF; $M = 81.07\%$, for

79

integrity), $t(12091) = -9.7$, $p < .001$, $r = .09$ for SIRF, $t(10841) = -5.33$, $p < .001$, $r = .05$ for integrity.

On average, mentors had 5 years of teaching experience (range = 0-26, prior and after imputation). In addition, mentors who had students with special education status had significantly more years of teaching ($M = 5.53$) than those who had students without special education status ($M = 4.92$), $t(13519) = -6.36$, $p < .001$, $r = .05$. Likewise, mentors who had students that stayed with them had significantly more years of teaching ($M = 7.81$) than their counterparts ($M = 4.01$), $t(10594) = 32.76$, $p < .001$, $r = .30$

Research Question 1. **Do students' social validity ratings in Year 1, the percentage of Check & Connect sessions participated, and change in mentors predict students' social validity ratings in Year 2 (Year 2 SIRF)?**

Simultaneous multiple regression with multiple imputation was conducted to examine whether students' social validity ratings in Year 1 (Year 1 SIRF), the percentage of Check & Connect sessions participated (two-year dosage), and change in mentors would predict students' social validity ratings in Year 2 (Year 2 SIRF). Pooled regression results (Table 2) indicated that Year 1 SIRF significantly correlated with Year 2 SIRF. As hypothesized, students' Year 1 SIRF, two-year dosage, and change in mentors significantly explained 15% of the variance in students' Year 2 SIRF, $F(3, 343.64) = 8.00$, $p < .001$. However, further examining each variable, only Year 1 SIRF made a significant contribution to the model, $t(59.23) = 4.25$, $p < .001$.

The b weights of each predictor are presented in Table 2. When holding the other predictors constant, as Year 1 SIRF increased by one unit, Year 2 SIRF increased by 0.29 units. Likewise, as two-year dosage increased by one unit, Year 2 SIRF increased by 0.03 units after controlling for the other two variables. Holding Year 1 SIRF and two-year dosage constant,

students who experienced change in mentors had lower Year 2 SIRF by 0.14 units than those who had the same mentors.

Research Question 2. Does students' behavior severity, measured by the teacher-rated BASC-2 (i.e., TRS), diagnostic status, mentors' years of teaching experience and the percentage of Check & Connect sessions participated predict students' social validity?

Hierarchical multiple regression with multiple imputation was applied to investigate how well student characteristics (i.e., behavior severity and special education diagnostic status) and mentor characteristics (i.e., years of teaching experience) predicted students' Year 2 SIRF after controlling for varied Check & Connect sessions in Year 2 (Year 2 Dosage). As shown in Table 3, none of the predictors significantly correlated with students' Year 2 SIRF. When first entered to the regression model, Year 2 Dosage did not significantly predict students' Year 2 SIRF and only explained 2% amount of the variance, $F(1,103.17) = 1.94$, $p = .17$. Contrary to the hypothesis, adding students' behavior severity and special education diagnostic status did not significantly explain the variance in students' Year 2 SIRF, $R^2 = .03$, $F(3, 566.15) = .60$, $p = .61$, and the change in variance was small (1%). Likewise, further adding mentors' years of teaching experience did not significantly increase the amount of variance in students' Year 2 SIRF, $R^2 = .04$, $F(1, 89.32) = .01$, $p = .93$, and the change in the variance remained small (1%). Table 4 shows that the b weights of each variable stayed the same with the inclusion of new variable(s) in each regression step.

Research Question 3. Does students' behavior severity, measured by the teacher-rated BASC-2 (i.e., TRS), students' diagnostic status, mentors' years of teaching experience, and the percentage of Check & Connect sessions participated predict mentors' social validity ratings in Year 2?

Hierarchical multiple regression with multiple imputation was conducted to examine how well the same predictors in Research Question 2 predicted mentors' Year 2 SIRF after controlling for Check & Connect dosage in Year 2. Similar to the findings in Research Question 2, the predictors did not significantly correlate with mentors' Year 2 SIRF (Table 5). After Year 2 dosage was entered, the regression model was nonsignificant and explained less than 1% of the variance in mentors' SIRF, $F(1, 112.57) = .08$, $p = .78$. Although adding students' behavior severity and special education diagnostic status improved the variance in mentors' SIRF to 4%, the model remained non-significant, $F(3, 385.7) = 1.65$, $p = .18$. Similarly, adding mentors' years of teaching experience did not significantly improve the prediction and only explained 5% of the variance in mentors' SIRF, $F(1, 120.18) = .06$, $p = .81$. The b weights of predictors in each step are presented in Table 6.

Research Question 4. Do mentors' social validity ratings in Year 2 predict their treatment fidelity in Year 2?

To examine the relation between mentors' SIRF and treatment integrity in Year 2, simple regression with multiple imputation was employed. As hypothesized, mentors' SIRF significantly predicted their treatment integrity, $F(1, 164.82) = 19.76$, $p < .001$. However, mentors' SIRF only explained 10% of the variance in their treatment integrity. When mentors' SIRF increased by one unit, mentors' treatment integrity increased by 0.67 units (Table 7).

Chapter V

Discussion

The study addressed gaps in the social validity literature as it pertained to Check & Connect, a mentoring-based intervention. In particular, the study examined: (a) whether students' social validity of Check & Connect was consistent over two years; (b) whether the variables investigated in analogue studies on social validity (i.e., behavior severity, special education status, and teachers' years of experience) held true for predicting students' and mentors' social validity ratings of Check & Connect; (c) whether dosage of Check & Connect and change in mentor affected students' and mentors' social validity; and (d) whether the relation between social validity and treatment integrity postulated in several theoretical models could be validated with real data.

Overall, students and mentors perceived Check & Connect as an acceptable intervention. Further analyses using multiple regression with multiple imputation showed that students' Year 1 SIRF, dosage received over 2 years, and change in mentor as a whole significantly predicted their Year 2 SIRF despite the small magnitude. However, when examining the relation between students' Year 2 SIRF and each predictor separately, only students' Year 1 SIRF significantly predicted their Year 2 SIRF. Moreover, after controlling for Year 2 dosage, students' behavior severity, special education status, and mentors' years of experience failed to predict students' Year 2 SIRF. For mentors, after holding Year 2 dosage constant, student characteristics and mentors' years of experience also failed to predict mentors' Year 2 SIRF. Nevertheless, as the theoretical models propose, mentors' Year 2 SIRF significantly predicted their Year 2 treatment integrity although the effect size was small. Taken together, the findings indicate that students' acceptability of Check & Connect in the previous year was more relevant to their acceptability in

the following year, and mentors' treatment integrity was more critical to their social validity than other influential variables of social validity identified in analogue studies.

Relations among Social Validity of Check & Connect, Dosage, and Change in Mentor

With respect to assessing the sustainability of social validity posited by Kern and Manz (2004), this study showed that students' positive social validity ratings of Check & Connect maintained over the 2-year intervention, and their social validity ratings in the previous year predicted their ratings in the second year. The findings suggest that high school students with problem behavior deemed Check & Connect as an acceptable mentoring intervention, and their prior perception determined if they continued liking the intervention. Therefore, it seems that helping students perceive positive relationships with mentors at the onset of the intervention would be vital for sustaining students' positive perception over time.

Contrary to the hypothesis, dosage and change in mentor did not predict students' acceptability in Year 2. Lack of predictive relations among dosage, change in mentor, and students' acceptability in Year 2 may be due to the way that dosage and change in mentor were measured. In this study, 2-year dosage was measured by averaging percentage of sessions attended across 2 years, which may mask the actual frequency of meeting. For example, although two students could have had an average of 50% dosage, student A may have 50% of dosage (18 meetings in 36 weeks) in Year 1 and 50% of dosage (18 meetings in 36 weeks) in Year 2, while Student B may have 75% of dosage (27 meetings in 36 weeks) in Year 1 and 25% of dosage (9 meetings in 36 weeks) in Year 2. The lack of a predictive relation in dosage may be also due to the lack of variation in dosage. That is, despite the range was large (i.e., from 0-100%), the majority of dosage scores clustered around 70-90% (see the histogram in Figure 8). Another explanation may be that the use of percentage may not capture the true dosage.

Although mentors were required to meet with students once per week at least for 10 min, the length of each meeting was not recorded. Measuring the time spent in the meeting may better reflect the dosage that students received. Similarly, some students experienced change in mentors more than twice, but this variable was quantified as a categorical measure (one, more than one). Alternatively, the absence of predictive relations of dosage and change in mentor on students' acceptability in Year 2 may suggest that frequency of student meetings with mentors or the switch of mentors did not matter as long as they perceived there was someone who cared about them.

Student and Mentor Characteristics and Social Validity Ratings

Student and teacher characteristics have been the focus of research interest in the social validity literature. This study also investigated these two variables but extended the literature from analogue to authentic and direct consumers' perspectives. The results revealed that students' behavior severity and special education status, and mentors' years of experience had no predictive effects on their Year 2 acceptability, suggesting that students viewed Check & Connect as acceptable regardless of their behavior severity, special education status, or their mentors' years of experience. Likewise, both novice and seasoned mentors perceived Check & Connect as acceptable regardless of students' behavior severity and their special education status.

The results of this study were consistent with the analogue studies in two aspects. First, the target student's behavior severity alone did not show statistical significant effect on student raters' acceptability (e.g., Elliott et al., 1986). Second, when examining student and teacher characteristics individually, the target student's behavior severity (e.g., Kutsick et al., 1991), special education label (e.g., Epstein et al., 1986; Stinnett et al., 2001), and teacher raters' years

85

of teaching experience (e.g., Power et al., 1995) was not significantly predictive of teacher raters' acceptability.

Interestingly, the results of this study differed from those in the Cowan and Sherian study (2003) despite that both studies assessed authentic consumers' acceptability utilized a hierarchical regression analysis. In this study, students' behavior severity did not significantly predict students' and mentors' acceptability whereas Cowan and Sherian found that students' behavior severity was significant, explaining 16% of the variance in the teachers' acceptability. The difference may result from the instruments used to measure behavior severity. In this study, the BASC-2, a norm-referenced instrument was used, while in Cowan and Sherian's study, behavior severity was determined solely by a questionnaire with one item on a 7-point Likert scale, which the researchers acknowledged has been notoriously unreliable. Another alternative explanation for the different findings in behavior severity may be due to the differences in interventions and the implementers. This study focused on a mentoring-based intervention, and mentors were volunteers who may have understood and accepted the behavior severity of their mentees. By contrast, Cowan and Sherian's study targeted a behavior-based intervention (i.e., conjoint behavioral consultation), in which the degree of supports or the intervention components are based on behavior severity. Moreover, the implementers were participating students' teachers, who had to manage problem behavior directly.

Similar to dosage discussed, the lack of predictive relations in student behavior severity and mentors' years of experience may be due to the limited variation in the variables. Despite a wide range in these variables (i.e., range = 41-104 for the Externalizing Problem Composite; range = 40-94 for the Internalizing Problem Composite; range = 0-26 for years of teaching experience), the majority of the scores clustered in 50-60 and 0-5 for behavior severity and years

of teaching, respectively (see histograms in Figure 8). Alternatively, the absence of predictive

relations in student behavior problem, special education status, and mentors' years of experience

may suggest these variables were not influential factors for social validity as some analogue

studies suggested (e.g., Elliott et al., 1986; Epstein et al., 1986; Kutsick et al., 1991; Power et al.,

1995; Stinnett et al., 2001). Given Check & Connect is a relationship-based intervention, factors

examined in the mentoring literature may be more relevant to students' and mentors' social

validity, such as perceived quality of mentoring relationships (e.g., Karcher, Nakkula, & Harris,

2005), students' prior relationships with others (e.g., Schwartz, Rhodes, Chan, & Herrera, 2011),

or mentors' motivations to become mentors (e.g., Caldarella, Gomm, Shatzer, & Wall, 2010).

