Published in final edited form as:

Prev Sci. 2015 January ; 16(1): 122–132. doi:10.1007/s11121-014-0488-9.

Observational Measures of Implementer Fidelity for a School-based Preventive Intervention: Development, Reliability and Validity

Wendi Cross¹, Jennifer West¹, Peter A. Wyman¹, Karen Schmeelk-Cone¹, Yinglin Xia², Xin Tu², Michael Teisl¹, C. Hendricks Brown³, and Marion Forgatch⁴

¹Department of Psychiatry, University of Rochester School of Medicine and Dentistry

²Department of Biostatistics and Computational Biology, University of Rochester School of Medicine and Dentistry

³Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine

⁴Oregon Social Learning Center

Abstract

Current measures of implementer fidelity often fail to adequately measure core constructs of adherence and competence, and their relationship to outcomes can be mixed. To address these limitations, we used observational methods to assess these constructs and their relationships to proximal outcomes in a randomized trial of a school-based preventive intervention (Rochester Resilience Project) designed to strengthen emotion self-regulation skills in 1st–3rd graders with elevated aggressive-disruptive behaviors. Within the intervention group ($n = 203$), a subsample ($n = 76$) of students was selected to reflect the overall sample. Implementers were 10 paraprofessionals. Videotaped observations of three lessons from Year 1 of the intervention (14 lessons) were coded for each implementer-child dyad on Adherence (content) and Competence (quality). Using multi-level modeling we examined how much of the variance in the fidelity measures was attributed to implementer and to the child within implementer. Both measures had large and significant variance accounted for by implementer (Competence, 68%; Adherence, 41%); child within implementer did not account for significant variance indicating that ratings reflected stable qualities of the implementer rather than the child. Raw Adherence and Competence scores shared 46% of variance ($r = .68$). Controlling for baseline differences and age, the amount (Adherence) and quality (Competence) of program delivered predicted children's enhanced response to the intervention on both child and parent reports after six months, but not on teacher report of externalizing behavior. Our findings support the use of multiple observations for measuring fidelity and that adherence and competence are important components of fidelity which could be assessed by many programs using these methods.

Keywords

implementer fidelity; measurement; adherence; competence; observation

There is general agreement that implementation quality has an impact on intervention outcomes (Aarons et al., 2010; Brownson et al., 2012; Durlak & DuPre, 2008; Fixsen et al., 2005; Glasgow et al., 2012; Proctor et al., 2011). Several frameworks have been proposed to describe the factors that contribute to the implementation process (e.g., Curran et al., 2008; Damschroder et al., 2009; Glasgow, & Boles, 1999; Stetler, et al., 2008). Implementation fidelity is fundamental to these models. Implementer fidelity focuses on program delivery by the individual interventionist. Currently, two primary constructs are considered to be central to the measurement of implementer fidelity: a) *adherence* to program content as specified in manuals, and, b) *competence* in program delivery which describes the quality of the implementation (Breitenstein, et al. 2010; Cross & West, 2011; Dumas et al., 2001; Durlak & DuPre, 2008; Forgatch & DeGarmo, 2011; Hepner et al., 2011; Hogue et al., 2008; Schoenwald, et al., 2010; Waltz et al., 1993). A third component, *treatment differentiation*, is the degree that treatments differ from one another on critical dimensions and is used to test the outcomes of two or more treatments (Schoenwald, 2010). Conceptually, it is important to consider both implementer adherence and competence because low levels of adherence or competence separately, or both together, may be related to poor outcomes. For example, low levels of adherence to the manual would compromise a test of program outcomes because the intervention content would not have been delivered and conclusions about effectiveness would be erroneous. On the other hand, an implementer may deliver a high degree of program content, and thus be rated as highly adherent, but in a way that is not consistent with the program model's expectation for quality or competence. Low levels of competence, such as insensitivity to the individual/group, rote program delivery that compromises the transactional nature of the intervention, and poor intervention timing are all likely to have an impact on outcomes (Cross & West, 2011). Studies where high ratings of adherence were associated with poor outcomes may be clarified if competence had been measured and found to be low (e.g., James et al., 2001).