Relation between Social Validity and Treatment Integrity

Similar to the survey by Miramontes et al. (2011), which found a significantly positive

relation between teachers' and school personnel's acceptability of Tier 1 SWPBS and their

implementation, this study also revealed a similar predictive relation in Check & Connect.

Despite no statistical significance, the results from the studies by Sterling-Turner and Watson

(2002) and Mautone et al. (2009) also showed a positive correlation between social validity and

treatment integrity. The results compiled from this small body of literature seem to suggest that

a positive correlation exists between social validity and treatment integrity as depicted in the

Reimers et al. (1987) model, depicted in Figure 2 (i.e., High acceptability → High compliance)

and the self-reinforcing model, depicted in Figure 3 (i.e., Social Acceptability by Treatment

Agent (B) → High Treatment Integrity (C)).

Further comparing the two models, Witt and Elliot's (1985) model did not directly

specify a reciprocal relation between acceptability of treatment and integrity of treatment.

However, the statistically significant correlation between acceptability and treatment integrity

87

found in this study suggests a fifth reciprocal relation can be added into Witt and Elliot's model. That is, a reciprocal line can be placed inside the model, indicating a reciprocal relation between *acceptability of treatment* and *integrity of treatment* in addition to the four reciprocal relations among the components (see Figure 1).

Implications for Practice

From the findings, several implications can be drawn for practice. First, based on the results of both student and mentor acceptability data, Check & Connect has good social validity as a mentoring-based for high school students with problem behavior. When selecting interventions, practitioners may consider Check & Connect as a possible option.

Second, since students' previous acceptability of Check & Connect affects their later acceptability, it may be important to help students buy into the intervention early on for their own benefit and the survival of the intervention. Possible strategies may include: (a) presenting Check & Connect to high school students with problem behavior in a non-stigmatizing way, (b) clarifying students' misconceptions of mentoring process, (c) matching students with mentors who have certain interaction styles or personalities that students prefer, (d) pairing students with adults with whom they already have established trusting relationships, or (e) training mentors to interact with students in their preferred ways. Furthermore, given the predictive relation of students' acceptability between the two time points, practitioners or researchers may consider utilizing students' acceptability data to detect any dissatisfaction that may lead to dropping out of the program and provide supports for students and mentors.

Third, the absence of predictive effects of dosage and change in mentors on students' acceptability suggests that how frequent students met with mentors and who their mentors were may not matter. Hence, practitioners may consider quality over quantity. That is, ensuring

students have a good quality mentoring experience may be the most important aspect of the intervention. This may be established through training mentors to have good reflective listening skills. Further, good quality mentoring processes that are tailored to students' needs may be more critical for female students and those with severe behavior problem in that these two types of students were prone to drop out of the intervention in the first year despite their positive perception of Check & Connect. It is speculated that the nature of Check & Connect is an instrumental oriented approach (i.e., focus on coaching and problem-solving). Female students may prefer a psychosocial approach (i.e., focus on emotional connectedness) while students with severe behavior problems may require more attunement from mentors. Therefore, providing additional psychosocial or individualized support may help female students and students with severe behavior problem sustain in the mentoring relationships. In addition to mentor training, practitioners may consider student preferences in the mentoring process, such as pairing female students with female mentors (Spencer & Liang, 2009) or matching students with severe problem behavior with mentors who have a strong background in a helping profession (DuBois, Holloway, Valentine, & Cooper, 2002).

Fourth, change in mentors and mentors' years of experience had no impact on students' and mentors' acceptability, which implies that the prescribed two years of commitment for mentors may not be necessary for social validity, and mentors' years of experience may not be an important criterion for mentoring. In fact, given instructional time constraints, the absence of a 2-year commitment may make becoming a mentor more appealing and doable for teachers. Moreover, when recruiting mentors, practitioners or researchers can target teachers with various years of experience, thereby increasing the potential number of available mentors.

Fifth, because of the predictive relation between social validity and treatment integrity, practitioners or researchers may consider using social validity data to gauge when to provide booster sessions for mentors before observing deteriorating implementation.

Strengths

The study contributes to the literature in several ways. First, the study addressed the analogue nature of social validity research by analyzing social validity of real intervention recipients. Data obtained from authentic consumers allow researchers to determine whether variables examined in the analogue studies remain applicable in real settings.

Second, the study addressed the validity limitations of SWPBS discussed by Kern and Manz (2004). Particularly, the study focused on the social validity of Check & Connect and extended the literature by examining students' perspectives over two years. Further, the study explored two possible variables of social validity: dosage and change in mentor, which were unique to Check & Connect and other mentoring-based interventions.

Third, from statistical analysis standpoint, the study employed multiple imputation to address potential bias from missing data. Along this line, the study validated the relation between social validity and treatment integrity with parametric analysis in contrast to the study by Miramontes et al. (2011), which applied a non-parametric approach to analyze the relation between social validity and treatment integrity.

Limitations

Inevitably, several limitations can be found in the study; therefore, the results should be interpreted with caution. The foremost is the way that the data were aggregated, particularly Student SIRF, dosage, and change in mentor. Students' SIRF scores were collapsed across waves to represent two yearly time points. That is, Waves 1 to 3 were collapsed to Year 1, and

Waves 4 and 5 were collapsed as Year 2. If a student's data from one wave were missing, then the next closest one or two were averaged. For example, Student A had data at Waves 2 to 5; therefore, the average score of Wave 2 and 3 would be Year 1 score, and the average of Wave 4 and 5 as Year 2 score. If Student B only had data at Waves 3 and 4, the data at the two waves would be used as Year 1 and 2 scores. As a result, each participant's yearly score did not reflect the same time point. Thus, it is unclear if the significant prediction between Year 1 and Year 2 SIRF would manifest across more frequent assessments. In addition, dosage was calculated based on whether or not weekly meetings were held. Although a weekly meeting was required, students were free to seek advice from their mentors more than once a week; however, the actual frequency and the length of the meeting were not documented. Consequently, the dosage analyzed in this study may not reflect the true dosage that students received. As discussed earlier, using percentage of sessions in which students participated to represent the dosage of Check & Connect may mask the true dosage across time, which may have varied. Further, change in mentor was quantified as a categorical variable (i.e., change or no change). In fact, some students experienced more than two mentors. It is unclear whether measuring change in mentor as the frequency of change would result in differences.

Second, the psychometric properties of Student SIRF remain unestablished despite acceptable internal consistency. It is unclear if Student SIRF shares similar factor structure as Mentor SIRF. Hence, caution is warranted in interpreting the significant prediction between Year 1 and Year 2 SIRF.

Third, the study only assessed two direct consumers' acceptability – students and mentors. It is unknown if indirect consumers, such as parents and school personnel, would also perceive Check & Connect as positively as students and mentors.

91

Fourth, the study only examined the quantitative aspect of the social validity of Check & Connect, measured by the total scores of Student and Mentor SIRF. The relation between quantitative SIRF data and qualitative feedback is unknown. That is, it is unclear whether qualitative data might provide a different view from students and mentors about Check & Connect acceptability. Moreover, since SIRF is a multi-dimensional measure, without teasing apart the factors, it is unclear which dimension is critical. That is, for the Mentor SIRF, it is unclear which dimension among *understanding*, *effectiveness, cost, time, willingness*, and *side effects* was the most important factor that predicted their treatment integrity. Likewise, for the Student SIRF, it is unclear which factors among *understanding, like, helpfulness, and side effects* were strong predictors between Year 1 and 2.

Fifth, the study only examined the relation between post-treatment social validity and treatment integrity. However, the Reimer, Wacker, and Kieppl's (1987) model (Figure 2) and the self-reinforcing model (Figure 3) depict the relation between pre-treatment social validity and treatment integrity. Therefore, it is unknown if mentors' pre-treatment SIRF would result in significant prediction in their treatment integrity.

Future Research

Based on the findings and the limitations of this study, four directions are suggested for future research pertaining to: (a) replication, (b) assessment, (c) methodology, and (d) utility. First, although this study showed that students' and mentors' previous social validity ratings were more relevant to integrity than student characteristics (i.e., behavior severity and special education status) or mentors' years of experience, replications with different datasets and mentoring-based interventions (e.g., Check In/Check Out) are needed to validate the relations

among these variables. Replications are also necessary to fully understand the relation between social validity and treatment integrity.

In the area of assessment, as Schwartz and Baer (1991) suggested, social validity should be included in a program assessment and assessed frequently across relevant consumers. Future research should continue assessing the social validity of Check & Connect frequently and at the same time point. Researchers may also consider assessing pre-treatment and post-treatment social validity among students, mentors, school personnel, and parents to capture a wide range of perspectives and to ascertain whether there is change over time. In addition to using self-report measures, researchers may consider incorporating objective measures, such as documenting the frequency and party that indicated unavailability for meeting and the reasons. Despite being a secondary measure, social validity assessment also requires psychometric rigor (Schwartz & Baer, 1991). Hence, establishing the psychometrics of the Student SIRF and the adult SIRF for indirect consumers such as parents and other school personnel is needed. In addition, under the SWPBS framework, Check & Connect is suggested as a Tier 2 intervention (Christenson, Stout, & Pohl, 2012). However, research within the context of tiered support remains scant. Whether the findings from this study would be replicated or would produce different results under the SWPBS framework deserves further investigations.

Regarding methodology, several future directions are recommended to illuminate the influencing variables on social validity and the social validity models. First, non-significant prediction in dosage may be due to the imprecision of measurement; therefore, future research may consider measuring dosage by the actual time that mentors spent with students. This may be accomplished through documenting the length of meeting time on the existing Check & Connect monitoring sheet or using a meeting log recorded by both students and mentors.

Second, future research may consider exploring other possible influencing variables of social validity. For example, matching mentors and students is commonly used in mentoring-based interventions. Future research may explore whether matching based on students' preferred interaction styles or exhibiting trusting relationships with other adults (i.e., students nominate adults they trust) would predict students' and mentors' social validity. As discussed earlier, future research may consider investigating the perceived quality of mentoring relationships, students' prior relationships with others, and mentors' motivations for volunteering on students' and mentors' social validity. If the predictive relations hold true in the aforementioned variables, researchers may further investigate if providing on-going training for mentors to improve the relationships (e.g., interact with students in their preferred ways) and to adjust their expectations for mentoring relationships would result in better mentors' and students' social validity. Likewise, researchers may examine whether providing support for students, particularly those who have a history of difficulties in maintaining relationships with others, would lead to better mentors' and students' social validity. In addition, this study only examined students' social validity over 2 years. It is unknown if the social validity of mentors who stayed longer than 2 years would differ. Thus, researchers may consider exploring mentors' social validity over time and the variables that encourage them to continue for a longer time period. Furthermore, this study did not examine any interaction effect among variables as in Cowan and Sheiran's study (2003). Researchers may consider exploring if the interaction between behavior severity and dosage would affect students' and mentors' acceptability.

In terms of social validity models, this study only tested the relation between social validity and treatment integrity. Based on Witt and Elliot's model, future research may test if effectiveness of treatment mediates the relation between social validity and treatment integrity.

94

Additionally, researchers may consider testing the Reimers et al. (1987) model, particularly the relations among understanding, acceptability, compliance, and effectiveness using structural equation modeling (SEM).

Future research may also consider employing a mixed methods design and include qualitative data. Focus group interviews could be conducted to gain insight from students and mentors, particularly the perspectives from females and students with severe problem behavior since they had a higher dropout rate than their counterparts. To obtain better estimation, researchers should explore and include auxiliary variables when using imputation or conduct imputation at the item level prior to aggregating the total scores.

Finally, for utility, the current study only examined the global relation between mentors' social validity and their treatment integrity of Check & Connect at one time point. Future research may consider exploring the relation between social validity and treatment integrity over time. Possible relations between social validity and treatment integrity can be conceptualized as a cross-lagged model illustrated in Figure 9. Future research may consider: (a) whether mentors' social validity and treatment integrity at the previous time point would predict their scores at later time point; (b) whether mentors' previous social validity (e.g., Time 1) would predict their later treatment integrity (e.g., Time 2); and (c) whether mentors' previous treatment integrity would predict their later social validity. If the predictive relation holds true, researchers may further investigate whether providing booster sessions to mentors or making adaptations would sustain later social validity and treatment integrity when compared to those who did not receive booster sessions or adaptations. In other words, data could be used to ascertain how the findings from social validity assessments can be used to gauge later treatment integrity or vice versa. Although Strain et al. (2009) illustrated the use of social validity to select goals, refine

intervention procedures, and promote treatment integrity in five cases, empirical studies with

rigorous designs and statistical analysis will be needed to elucidate the utility of social validity.

Conclusion

Check & Connect shows promise as an intervention for disengaged high school students

with problem behavior. Nevertheless, little is known about its social validity, which reflects the

dearth of authentic consumer perspectives in the social validity literature, particularly students'

perspectives over a long period of time. This study provides a glimpse of students' and mentors'

authentic perspectives toward Check & Connect through a longitudinal dataset. In general,

Check & Connect is a mentoring-based intervention that appears to have good social validity,

evidenced by positive ratings from students and mentors and the sustainability of students'

positive ratings over 2 years of intervention. This study also revealed interesting findings that

frequency of student meetings with mentors and change in mentors did not affect their

acceptability; students and mentors also perceived Check & Connect positively regardless of

their characteristics (i.e., behavior severity and special education status, and years of teaching

experience for students and teachers, respectively). As the conceptual models delineate, mentors

who rated Check & Connect as acceptable were more likely to implement it with integrity.

Building upon the findings, this study offers directions for future research, including the

need for replicating the results with Check & Connect or other mentoring-based interventions.

To accomplish this goal, it is necessary to establish psychometric rigor of social validity

assessment that allows reliable measurement at multiple times across different consumers.

In addition to frequent assessment, it is recommended to consider a mixed methods design to

gain insight from different consumers in different ways. Other considerations include exploring

additional potential variables that influence consumers' acceptability of Check & Connect and

96

testing the theoretical models with advanced statistical analysis and examining possible mediating effects.

Following Cowan and Sherian (2003), this study also bridges some gaps in analogue and applied studies. To further advance knowledge of social validity and its utility in the intervention research, extending the current study and the endeavor of assessing and reporting social validity will be needed, or, as Finny (1991) advised, "Keep asking them" (p.248).

References

Adkins, V. K. (1997).   Social validity and naturalistic ethics: Wolf and Quine. *Behavior and Social Issues, 7*, 153-157.

Ahrens, K., DuBois, D. L., Lozano, P., & Richardson, L. P. (2010). Naturally acquired mentoring relationships and young adult outcomes among adolescents with learning disabilities. *Learning Disabilities Research & Practice, 25*(4), 207-216. doi: 10.1111/j.1540-5826.2010.00318.x

Allinder, R. M., & Oats, R. G. (1997). Effects of acceptability on teachers' implementation of curriculum-based measurement and student achievement in mathematics computation. *Remedial and Special Education, 18*(2), 113-120. doi: 10.1177/074193259701800205

Allison, P. D. (2001). *Missing data* (Vol. 136). Sage publications.

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools, 45*(5), 369-386. doi: 10.1002/pits.20303

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*(1), 91-97. doi: 10.1901/jaba.1968.1-91

Beam, M. R., Chen, C., & Greenberger, E. (2002). The nature of adolescents' relationships with their "very important" nonparental adults. *American Journal of Community Psychology, 30*(2), 305-325. doi: 10.1023/A:1014641213440

Bowman, L. J. (2005). Grade retention: Is it a help or hindrance to student academic success? *Preventing School Failure: Alternative Education for Children and Youth, 49*(3), 42-46. doi: 10.3200/PSFL.49.3.42-46

Bowerman, B. L., & O'Connell, R. T. (1990). *Linear statistical models: An applied approach* (2nd ed.). Belmont, CA: Duxbury.

Brener, N. D., Kann, L., McManus, T., Kinchen, S. A., Sundberg, E. C., & Ross, J. G. (2002). Reliability of the 1999 youth risk behavior survey questionnaire. *Journal of Adolescent Health, 31*(4), 336-342. doi: http://dx.doi.org/10.1016/S1054-139X(02)00339-7

Bridgeland, J. M., DiIulio Jr, J. J., & Morison, K. B. (2006). The silent epidemic: Perspectives of high school dropouts. *Civic Enterprises.*

Britner, P. A., & Kraimer-Rickaby, L. (2009). Mentoring and its role in promoting academic and social competency. In Gullotta, T. P., Bloom, M., Gullotta, C. F., & Messian, J. C., (Eds), *A blueprint for promoting academic and social competence in after-school programs* (pp. 79-100). New York: Springer.

Bumbarger, B. K. (2014). Understanding and promoting treatment integrity in prevention. In Hagermoser Sanetti, L. M., & Kratochwill, T. R. (Eds.), *Treatment integrity: A foundation for evidence-based practice in applied psychology* (pp. 35-54). Washington, DC, US: American Psychological Association.

Chapman, C., Laird, J., Ifill, N., & KewalRamani, A. (2011). Trends in High School Dropout and Completion Rates in the United States: 1972–2009 (NCES 2012-006). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch.

Carr, J. E., Austin, J. L., Britton, L. N., Kellum, K. K., & Bailey, J. S. (1999). An assessment of social validity trends in applied behavior analysis. *Behavioral Interventions, 14*(4), 223-231. doi: 10.1002/(SICI)1099-078X(199910/12)14:4<223::AID-BIN37>3.0.CO;2-Y

Caldarella, P., Gomm, R. J., Shatzer, R. H., & Wall, D. G. (2010). School-Based Mentoring: A study of volunteer motivations and benefits. *International Electronic Journal of Elementary Education, 2*(2), 199-216.

Center for Adolescent Research in Schools (2011). Classroom-based Interventions Manual.

Centers for Disease Control and Prevention (2004). Youth Risk Behavior Surveillance – United States, 2003. Morbidity and Mortality Weekly Report, 53(SS2), 1–100.

Cheney, D., Lynass, L., Flower, A., Waugh, M., Iwaszuk, W., Mielenz, C., & Hawken, L. (2010). The check, connect, and expect program: A targeted, tier 2 intervention in the schoolwide positive behavior support model. *Preventing School Failure, 54*(3), 152-158.

Christenson, S., Stout, K., & Pohl, A. (2012). Check & Connect: A comprehensive student engagement intervention: Implementing with fidelity. *Minneapolis, MN: University of Minnesota, Institute on Community Integration.*

Clark, L., & Elliott, S. N. (1988). The influence of treatment strength information on knowledgeable teachers' pretreatment evaluations of social skills training methods. *Professional School Psychology, 3*(4), 241-251. doi: 10.1037/h0090567

Clarke, S., & Dunlap, G. (2008). A descriptive analysis of intervention research published in the journal of positive behavior interventions: 1999 through 2005. *Journal of Positive Behavior Interventions, 10*(1), 67-71. doi: 10.1177/1098300707311810

Cowan, R. J., & Sheridan, S. M. (2003). Investigating the acceptability of behavioral interventions in applied conjoint behavioral consultation: Moving from analog conditions to naturalistic settings. *School Psychology Quarterly, 18*(1), 1-21.

Dong, Y., & Peng, C. (2013). Principled missing data methods for researchers. Springerplus 2: 222. Published online May 14, 2013. doi: 10.1186/ 2193-1801-2-222.

DuBois, D. L., Doolittle, F., Yates, B. T., Silverthorn, N., & Tebes, J. K. (2006). Research

    methodology and youth mentoring. *Journal of Community Psychology, 34*(6), 657-676. doi:

    10.1002/jcop.20122

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of

    mentoring programs for youth: A meta-analytic review. *American Journal of Community*

    Psychology, 30(2), 157-197. doi: 10.1023/A:1014628810714

DuBois, D. L., & Silverthorn, N. (2005). Natural mentoring relationships and adolescent health:

    Evidence from a national study. *American Journal of Public Health, 95*(3), 518-524. doi:

    10.2105/AJPH.2003.031476

Dunn, C., Chambers, D., & Rabren, K. (2004). Variables affecting students' decisions to drop out

    of school. *Remedial and Special Education, 25*(5), 314-323. doi:

    10.1177/07419325040250050501

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. doi: 10.1007/s10464-008-

    9165-0

Eckert, T. L., & Hintze, J. M. (2000). Behavioral conceptions and applications of acceptability:

    Issues related to service delivery and research methodology. *School Psychology Quarterly,*

    15(2), 123-148. doi: 10.1037/h0088782

Ehrhardt, K. E., Barnett, D. W., Lentz Jr., F. E., Stollar, S. A., & Reifin, L. H. (1996). Innovative

    methodology in ecological consultation: Use of scripts to promote treatment acceptability

    and integrity. *School Psychology Quarterly, 11*(2), 149-168. doi: 10.1037/h0088926

Elliott, S. N., & Treuting, M. V. (1991). The behavior intervention rating scale: Development and validation of a pretreatment acceptability and effectiveness measure. *Journal of School Psychology, 29*(1), 43-51. doi: 10.1016/0022-4405(91)90014-I

Elliott, S. N., Turco, T. L., & Gresham, F. M. (1987). Consumers' and clients' pretreatment acceptability ratings of classroom group contingencies. *Journal of School Psychology, 25*(2), 145-153. doi: http://dx.doi.org/10.1016/0022-4405(87)90023-9

Elliott, S. N., Witt, J. C., Galvin, G. A., & Moe, G. L. (1986). Children's involvement in intervention selection: Acceptability of interventions for misbehaving peers. *Professional Psychology: Research and Practice, 17*(3), 235-241. doi: 10.1037/0735-7028.17.3.235

Elliott, S. N., Witt, J. C., Galvin, G. A., & Peterson, R. (1984). Acceptability of positive and reductive behavioral interventions: Factors that influence teachers' decisions. *Journal of School Psychology, 22*(4), 353-360. doi: 10.1016/0022-4405(84)90022-0

Epstein, M. H., Matson, J. L., Repp, A., & Helsel, W. J. (1986). Acceptability of treatment alternatives as a function of teacher status and student level. *School Psychology Review, 15*(1), 84-90.

Fairbanks, L. D., & Stinnett, T. A. (1997). Effects of professional group membership, intervention type and diagnostic label on treatment. *Psychology in the Schools, 34*(4), 329.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160.