Although implementer fidelity is theoretically related to clinical outcomes, studies have shown mixed findings regarding this relationship (Byrnes et al., 2010; Hamre et al., 2010; Webb et al., 2010). What accounts for this? One factor may be related to the failure to measure both constructs, adherence to content and competence, in implementer fidelity (Liber et al., 2010). Measurement methods may also account for the apparent disconnect between implementer fidelity and program outcomes. For example, fidelity measures based on implementer self-reports of fidelity are likely to be biased and positively skewed (Carroll et al., 2000; Lillehoj et al., 2004; Moore et al., 2009). Consequently, videotaped or audiotaped interactions, coded by independent observers, are currently the optimal standard for fidelity assessment (Schoenwald & Garland, 2013; Schoenwald et al., 2011; Synder et al., 2006). In some cases, low measurement reliability and validity may also contribute to the mixed findings regarding the relationship between program fidelity and outcomes (Schoenwald, 2011).

The present study

The current study sought to address gaps in fidelity measurement and contribute to the understanding of the relationship between fidelity and outcomes in the context of a

randomized trial of a school-based preventive intervention delivered by paraprofessionals (Rochester Resilience Project, RRP; Cross & West, 2011; Wyman et al., 2010; Wyman et al., 2013). Our first research aim was to develop reliable observational fidelity measures of implementer adherence and competence, and to examine the relationship between them. To do this, we first used rigorous observational methods to code 206 sessions on both adherence and competence and analyzed these data to assess variation across implementer, lesson, and child. This was used to conduct formal tests that the underlying constructs of adherence and competence were reasonably constant over lesson and child. We wanted assurance that the measures captured implementer behaviors rather than other factors such as child variables. The second aim was to conduct an initial test of criterion-related validity of these measures by using them to predict improvements in child behavior after one year of the two-year intervention. Although the two measures developed may be specific to our intervention (particularly for adherence), the methods are generalizable to other program developers and researchers.

Method

The intervention: The Rochester Resilience Project

This study of implementer fidelity occurred in the context of a randomized trial of the Rochester Resilience Project (RRP; Wyman et al., 2010; Wyman et al., 2013), a school-based indicated preventive intervention for 1st–3rd grade students with elevated aggressive-disruptive behaviors, which are associated with long-term problems (Kellam et al., 1998). The intervention has three targets: child and parent components for selected children, and a universal classroom component. The current study focused solely on implementer fidelity delivering the individual child component of the intervention, which teaches hierarchically-ordered skills from emotion self-monitoring to behavioral and cognitive strategies, to assist children in using these skills in real world settings at school and home. Analyses focus on proximal outcomes in the first year of the intervention during which implementers (known as Resilience “Mentors”) meet individually with children for fourteen, 25-minute weekly lessons. The parent and classroom components, which were not the foci of this study, are briefer and designed to encourage adults to support the child’s skills. (For a full description of RRP see Wyman et al., 2013).

Ten implementers (“Mentors”)¹ delivered the school-based intervention and all participated in this fidelity study. Previously they had positions as school-based paraprofessionals (e.g., classroom aide) in a variety of schools. These paraprofessionals participated in extensive training (approximately 200 hours). All implementers were employed full time with the study for at least one year, with three implementers not retained for all four years of the trial. Implementer demographics are provided in Table 1. Mentors delivered all three intervention components (child, parent, classroom) for their assigned children, with permission to videotape lessons with children. Videos were stored on a secure server.

Details about child recruitment procedures for the randomized control study are provided elsewhere (Wyman et al., 2012; Wyman, et al., 2010). Briefly, over three school years, 2213

¹We use the generic term “implementer” and the program specific term “Mentor” interchangeably.

kindergarten-3rd grade students across five schools were screened for eligibility for the Resilience Project using a modified Teacher Observation of Classroom Adaptation – Revised (TOCA; Werthamer-Larsson et al., 1991), administered to all teachers for all students. A total of 37.7% (834) of the students scored in the top tercile on the Authority Acceptance (AA) subscale that measures aggressiveness. After exclusions (e.g., sibling participating), 651 students were eligible and 403 (62% of eligible students with these high aggressive scores) were enrolled and randomized into the intervention study (individual Mentoring $n = 203$; control group $n = 200$). Fully 97% (197) of the students in the intervention (“Mentored”) group received the entire Year 1 of the intervention. The current study focuses on implementer fidelity during the first year of the program with just over a third ($n = 76$) of the Mentored children. This subset was chosen so that there would be three measures per child in our 206 coded sessions for fidelity.

Measurement development

Development of implementer fidelity measures in the one-to-one child component occurred during a pre-implementation year of the program, before enrollment in the trial began. In order to capture two critical dimensions of implementer fidelity, measures were developed for the purpose of coding implementer *adherence* to the manual and *competence* in delivering the intervention material (Schoenwald et al., 2010). We assessed fidelity in one intervention; therefore, we did not assess treatment differentiation.