Fawcett, S. B. (1991). Social validity: A note on methodology. *Journal of Applied Behavior Analysis, 24*(2), 235-239. doi: 10.1901/jaba.1991.24-235

Finn, C. A., & Sladeczek, I. E. (2001). Assessing the social validity of behavioral interventions:
 A review of treatment acceptability measures. *School Psychology Quarterly, 16*(2), 176-
 206. doi: 10.1521/scpq.16.2.176.18703

Finn, J. D. (1993). *School engagement & students at risk.* Washington, DC: National Center for
 Education Statistics.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59*(2), 117-142.
 doi: 10.3102/00346543059002117

Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure.
 Journal of Applied Psychology, 82(2), 221-234. doi: 10.1037/0021-9010.82.2.221

Finney, J. W. (1991). On futher development of the concept of social validity. *Journal of Applied
 Behavior Analysis, 24*(2), 245-249. doi: 10.1901/jaba.1991.24-245

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the
 concept, state of the evidence. *Review of Educational Research, 74*(1), 59-109. doi:
 10.3102/00346543074001059

Fuqua, R. W., & Bachman, J. (1986). Some factors limiting the applicability of applied
 behavioral research. In Poling, A., & Fuqua, R. W. (Eds.), *Research methods in applied
 behavior analysis: issues and advances* (pp. 85-98). New York: Plenum Press.

Gresham, F. M. (1996). Treatment integrity in single-subject research. In R. D. Franklin, D. B.
 Allison & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 93-117).
 Mahwah, NJ: Lawrence Erlbaum.

Gresham, F. M. (2014). Measuring and analyzing treatment integrity data in research. In
 Hagermoser Sanetti, L. M., & Kratochwill, T. R. (Eds.), *Treatment Integrity: A foundation*

for evidence-based practice in applied psychology (pp. 109-130). Washington, DC, US: American Psychological Association.

Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice, 15*(4), 198-205.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*(3), 206-213.

Grossman, J. B., & Bulle, M. J. (2006). Review of what youth programs do to increase the connectedness of youth with adults. *Journal of Adolescent Health, 39*(6), 788-799. doi: doi:10.1016/j.jadohealth.2006.08.004

Grossman, J. B., & Rhodes, J. E. (2002). The test of time: Predictors and effects of duration in youth mentoring relationships. *American Journal of Community Psychology*, *30*(2), 199-219.

Hagermoser 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, 37*(4), 195-202. doi: 10.1177/1534508411432466

Harris, K. R., Preller, D. M., & Graham, S. (1990). Acceptability of cognitive-behavioral and behavioral interventions among teachers. *Cognitive Therapy and Research, 14*(6), 573-587. doi: 10.1007/BF01173365

Hawkins, R. P. (1991). Is social validity what we are interested in? argument for a functional

  approach. *Journal of Applied Behavior Analysis, 24*(2), 205-213. doi: 10.1901/jaba.1991.24-

  205

Herrera, C. (1999). *School-based mentoring: A first look into its potential.* Public/Private

  Ventures, Philadelphia, PA.

Herrera, C., Sipe, C. L., & McClanahan, W. S. (2000). *Mentoring school-age children:*

  Relationship development in community-based and school-based programs. Public/Private

  Ventures, Philadelphia, PA.

Herrera, C., Grossman, J. B., Kauh, T. J., & McMaken, J. (2011). Mentoring in schools: An

  impact study of big brothers big sisters school-based mentoring. *Child Development, 82*(1),

  346-361. doi: 10.1111/j.1467-8624.2010.01559.x

Hoagwood, K., Horwitz, S., Stiffman, A., Weisz, J., Bean, D., Rae, D., . . . Leaf, P. (2000).

  Concordance between parent reports of children's mental health services and service

  records: The services assessment for children and adolescents (SACA). *Journal of Child*

  and Family Studies, 9(3), 315-331.

Hollar, D. (2005). Risk behaviors for varying categories of disability in NELS:88. *Journal of*

  School Health, 75(9), 350-358. doi: 10.1111/j.1746-1561.2005.00049.x

Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for school-

  wide positive behavior support. *Focus on Exceptional Children, 42*(8), 1-14.

Houten, R. V. (1979). Social validation: The evolution of standards of competency for target

  behaviors. *Journal of Applied Behavior Analysis, 12*(4), 581-591. doi:

  10.1901/jaba.1979.12-581

Institute for Digital Research and Education (IDRA), UCLA: Statistical Consulting Group (2015, July). *Statistical Computing Seminars: Missing Data in SAS Part 1*. Retrieved from http://www.ats.ucla.edu/stat/sas/seminars/missing_data/mi_new_1.htm

Jekielek, S., Moore, K. A., & Hair, E. C. (2002). *Mentoring programs and youth development: A synthesis.* Washington, DC: Child Trends.

Jensen, P. S., Hoagwood, E. K., Roper, M., Arnold, L. E., Obert, C., Crowe, M., . . . Wells, K. (2004). The services for children and Adolescents–Parent interview: Development and performance characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry, 43*(11), 1334-1344. doi: http://dx.doi.org/10.1097/01.chi.0000139557.16830.4e

Karcher, M. J., Nakkula, M. J., & Harris, J. (2005). Developmental mentoring match characteristics: Correspondence between mentors' and mentees' assessments of relationship quality. *Journal of Primary Prevention*, *26*(2), 93-110.

Karpinski, M. J., Neubert, D. A., & Graham, S. (1992). A follow-along study of postsecondary outcomes for graduates and dropouts with mild disabilities in a rural setting. *Journal of Learning Disabilities, 25*(6), 376-385. doi:10.1177/002221949202500606

Kaufman, J. S., Jaser, S. S., Vaughan, E. L., Reynolds, J. S., Di Donato, J., Bernard, S. N., & Hernandez-Brereton, M. (2010). Patterns in office referral data by grade, Race/Ethnicity, and gender. *Journal of Positive Behavior Interventions, 12*(1), 44-54.

Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification, 1*(4), 427-452. doi: 10.1177/014544557714001

Kazdin, A. E. (1981). Acceptability of child treatment techniques: The influence of treatment efficacy and adverse side effects. *Behavior Therapy, 12*(4), 493-506. doi:10.1016/S0005-7894(81)80087-1

Kennedy, C. H. (1992). Trends in the measurement of social validity. *The Behavior Analyst, 15*(2), 147-156.

Kennedy, C. H. (2005). Social validity. In Kennedy, C. H., *Single-case designs for educational research* (pp. 218-233). Boston, MA: Allyn and Bacon.

Kent, K. M., Pelham Jr., W. E., Molina, B. S. G., Sibley, M. H., Waschbusch, D. A., Yu, J., . . . Karch, K. M. (2011). The academic experience of male high school students with ADHD. *Journal of Abnormal Child Psychology, 39*(3), 451-462. doi: 10.1007/s10802-010-9472-4

Kern, L., Evans, S. W., & Lewis, T. J. (2011). Description of an iterative process for intervention development. *Education and Treatment of Children, 34*(4), 593-617.

Kern, L. & Manz, P. M. (2004). A look at current validity issues of school-wide behavior support. *Behavioral Disorders, 30,* 47-59.

King, H. A., & Bosworth, H. (2014). Treatment fidelity in health services research. In Hagermoser Sanetti, L. M., & Kratochwill, T. R. (Eds.), *Treatment integrity: A foundation for evidence-based practice in applied psychology* (pp. 15-33). Washington, DC, US: American Psychological Association.

Kortering, L. J., & Braziel, P. M. (1999). School dropout from the perspective of former students: Implications for secondary special education programs. *Remedial and Special Education, 20*(2), 78-83. doi: 10.1177/074193259902000203

Kortering, L. J., & Christenson, S. (2009). Engaging students in school and learning: The real deal for school completion. *Exceptionality, 17*(1), 5-15. doi: 10.1080/09362830802590102

Krezmien, M. P., Leone, P. E., & Achilles, G. M. (2006). Suspension, race, and disability: Analysis of statewide practices and reporting. *Journal of Emotional and Behavioral Disorders, 14*(4), 217-226. doi: 10.1177/10634266060140040501

Kutsick, K. A., Witt, J. C., & Gutkin, T. B. (1991). The impact of treatment development

process, intervention type, and problem severity on treatment acceptability as judged by

classroom teachers. *Psychology in the Schools, 28*(4), 325-331. doi: 10.1002/1520-

6807(199110)28:4<325::AID-PITS2310280407>3.0.CO;2-Y

Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity:

An essential—but often forgotten—component of school-based interventions. *Preventing*

School Failure: Alternative Education for Children and Youth, 48(3), 36-43.

Ledford, J. R., & Wolery, M. (2013). Procedural fidelity: An analysis of measurement and

reporting practices. *Journal of Early Intervention, 35*(2), 173-193. doi:

10.1177/1053815113515908

Lehr, C. A., Sinclair, M. F., & Christenson, S. L. (2004). Addressing student engagement and

truancy prevention during the elementary school years: A replication study of the check &

connect model. *Journal of Education for Students Placed at Risk, 9*(3), 279-301.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing

values. *Journal of the American Statistical Association*, *83*(404), 1198-1202.

March, J. S., Parker, J. D. A., Sullivan, K., Stallings, P., & Conners, C. K. (1997). The

multidimensional anxiety scale for children (MASC): Factor structure, reliability, and

validity. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(4), 554-

565. doi: http://dx.doi.org/10.1097/00004583-199704000-00019

Martens, B. K., & Meller, P. J. (1989). Influence of child and classroom characteristics on

acceptability of interventions. *Journal of School Psychology, 27*(3), 237-245. doi:

http://dx.doi.org/10.1016/0022-4405(89)90038-1

Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist, 56*(3), 227-238. doi: 10.1037/0003-066X.56.3.227

Masten, A. S., & Powell, J. L. (2003). A resilience framework for research, policy, and practice. In Luthar, S. (Ed.), *Resilience and vulnerability: Adaptation in the context of childhood adversities*, (pp. 1-25). New York, NY: Cambridge University Press.

Mautone, J. A., DuPaul, G. J., Jitendra, A. K., Tresco, K. E., Junod, R. V., & Volpe, R. J. (2009). The relationship between treatment integrity and acceptability of reading interventions for children with attention-Deficit/Hyperactivity disorder. *Psychology in the Schools, 46*(10), 919-931. doi: 10.1002/pits.20434

Maynard, B. R., Salas-Wright, C. P., & Vaughn, M. G. (2014). High school dropouts in emerging adulthood: Substance use, mental health problems, and crime. *Community Mental Health Journal (no pagination specified).* doi: http://dx.doi.org/10.1007/s10597-014-9760-5

Menard, S. (1995). *Applied logistic regression analysis*. Sage university paper series on quantitative application in the social sciences, 07-106. Thousand Oaks, CA: Sage.

McNamara, J. K., & Willoughby, T. (2010). A longitudinal study of risk-taking behavior in adolescents with learning disabilities. *Learning Disabilities Research & Practice, 25*(1), 11-24. doi: 10.1111/j.1540-5826.2009.00297.x

McNamara, J., Vervaeke, S., & Willoughby, T. (2008). Learning disabilities and risk-taking behavior in adolescents: A comparison of those with and without comorbid attention-Deficit/Hyperactivity disorder. *Journal of Learning Disabilities, 41*(6), 561-574.

Miltich Lyst, A., Gabriel, S., O'Shaughnessy, T. E., Meyers, J., & Meyers, B. (2005). Social validity: Perceptions of check and connect with early literacy support. *Journal of School Psychology, 43*(3), 197-218. doi: http://dx.doi.org/10.1016/j.jsp.2005.04.004

Miramontes, N. Y., Marchant, M., Heath, M. A., & Fischer, L. (2011). Social validity of a

    positive behavior interventions and support model. *Education & Treatment of Children,*

    34(4), 445-468.

Newman, L., Wagner, M., Knokey, A., Marder, C., Nagle, K., Shaver, D., & Wei, X. (2011).

    The post-high school outcomes of young adults with disabilities up to 8 years after high

    school: A report from the national longitudinal transition study-2 (NLTS2). NCSER 2011-

    3005. *National Center for Special Education Research.*

O'Dell, S. M., Vilardo, B. A., Kern, L., Kokina, A., Ash, A. N., Seymour, K. J., . . . Thomas, L.

    B. (2011). JPBI 10 years later: Trends in research studies. *Journal of Positive Behavior*

    Interventions, 13(2), 78-86. doi:10.1177/1098300710385346

Perepletchikova, F. (2011). On the topic of treatment integrity. *Clinical Psychology: Science and*

    Practice, 18(2), 148-153. doi: 10.1111/j.1468-2850.2011.01246.x

Perepletchikova, F. (2014). Assessment of treatment integrity in psychotherapy research. In

    Hagermoser Sanetti, L. M., & Kratochwill, T. R. (Eds.), *Treatment integrity: A foundation*

    for evidence-based practice in applied psychology (pp. 131-157). Washington, DC, US:

    American Psychological Association.

Perepletchikova, F., Hilt, L. M., Chereji, E., & Kazdin, A. E. (2009). Barriers to implementing

    treatment integrity procedures: Survey of treatment outcome researchers. *Journal of*

    Consulting and Clinical Psychology, 77(2), 212-218. doi: 10.1037/a0015232

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues

    and research recommendations. *Clinical Psychology: Science and Practice, 12*(4), 365-383.

    doi: 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 & Clinical Psychology, 75(6), 829-841. doi: 10.1037/0022-006X.75.6829

Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables

    in behavior analysis. *Journal of Applied Behavior Analysis, 15*(4), 477-492. doi:

    10.1901/jaba.1982.15-477

Pisecco, S., Huzinec, C., & Curtis, D. (2001). The effect of child characteristics on teachers'

    acceptability of classroom-based behavioral strategies and psychostimulant medication for

    the treatment of ADHD. *Journal of Clinical Child Psychology, 30*(3), 413-421.

Power, T. J., Hess, L. E., & Bennett, D. S. (1995). The acceptability of interventions for

    attention-deficit hyperactivity disorder among elementary and middle school teachers.

    Journal of Developmental and Behavioral Pediatrics, 16(4), 238-243. doi:

    10.1097/00004703-199508000-00005

Quinn, M. M., Rutherford, R. B., Leone, P. E., Osher, D. M., & Poirier, J. M. (2005). Youth with

    disabilities in juvenile corrections: A national survey. *Exceptional Children, 71*(3), 339-345.

    doi: 10.1177/001440290507100308

Reimers, T. M., & Wacker, D. P. (1988). Parents' ratings of the acceptability of behavioral

    treatment recommendations made in an outpatient clinic: A preliminary analysis of the

    influence of treatment effectiveness. *Behavioral Disorders, 14*(1), 7-15.

Reimers, T. M., Wacker, D. P., & Koeppl, G. (1987). Acceptability of behavioral interventions:

    A review of the literature. *School Psychology Review, 16*(2), 212-227.

Reschly, A. L., & Christenson, S. L. (2006). Prediction of dropout among students with mild disabilities: A case for the inclusion of student engagement variables. *Remedial and Special Education, 27*(5), 276-292. doi: 10.1177/07419325060270050301

Reschly, A. L., & Christenson, S. L. (2006). Prediction of dropout among students with mild disabilities: A case for the inclusion of student engagement variables. *Remedial and Special Education, 27*(5), 276-292. doi: 10.1177/07419325060270050301

Reynolds, C. R., & Kamphaus, R. W. (2004). *Manual for the behavior assessment system for children* (2nd ed.). Circle Pines, MN: AGS Publishing.

Reynolds, W. M. (2002). *Reynolds adolescent depression scale-second edition: Professional manual* (2nd ed.). Odessa, FL: Psychological Assessment Resources.

Rhodes, J. E. (2002). *Stand by me: The risks and rewards of mentoring today's youth*. Cambridge, MA, US: Harvard University Press.

Rhodes, J. E. (2005). A model of youth mentoring. In DuBois, D. L., & M. J. Karcher, M. J. (Eds.), *Handbook of youth mentoring* (pp. 30-43). Thousand Oaks, CA: Sage.

Rhodes, J., & Lowe, S. R. (2008). Youth mentoring and resilience: Implications for practice. *Child Care in Practice, 14*(1), 9-17. doi: 10.1080/13575270701733666

Rogers-Warren, A., & Baer, D. M. (1976). Correspondence between saying and doing: Teaching children to share and praise. *Journal of Applied Behavior Analysis, 9*(3), 335-354. doi: 10.1901/jaba.1976.9-335

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581-592.

Sanetti, L. M. H., Dobey, L. M., & Gallucci, J. (2014). Treatment integrity of interventions with children in school psychology international from 1995–2010. *School Psychology International, 35*(4), 370-383. doi: 10.1177/0143034313476399

Sanetti, L. M. H., 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*(1), 29-46. doi: 10.1177/1098300711405853

Sanetti, L. M. H., 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*(1), 72-84.

Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38*(4), 445-459.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147-177. doi:10.1037/1082-989X.7.2.147

Schlosser, R. (2002). On the importance of being earnest about treatment integrity. *Augmentative and Alternative Communication, 18*(1), 36-44.

Schneider, B. H., Kerridge, A., & Katz, J. (1992). Teacher acceptance of psychological interventions of varying theoretical orientation. *School Psychology International, 13*(4), 291-305. doi: 10.1177/0143034392134001

Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review, 38*(4), 460-475.

Schwartz, I. S., & Baer, D. M. (1991). Social validity: Is current practice state of the art? *Journal of Applied Behavior Analysis, 24*(2), 189-204. doi: 10.1901/jaba.1991.24-189

Schwartz, S. E., Rhodes, J. E., Chan, C. S., & Herrera, C. (2011). The impact of school-based mentoring on youths with different relational profiles. *Developmental psychology*, *47*(2), 450.

Seidel, J. F., & Vaughn, S. (1991). Social alienation and the learning disabled school dropout. *Learning Disabilities Research & Practice, 6*(3), 152-157.

Seligson, J. L., Huebner, E. S., & Valois, R. F. (2003). Preliminary validation of the brief multidimensional students' life satisfaction scale (BMSLSS). *Social Indicators Research, 61*(2), 121-145.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA, US: Houghton, Mifflin and Company.

Sinclair, M. F., Christenson, S. L., Lehr, C. A., & Anderson, A. R. (2003). Facilitating student engagement: Lessons learned from check & connect longitudinal studies. *The California School Psychologist, 8*(1), 29-41.

Sinclair, M. F., Christenson, S. L., Elevo, D. L., & Hurley, C. M. (1998). Dropout prevention for youth with disabilities: Efficacy of a sustained school engagement procedure. *Exceptional Children, 65*(1), 7-21.

Sinclair, M. F., Christenson, S. L., Lehr, C. A., & Anderson, A. R. (2003). Facilitating student engagement: Lessons learned from check & connect longitudinal studies. *California School Psychologist, 8*, 29-41.

Sinclair, M. F., Christenson, S. L., & Thurlow, M. L. (2005). Promoting school completion of urban secondary youth with emotional or behavioral disabilities. *Exceptional Children, 71*(4), 465-482.

Smith, S. W., Daunic, A. P., & Taylor, G. G. (2007). Treatment fidelity in applied educational research: Expanding the adoption and application of measures to ensure evidence-based practice. *Education & Treatment of Children, 30*(4), 121-134.

Spencer, R., & Liang, B. (2009). "She gives me a break from the world": Formal youth

    mentoring relationships between adolescent girls and adult women. *The Journal of Primary*

    Prevention, *30*(2), 109-130.

Sterling-Turner, H. E., & Watson, T. S. (2002). An analog investigation of the relationship

    between treatment acceptability and treatment integrity. *Journal of Behavioral Education,*

    11(1), 39-50.

Stinnett, T. A., Crawford, S. A., Gillespie, M. D., Cruce, M. K., & Langford, C. A. (2001).

    Factors affecting treatment acceptability for psychostimulant medication versus

    psychoeducational intervention. *Psychology in the Schools, 38*(6), 585-591. doi:

    10.1002/pits.1045

Strain, P. S., Barton, E. E., & Dunlap, G. (2012). Lessons learned about the utility of social

    validity. *Education and Treatment of Children, 35*(2), 183-200.

Sugai, G., Lewis-Palmer, T., Todd, A. W., & Horner, R. H. (2001). School-wide evaluation tool

    (SET). *Eugene, OR: Center for Positive Behavioral Supports, University of Oregon,*

Sum, A., Khatiwada, I., McLaughlin, J., & Palma, S. (2009). The consequences of dropping out

    of high school. *Center for Labor Market Studies Publications* (Paper 23). Retrieved from

    http://hdl.handle.net/2047/d20000596

Turco, T. L., & Elliott, S. N. (1986a). Assessment of students' acceptability ratings of teacher-

    initiated interventions for classroom misbehavior. *Journal of School Psychology, 24*(3),

    277-283.

Turco, T. L., & Elliott, S. N. (1986b). Students' acceptability ratings of interventions for

    classroom misbehaviors: A study of well-behaving and misbehaving youth. *Journal of*

    Psychoeducational Assessment, 4(4), 281-289.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse (2011, October). *Children Classified as Having an Emotional Disturbance intervention report: Check & Connect*. Retrieved from http://whatworks.ed.gov.

U.S. Department of Education, National Center for Education Statistics (2014, January). *Fast Facts*. Retrieved from http://nces.ed.gov/fastfacts/display.asp?id=16.

U.S. Department of Labor Statistics (2012, February). *Bureau of Labor Statistics* (BLS). Retrieved from www.bls.gov/data

Vereb, R. L., & DiPerna, J. C. (2004). Teachers' knowledge of ADHD, treatments for ADHD, and treatment acceptability: An initial investigation. *School Psychology Review, 33*(3), 421-428.

Von Brock, M. B., & Elliott, S. N. (1987). Influence of treatment effectiveness information on the acceptability of classroom interventions. *Journal of School Psychology, 25*(2), 131-144. doi: http://dx.doi.org/10.1016/0022-4405(87)90022-7

Wagner, M., Newman, L., D'Amico, R., Jay, E. D., Butler-Naylin, P., Marder, C., & Cox, R. (1991). Youth with disabilities: How are they doing? The first comprehensive report from the national longitudinal transition study of special education students. (SRI International, Contract No. 300-87-0054). Washington, DC: U.S. Department of Education, Office of Special Education Programs.

Wheeler, M., Keller, T., & DuBois, D. L. (2010). Review of three recent randomized trials of school-based mentoring. Social Policy Report, 24, 1–21. Retrieved from http://www .srcd.org

Wilkinson, L. A. (2007). Assessing treatment integrity in behavioral consultation. *International Journal of Behavioral Consultation and Therapy, 3*(3), 420-432. doi: 10.1037/h0100816

Winett, R. A., Moore, J. F., & Anderson, E. S. (1991). Extending the concept of social validity: Behavior analysis for disease prevention and health promotion. *Journal of Applied Behavior Analysis, 24*(2), 215-230. doi: 10.1901/jaba.1991.24-215

Witt, J. C., & Elliott, S. N. (1985). Acceptability of classroom intervention strategies. In TR Kratochwill, T. R. (Ed.), *Advances in School Psychology, (Vol. 4*, 251-288). NJ: Lawrence Erlbaum.

Witt, J. C., & Robbins, J. R. (1985). Acceptability of reductive interventions for the control of inappropriate child behavior. *Journal of Abnormal Child Psychology, 13*(1), 59-67.

Witt, J. C., Elliott, S. N., & Martens, B. K. (1984). Acceptability of behavioral interventions used in classrooms: The influence of amount of teacher time, severity of behavior problem, and type of intervention. *Behavioral Disorders, 9*(2), 95-104.