Adherence—Because each lesson has content specified in the manual, unique measures of adherence were developed for each lesson. Lessons were chosen for analyses because they addressed key intervention skills and reflected phases (early, middle, late) of the first year of the intervention, as well as a broad range of anticipated difficulty of implementation. Videotapes of five lessons were used to produce fidelity ratings. Each lesson from the program manual (Cross & Wyman, 2004) follows an identical structural format, and each adherence measure was designed to assess the degree to which implementers administered each lesson as described in the manual (see Supplemental Material for the Adherence measure and coding system for Lesson 8; other adherence measures may be obtained through the authors). Lesson segments are: (a) Feelings “check-in”, (b) Review (of last lesson), (c) Introduction to the skill, (d) Teaching the skill, (e) Actively rehearsing the skill, and (f) Reviewing and generalizing the skill for future use. Given that each lesson focuses on novel content, distinctive criteria were identified for adherence to the manual when delivering each segment. One or more items are rated on each of these six segments, with ratings indicating that specific lesson material was not at all observed (0), partially observed (1), or fully observed (2). In addition to segment ratings, one point is given if implementers adhere to the designated timeframe of 25 $\pm$ 3 minutes. Scores for overall adherence for a given lesson are determined by transforming the total number of points earned into a “percent of content delivered” score. Raw scores may also be reported for the adherence measures. Given that adherence items code the presence or absence of events (i.e., causal indicators), measuring internal consistency using Cronbach’s alpha is not appropriate (Smith & McCarthy, 1995; Streiner, 2003).

Competence—Unlike implementer adherence, the criteria used to determine the degree to which lessons were implemented competently are consistently applied to every lesson. Therefore, a single measure was developed to rate implementer competence for each of the coded lessons (see Supplemental Materials for the Competence measure and scoring system). Based on our intervention theory (Wyman et al., 2010), literature review of core competencies for working with children, and our experience with paraprofessionals, seven elements of implementer competence were identified as necessary for high-quality intervention delivery: (a) Emotional responsiveness (ability to respond empathically); (b) Boundaries (ability to maintain appropriate psychological/physical boundaries that promote the child's autonomy); (c) Language (use of developmentally appropriate language to convey concepts/skills and the implementer's empathy); (d) Pacing (ability to strategically and sensitively adjust lesson pace to the needs of the child); (e) Active learning/practice (ability to effectively use interactive strategies, such as demonstrations and role-plays, to introduce, teach and reinforce concepts/skills); (f) Individualizing/tailoring (ability to flexibly tailor teaching of the concepts/skills for the child and his/her context); and, (g) Use of in vivo learning opportunities (ability to use spontaneous material such as the child's presentation to teach or reinforce a skill). Based on previous work by Forgatch and colleagues (Forgatch et al., 2005), domains of competence were rated using a 9-point scale categorized as: Good work (7–9), Adequate work (4–6), and Needs work (1–3).

Coding procedures

Five coders (two developers and three trained coders) rated entire 25-minute lessons. Trainees reviewed the program manual for familiarity with the requirements of the intervention, and initial exposure to coding rules and activities occurred in a group format where trainees coded alongside the measure developers. Group codes and the rationale for each scoring decision were discussed. Once a coder reached an acceptable level of reliability ($> .80$) with gold standard tapes, s/he was approved for independent coding and coded the same sessions as experienced coders. Coder disagreement was discussed with the coding team to reach a consensus. Codes were tracked over time and periodically analyzed for inter-rater reliability using intra-class correlations (ICC; Cicchetti, 1994). Coders were assigned to rate adherence or competence (sometimes both) for a variety of lessons.

Mentors shared cameras for recording at times due to technical or scheduling difficulties. Thus, most but not all lessons were recorded and retained. We planned to code an early, middle and late session for each randomly selected child, blocked by gender, grade, and baseline teacher rating. Videotaped lessons were coded for five of 14 intervention lessons and randomly assigned to the coding team. We identified comparable lessons delivered close in time that focused on the same skill (i.e., 5 and 6; 8 and 9) to ensure that each dyad had three equivalent observations for coding. For each child participant in the study, three lessons from the five selected for coding from the first year of the program were rated (lesson 5 or 6, 8 or 9, and 11) except three children who had only two lessons coded due to technical omissions. Coders overlapped on 50% of observations for coding. Coder group drift was assessed by re-coding 10% of the rated lessons throughout the 2 years of coding. An average of 21 sessions per implementer were coded with a range of 9 to 35 observations

($M = 20.6$; $SD = 7.5$). We coded 206 videotaped sessions reflecting a representative sample of (76 of 203) implementer-child dyads.