Witt, J. C., & Martens, B. K. (1983). Assessing the acceptability of behavioral interventions used in classrooms. *Psychology in the Schools, 20*(4), 510-517. doi: 10.1002/1520-6807(198310)20:4<510::AID-PITS2310200420>3.0.CO;2-1

Witt, J. C., Martens, B. K., & Elliott, S. N. (1984). Factors affecting teachers' judgments of the acceptability of behavioral interventions: Time involvement, behavior problem severity, and type of intervention. *Behavior Therapy, 15*(2), 204-209. doi: 10.1016/S0005-7894(84)80022-2

Witt, J. C., Moe, G., Gutkin, T. B., & Andrews, L. (1984). The effect of saying the same thing in different ways: The problem of language and jargon in school-based consultation. *Journal of School Psychology, 22*(4), 361-367. doi: http://dx.doi.org/10.1016/0022-4405(84)90023-2

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied

    behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*(2), 203-214.

    doi: 10.1901/jaba.1978.11-203

Woodcock, R. W., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson tests of achievement*

    Itasca, IL: Riverside Publishing.

Woolley, M. E., & Bowen, G. L. (2007). In the context of risk: Supportive adults and the school

    engagement of middle school students. *Family Relations, 56*(1), 92-104. doi:

    10.1111/j.1741-3729.2007.00442.x

Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of

    successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and*

    Clinical Psychology, 49(2), 156-167. doi: 10.1037/0022-006X.49.2.156

Zablocki, M., & Krezmien, M. P. (2013). Drop-out predictors among students with high-

    incidence disabilities: A national longitudinal and transitional study 2 analysis. *Journal of*

    Disability Policy Studies, 24(1), 53-64. doi: 10.1177/1044207311427726

Zand, D. H., Thomson, N., Cervantes, R., Espiritu, R., Klagholz, D., LaBlanc, L., & Taylor, A.

    (2009). The mentor–youth alliance: The role of mentoring relationships in promoting youth

    competence. *Journal of Adolescence, 32*(1), 1-17. doi:

    http://dx.doi.org/10.1016/j.adolescence.2007.12.006

Zimmerman, M. A., Bingenheimer, J. B., & Notaro, P. C. (2002). Natural mentors and

    adolescent resiliency: A study with urban youth. *American Journal of Community*

    Psychology, 30(2), 221-243. doi: 10.1023/A:1014632911622

Table 1

Pooled, Descriptive Statistics for Each Variable (N = 339)

Variable	Overall M (SE)	Special Education Status		Change in Mentor	
		With	Without	Same	Change
Students					
Year 2 EPC	62.16 (1.02)	64.64 (0.19)	59.36 (0.19)	63.98 (0.22)	61.00 (0.17)
Year 2 IPC	56.07 (0.71)	58.85 (0.13)	52.91 (0.13)	57.00 (0.16)	55.47 (0.12)
Year 1 Dosage (%)	70.65 (1.72)	72.11 (0.32)	69.00 (0.36)	77.29 (0.32)	66.41 (0.33)
Year 2 Dosage (%)	65.27 (1.53)	65.72 (0.28)	64.77 (0.30)	70.07 (0.29)	62.21 (0.27)
Two-Year Dosage (%)	67.96 (1.27)	68.92 (0.23)	66.89 (0.26)	73.68 (0.24)	64.31 (0.23)
Year 1 Student SIRF	33.94 (0.43)	33.60 0.08	34.32 (0.08)	34.11 (0.08)	33.82 (0.08)
Year 2 Student SIRF	34.86 (0.38)	34.51 (0.07)	35.26 (0.07)	35.19 (0.07)	34.65 (0.06)
Mentors					
Year 2 Mentor SIRF	112.12 (0.81)	113.73 (0.13)	110.31 (0.16)	110.92 (0.15)	112.90 (0.14)
Year 2 Treatment Integrity (%)	80.13 (1.72)	82.01 (0.29)	77.99 (0.33)	78.64 (0.36)	81.07 (0.28)
Mentors' Years of Teaching	5.24 (0.45)	5.53 (0.07)	4.92 (0.07)	7.18 (0.08)	4.01 (0.06)

Note. EPC = Externalizing Problem Composite score; IPC = Internalizing Problem Composite score; SIRF = School Intervention Rating Form; mean difference between with and without special education status across variables are significant, *p* <.01; mean difference between same and change in mentor were significant, *p* < .01.

Table 2

Pooled, Correlation Matrix and Simultaneous Multiple Regression Analysis Summary for Students' Year 2 SIRF and Its Predictors (N = 339)

Variable	1	2	3	B	SE
Students' Year 2 SIRF	.36**	.16	-.05		
Predictor					
Constant				22.79	2.91
1. Students' Year 1 SIRF		.11	-.02	0.29	0.07
2. Two Year Dosage			-.23	0.03	0.02
3. Change Mentor				-0.14	0.82
(1 = change; 0 = no change)					

Note. ** $p < .001$. $R^2 = .15$. SIRF = School Intervention Rating Form.

Table 3

Pooled Correlation Matrix for Students' Year 2 SIRF and Its Predictors (N = 339)

Variable	1	2	3	4	5
Students' Year 2 SIRF	.12	-.08	-.02	-.07	.001
1. Year 2 Dosage		-.07	-.08	.02	.15
2. Externalizing Behavior			.42**	.17*	.05
3. Internalizing Behavior				.27**	-.01
4. Special Ed. Status (1 = with; 0 = without)					.06
5. Mentors' Years of Teaching					

Note. * $p < .05$; ** $p < .001$; SIRF = School Intervention Rating Form.

Table 4

Pooled, Hierarchical Multiple Regression Analysis Summary for Students' Year 2 SIRF and Its Predictors (N = 339)

Variable	Step 1			Step 2			Step 3		
	B	SE	p	B	SE	p	B	SE	p
Constant	33.08	1.41		34.23	2.79		34.23	2.79	
1. Year 2 Dosage		0.02	.17	0.03	0.02	.19	0.03	0.02	.19
2. Externalizing Behavior				-0.03	0.03	.37	-0.03	0.03	.37
3. Internalizing Behavior				0.02	0.04	.67	0.02	0.04	.67
4. Special Ed. Status (1 = with; 0 = without)				-0.72	0.81	.38	-0.72	0.81	.38
5. Mentors' Years of teaching							-0.01	0.09	.93
R^2		.02			.03			.04	
ΔR^2					.01			.01	

Table 5

Pooled, Correlation Matrix between Mentors' Year 2 SIRF and Its Predictors (N = 339)

Variable	1	2	3	4	5
Mentors' Year 2 SIRF	.02	-.07	.04	.14	-.02
1. Year 2 Dosage		-.07	-.08	.14	-.02
2. Externalizing Behavior			.42**	.17*	.05
3. Internalizing Behavior				.27**	-.01
4. Special Ed. Status (1 = with; 0 = without)					.06
5. Mentors' Years of Teaching					

*Note. * p < .05; ** p < .001.*

Table 6

Pooled, Hierarchical Multiple Regression Summary for Mentors' Year 2 SIRF and Its Predictors (N = 339)

Variable	Step 1			Step 2			Step 3		
	B	SE	p	B	SE	p	B	SE	p
Constant	111.38	2.96		112.19	6.70		112.19	6.70	
1. Year 2 Dosage		0.04	.78	0.01	0.04	.83	0.01	0.04	.83
2. Externalizing Behavior				-0.09	0.07	.20	-0.09	0.07	.20
3. Internalizing Behavior				0.05	0.09	.57	0.05	0.09	.57
4. Special Ed. Status (1 = with; 0 = without)				3.54	1.84	.06	3.54	1.84	.06
5. Mentors' Years of teaching							-0.04	0.17	.81
R^2		-			.04			.05	
ΔR^2					.04			.01	

Table 7

Pooled, Correlation Coefficient and Simple Regression Summary for Mentors' Year 2 SIRF and Their Year 2 Treatment Integrity (N = 339)

Variable	1	B	SE
Year 2 Treatment Integrity	.31**		
Predictors			
Constant		4.67	17.14
1. Year 2 Mentor SIRF		0.67	0.15

*Note. ** p < .001. R^2 = .10; SIRF = School Intervention Rating Form.*
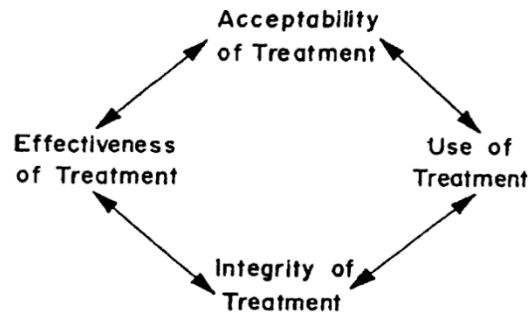
Figure 1. Witt and Elliott's (1985) model of treatment acceptability
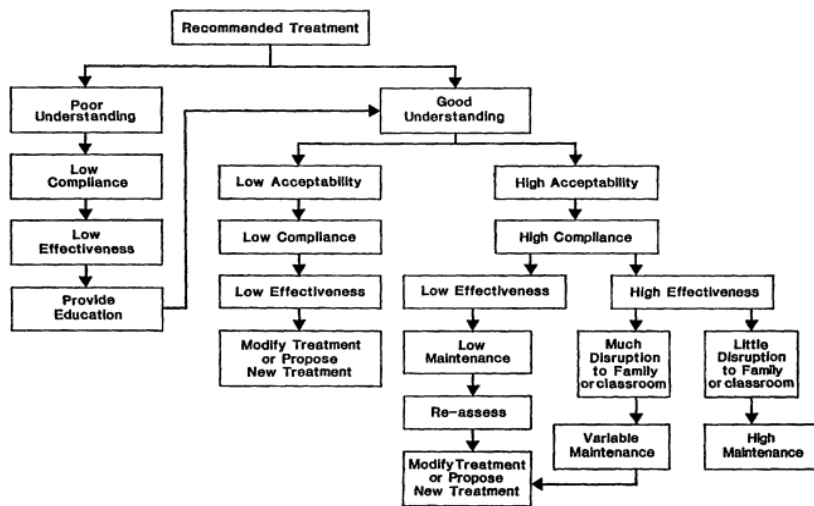
Figure 2. Reimers, Wacker, and Koeppl's (1987) model of treatment acceptability
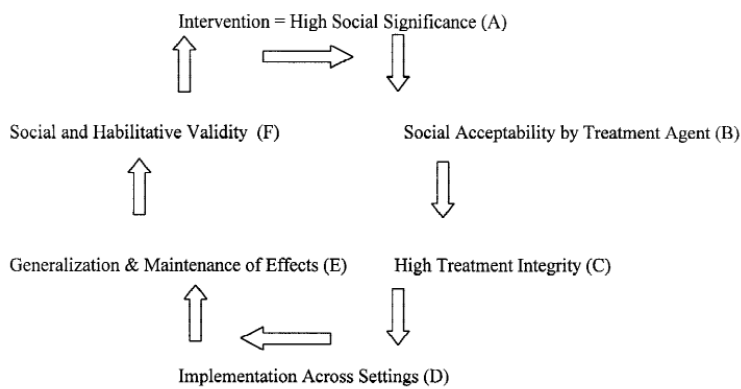
Figure 3. The self-reinforcing model of social validity

123

CARS
Check & Connect
TREATMENT INTEGRITY FORM

Draft

Student ID # : [ ][ ][ ][ ][ ][ ]   Site ID: [ ][ ]   Date : [ ][ ] / [ ][ ] / [ ][ ][ ][ ]

Teacher ID # : [ ][ ][ ][ ][ ][ ]   School ID: [ ][ ]   CARS Staff ID : [ ][ ][ ][ ][ ][ ][ ]

Integrity Number: 1 ○   2 ○   3 ○   4 ○   5 ○   6 ○   7 ○   8 ○   9 ○   Booster Number: B1 ○   B2 ○   B3 ○

Instructions: Collect a copy of the Check and Connect monitoring form at the end of the month. Based on the information recorded by the teacher on the form, mark Yes or No. At least once per month, observe a Check & Connect Session in person. Please indicate below which week you conducted your observation. If a mentor sees more than one CARS student, you only need to observe one session for the mentor per month (not one per student per month).

	WEEK 1		WEEK 2		WEEK 3		WEEK 4	
Student available for mentoring each week? If "N" bubbled in any of the weeks, record the number representing the reason why student was not available (1=Absent, 2=Suspended, 3= Competing school activity-e.g., testing; 4=Student doing well; weekly meetings not necessary) Look for student initials on permanent product for confirmation of mentoring process.	Yes ○	No ○	Yes ○	No ○	Yes ○	No ○	Yes ○	No ○
		1 ○		1 ○		1 ○		1 ○
		2 ○		2 ○		2 ○		2 ○
		3 ○		3 ○		3 ○		3 ○
		4 ○		4 ○		4 ○		4 ○

	WEEK 1		WEEK 2		WEEK 3		WEEK 4	
	O	P	O	P	O	P	O	P
Observed Session? Bubble "O" for observed & "P" for permanent product.	○	○	○	○	○	○	○	○

Mentor met with the student. (See teacher and student initials on C&C monitoring form).

	WEEK 1		WEEK 2		WEEK 3		WEEK 4	
1. Bubble **Y** if mentor met with the student _at least one time_ per week. If **"Y"**, continue to question 2. If **"N"**, mark all questions below **"N"** and implement booster session.	Y	N	Y	N	Y	N	Y	N
	○	○	○	○	○	○	○	○

Figure 4. Check & Connect treatment integrity checklist

124

```
                    ┌─────────────────────┐
                    │  Not met/Excluded   │
                    │    (N = 201)        │
                    └─────────────────────┘
┌──────────────┐                              ┌─────────────────────┐
│              │                              │  Treatment Group    │
│ Four criteria│                              │    (N = 341)        │
│              │                              └─────────────────────┘
└──────────────┘    ┌─────────────────────┐
                    │ Met/Included (N=647)│   ┌─────────────────────┐
                    │                     │   │   Control Group     │
                    └─────────────────────┘   │    (N = 306 )       │
                                              └─────────────────────┘
```

Figure 5. Number of students participated in CARS during each phase

Check & Connect Monitoring Sheet Key: Write in totals per day (or week) OR: ✓=Indicates behavior is occurring /= Indicates behavior is not occurring

Month_____ Student _____ ID_____ Mentor _____

CHECK	M	T	W	Th	F	Wk	M	T	W	Th	F	Wk	M	T	W	Th	F	Wk	M	T	W	Th	F	Wk	M	T	W	Th	F	Wk	Risk Indicator ∨	
Tardy																															≥ 5	
Skip																															≥ 3	
Absent																															≥ 4	
Behavior referral																															≥ 4	
Detention																															≥ 2	
In-school suspension																															≥ 2	
Out-of-school suspension																															≥ 2	

Failing classes (monthly) _____ D's _____ F's — risk indicator ≥ 1 F and/or ≥ 2 D's per grading period

Behind in credits (monthly) _____ Credits earned out of _____ total — earning <80% of possible credits per grading period

Missing Assignments (bi-monthly-no. assign. Completed /total no. of assign.) — English | Math | Social Studies | Science | Other — Turning in assignments an average of below 90% across classes

CONNECT

BASIC

	M	T	W	Th	F	Wk	M	T	W	Th	F	Wk	M	T	W	Th	F	Wk	M	T	W	Th	F	Wk	M	T	W	Th	F	Wk
Provided regular feedback																															
Problem-solved about risk																															

INTENSIVE

	M	T	W	Th	F	Wk	M	T	W	Th	F	Wk	M	T	W	Th	F	Wk	M	T	W	Th	F	Wk	M	T	W	Th	F	Wk
Organizational Skills																															
Study Skills																															
_____																															
_____																															
_____																															

< High risk for month

List classes where student is currently:

1. receiving a grade of D or lower: _____ _____ _____ _____ _____

2. has 10% or more missing assignments: _____ _____ _____ _____ _____

3. has had discipline problems (e.g., received ODR, teacher reported problem behaviors): _____ _____ _____ _____ _____

(CARS USE ONLY) List classes indicated for CARS intervention (where any of the above indicators occurred): _____ _____ _____ _____ _____

Core Components: 1. Record progress for all risk areas as indicated.
2. Schedule a set time to meet with the student weekly.
3. Meet for at least 10 – 15 minutes.
4. Connect and problem-solve high risk areas.

Record focus of meeting each week. Provide positive feedback for improvements or continued success. Problem-solve for risk areas. Initial at the close of the meeting and confirm next check-in date.	Mentor Initials	Student Initials
Week of		
Week of		
Week of		
Week of		
Week of		

Figure 6. Check & Connect monitoring sheet

126

Treatment Integrity Problem Solving/Booster Session Planning

Purpose: Teacher is implementing the intervention below determined integrity percentage. Meet with teacher to discuss the intervention regarding missing/incorrect steps, or reasons why intervention is not in place.

Re-Training: Are additional explanations of specific components and steps needed for the teacher, staff, or student? — Yes → Review intervention procedures; explain basic components and steps of intervention

No ↓

Are in vivo reminders needed to implement or carry out the steps of the intervention? — Yes → Implement prompts to remind teacher, staff, or peers to use intervention

No ↓

Is additional explicit training in the form of modeling needed for the teacher, staff, or student? — Yes → Model the intervention for the teacher, staff, or student

No ↓

Feasibility: Can the teacher/staff feasibly implement the intervention in the given situation? — No → ☐ Are there portions or steps of the intervention that can be modified? ☐ Can other teachers, staff, or peers implement the intervention? — Yes → Make appropriate changes to the current intervention — No → Go to "Implementation" box below

Yes ↓

Acceptability: Is intervention acceptable to the teacher and/or student? — No → ☐ Can the intervention be made more acceptable to the teacher or student? — Yes → Make appropriate changes to the current intervention — No

Yes ↓

Implementation: Have two booster sessions been tried and teacher still does not implement intervention with determined integrity? — No → Implement booster session and monitor treatment integrity with 3 consecutive observations

Yes ↓

Discontinue additional training; continue to monitor monthly and consider the next recommended intervention based on classroom-assessment

Collect 3 additional treatment integrity checks and schedule another problem solving meeting if integrity is below determined % on any of the observations
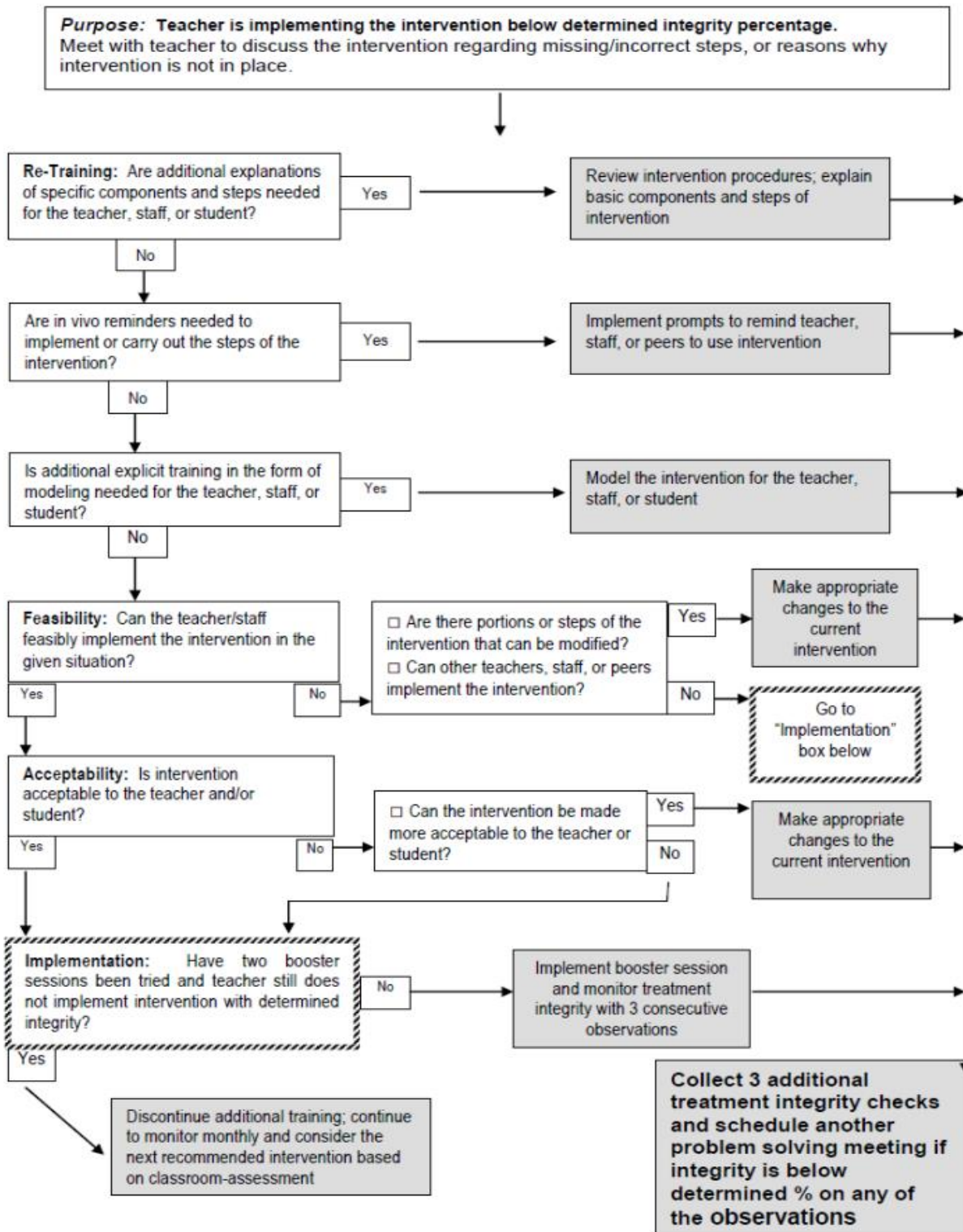
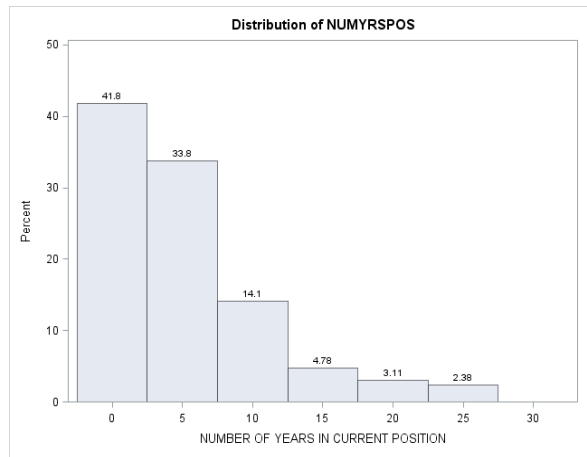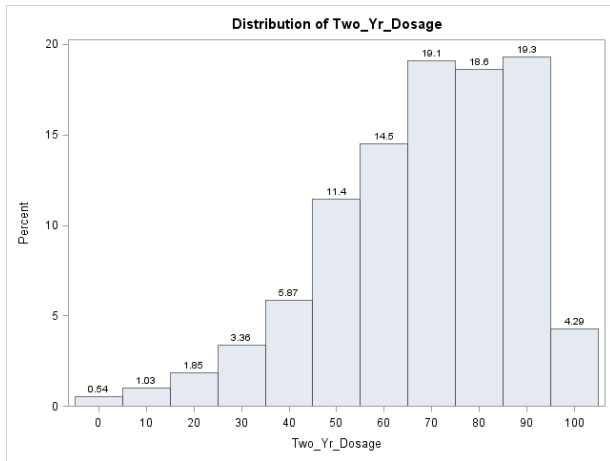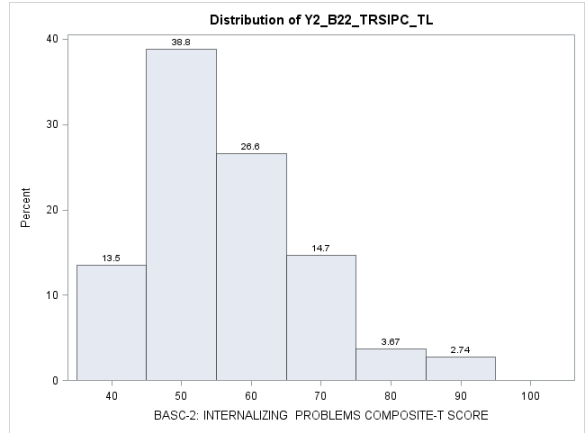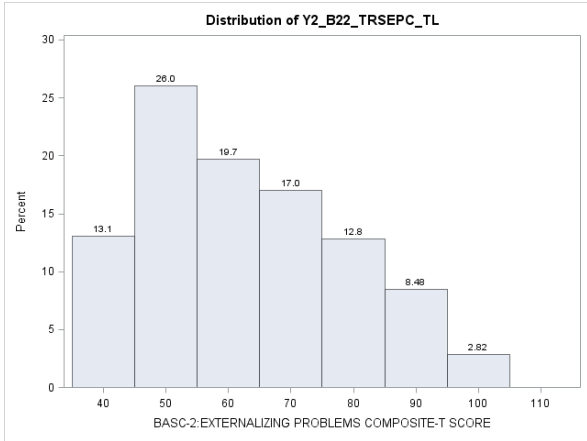*Figure 7*. Check & Connect treatment integrity problem solving process

Figure 8. Histograms of externalizing behavior (left top panel), internalizing behavior (right top panel), two-year dosage (left bottom panel), and mentors' years of teaching experience (right bottom panel)
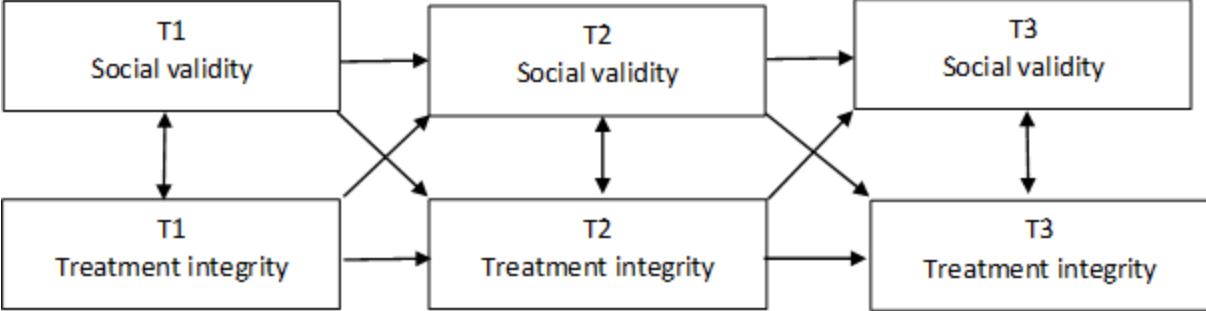
Figure 9. Cross-lagged model of social validity and treatment integrity

Shu-Chen Tsai

EDUCATION

2006   **California State University at Los Angeles**
            Master of Arts in Special Education

1998   **National Changhua University of Education**, Taiwan, R.O.C.
            Bachelor of Education in Special Education

CERTIFICATION

2007   Board Certified Behavior Analyst (BCBA)
            Certification number: 1-07-3294

1999   Certified Secondary Special Education in Taiwan
            Certification number: 特中第 8800874 號

RESEARCH EXPERIENCE: Autism and EBD

2010-Present   Graduate Research Assistant, Lehigh University
               Peer-Mediated Social Skills Project (PI: Dr. Linda Bambara)

2009-2010   Graduate Assistant, Centennial School of Lehigh University

PUBLICATIONS

Tsai, S., Kern, L., & George, M. P. (in preparation). Further evaluation of the effects of repeated reading on middle school students with emotional and behavioral disorders.

Bambara, L., Cole, C. L., Kunsch, C, **Tsai, S**., & Ayad, E. (2016). Improving the conversational skills of high school students with autism spectrum disorder. *Research in Autism Spectrum Disorders, 27,* 29-43. doi: 10.1016/j.rasd.2016.03.003

Wood, K. B., Drogan, R. R., **Tsai, S**., Sharp, E., & Custer, B. (in preparation). Comparing the effects of content and behavioral priming on the engagement of young children: A preliminary study.

2012   Book chapter translation into Chinese:

Cooper, J. O.、Heron, T. E. 、Heward, W. L. (2007)。篩選與定義目標行為 (蔡淑珍譯)。應
    用行為分析(上下) (鳳華等編譯)。台中市: 美國展望教育中心。
[Cooper, J. O., Heron, T. E., Heward, W. L. (2007). Selecting and defining target behaviors. In
    Cooper, J. O., Heron, T. E., & Heward, W. L. (Eds). *Applied Behavior Analysis* (2nd ed.)
    (pp. 48-71). New Jersey: Pearson.]

EDITORIAL EXPERIENCE

2012   Guest Reviewer for Journal of Behavioral Education

PRESENTATIONS

Tsai, S. & Kern, L. (2016, May). *Revisiting Social Validity and Treatment Integrity: A View from
    Empirical Data.* Poster will be presented at the 42nd Annual Convention of the Association
    for Behavior Analysis International, Downtown Chicago, IL.

Bambara, L. M., Chovanes, J., Thomas, A., & **Tsai, S**. (2016, April). *Effective Peer Strategies
    for Enhancing the Conversational Skills of Adolescents With Autism*. Paper presented at the
    Conference of the Council for Exceptional Children, St. Louis, MO.

Tsai, S. & Kern, L. (2016, March). *An Investigation of Social Validity and Treatment Integrity of
    Check & Connect.* Paper will be presented at the 13th International Conference of the
    Association for Positive Behavior Support, San Francisco, CA.

Bambara, L. M., Cole, C., Chovanes, J., Thomas, A., Phipps, K., **Tsai, S**., Ayad, E., & Telesford,
    A. (2016, January). *Improving the Conversational Skills of High School Students with
    Autism through Peer Mediation*. Poster was presented at the 10th Annual Autism Conference
    of the Association for Behavior Analysis International, New Orleans, LA.

Bambara, L. M., Cole, C., Kunsch, C., **Tsai, S**., & Ayad, E. (2014, April). *Helping High School
    Students With Autism Become Active Conversationalists Through Peer-Mediation*. Paper
    presented at the Conference of the Council for Exceptional Children, Philadelphia, PA.

Bambara, L. M., Cole, C., Kunsch, C., **Tsai, S**., & Ayad, E. (2014, March). *Helping High School
    Students with Autism Become Active Conversationalists through Peer-Mediation*. Paper
    presented at the 11th International Conference of the Association for Positive Behavior
    Support, Chicago, IL.

Bambara, L. M., Cole, C., Kunsch, C., **Tsai, S**., & Ayad, E. (2013, December). *Peer Intervention
    to Improve the Conversational Skills of High Schoolers with Autism*. Paper presented at

TASH 2013 Conference, Chicago, IL.

Bambara, L. M., Cole, C., Kunsch, C., Drogan, R., & **Tsai, S**. (2013, March). *Supporting the Conversational Skills of High School Students with Autism through Peer-Mediation*. Paper presented at the 10th International Conference of the Association for Positive Behavior Support, San Diego, CA.

Bambara, L. M., Cole, C., Kunsch, C., Drogan, R., & **Tsai, S**. (2012, November). *Teaching Peers to Support the Conversational Skills of High School Students with Autism*. Paper presented at TASH 2012 Conference, Long Beach, CA.

Tsai, S., Beiter, K., George, M.P., & Kern, L. (2010, April). *Further validation of repeated reading on students with emotional and behavioral disorders*. Poster presented at the Conference of the Council for Exceptional Children, National Harbor, MD.

Tsai, S., Symon, J., Menzies, H., & Mayer, G. (2007, May). *Increasing Social Interactions and Classroom Behavior of Children with Autism through Social Story Intervention Package*. Poster presented at the 33rd Annual Convention of the Association for Behavior Analysis International, San Diego, CA.

Tsai, S., Symon, J., Menzies, H., & Mayer, G. (2007, March). *An Effective Intervention Package to Increase Appropriate Behavior for Children with Autism*. Poster presented at the 4th International Conference of the Association for Positive Behavior Support, Boston, MA.

Tsai, S., Symon, J., Menzies, H., & Mayer, G. (2007, Feb). *Social Story Intervention Package on Increasing Social Interactions and Classroom Behavior of Children with Autism*. Poster presented at the 25th Annual Conference of the California Association for Behavior Analysis, San Francisco, CA.

TRAINING WORKSHOPS CONDUCTED

Bambara, L. M., Ayad, E., Lubar, A., & **Tsai, S.** (April 8, 2014). *Peer Intervention to Improve the Conversation Skills of High School Students with Autism*. Webinar training presented at TASH.

COLLEGE TEACHING/ADVISING EXPERIENCE

Graduate-level courses co-taught with faculty at Lehigh University:

2016   **Single-Subject Research Design** (EDUC 461)

2013   **Positive Behavior Support** (SpEd 432)

2011   **Classroom Management** (SpEd 444)
               Academic Interventions: PreK-8 (SpEd 419)

2010   **Advanced Inclusionary Practices in K-12** (SpEd 465)
                 Applied Behavior Analysis (SchP/SpEd 402)

Undergrad-level advising at California State University, Los Angeles:

2004-2005   **Peer Mentor**
             Partnership for Academic Learning and Success (PALS) Mentoring Program

PreK-12 WORKING EXPERIENCE

2008- 2009   Behavior Consultant
             SEEK Education, Inc., San Gabriel, CA

2006- 2008   Behavior Interventionist
             Autism Spectrum Therapies, Culver City & Arcadia, CA

1999- 2003   Special Education Teacher
             Tz-Chiang Junior High School, Taipei, Taiwan
             Chin-Shan Junior High School, Taipei, Taiwan

1998- 1999   Special Education Intern
             National Changhua Special School, Changhua, Taiwan

AWARDS/GRANTS

2010   Multicultural Center Grant ($1000), Lehigh University
2010   The Villas-Woodring Scholarship, Lehigh University
2009   The Villas-Woodring Scholarship, Lehigh University
2005   International Student Tuition Waiver ($6000), California State University at Los Angeles
2004   Apple Award for Best Peer Mentor, California State University at Los Angeles

UNIVERSITY SERVICE

2012   Tutor for Lehigh Prison Project, Lehigh University
2010-2011   Department representative for Graduate Student Senate, Lehigh University

PROFESSIONAL MEMBERSHIPS

Association for Behavior Analysis International (ABAI)
Association for Positive Behavior Support (APBS)
Council of Exceptional Children (CEC)
TASH