Coders also rated the child on one item that reflects behavior that might be challenging to even the most skillful implementer using a 3-point scale from “not behaviorally challenging” to “very behaviorally challenging.” Because only 1.3% of observed sessions rated a child as “very behaviorally challenging”, this category was combined with “somewhat” challenging into one category (high challenge). Twenty-eight percent (28%) of our sample was dichotomously coded as “high” on child behavior challenge. The majority of children demonstrated challenging behavior on only one observation and a minority ($n = 7$) demonstrated challenging behaviors on all 3 observations. Gender was not related to child challenge scores ($\chi^2(1) = 0.24$, ns). The resulting dichotomous Child Challenge score was used in the multi-level analyses of fidelity.

Data collection

Child participants—Seventy-six of the 203 mentored students (37.4%) were selected for the current study with the following blocking criteria: (a) grade, gender, and baseline TOCA scores to reflect the total child sample, and (b) equal number of dyads coded across the duration of the trial per implementer. Demographics of the sub-sample are provided in Table 1. A comparison of students in the observed group to the overall sample showed they did not differ significantly on sex, age, race/ethnicity or grade at time of enrollment.

Outcome Measures—Six outcome measures from the RCT were examined at the end of the first intervention year to assess criterion-related validity of the fidelity measures. (See Wyman et al., 2013 for details about outcomes measures). Briefly, teacher-reported child aggression was assessed with a shortened version of the Teacher Observation of Classroom Adaptation – Revised (TOCA; Werthamer-Larsson et al., 1991), including the 7-item Authority Acceptance scale (TOCA:AA). The AA measure was completed before a child’s random assignment to condition (baseline) and 6 months after the intervention was initiated to measure proximal outcomes. We found no effect for gender on AA. Child-rated symptoms were measured by individual administration of the Dominic Interactive (DI; Valla et al., 2000), a computerized self-report interview for 6 – 11 year old children that assesses symptoms of seven DSM-IV (American Psychiatric Association, 1994) diagnostic categories that resulted in two subscale totals: Internalizing and Externalizing. Parent ratings on child behavior were three scales from the Youth Outcome Questionnaire 2.0 (YOQ; Wells et al., 1996) at baseline and 6 months: Intrapersonal Distress (Internalizing), Behavioral Dysfunction, and Conduct Problems. Parent and child measures were administered before the intervention and after Year 1. All scales had good internal reliability (Wyman et al., 2013).

Statistical Analyses

We first examined descriptive summaries for the Adherence and Competence measures of implementer fidelity, and inter-rater reliability using intra-class correlation (ICC; Cicchetti, 1994). An exploratory factor analysis (EFA; Floyd & Widaman, 1995) was conducted to assess the factor structure of the Competence measure across lessons. The factor structure of

the Competence measure from EFA was validated using internal consistency coefficients. Because the items for Adherence measures are causal indicators, and not expected to be highly correlated with each other (Smith & McCarthy, 1995), internal consistency analysis was not conducted for these measures. Multi-level modeling (child, implementer) was conducted using the SAS MIXED procedure to examine how much of the variance in Competence and Adherence measures was attributed to the implementer. We used multilevel analyses to validate that our implementer fidelity measures reflect mentor's performance and not other variables such as those accounted for by fixed effects or by the child within mentor. Implementer and child within Mentor were entered as random effects in the analyses. At the same time, we tested influence of fixed effects to partial out their effects and allow us to examine only implementers' behavior with respect to fidelity measurement: lesson number, the implementer's years in the study, child gender, baseline aggression measured by the TOCA:AA, child age, and observed Child Challenge scores. We did not include a third level (observation) as no variance was available after accounting for lesson and child challenge. The same analytic strategy was applied to both implementer fidelity measures. Structural equation models (SEM), implemented in MPlus (Version 6, Muthén & Muthén, 1998–2010), were used to test criterion-related validity of the measures by creating latent variables for Adherence and Competence within student to predict 6-month outcome measures from three sources of measurement (child, parent, teacher). Three indicators consisting of Adherence or Competence scores for each mentor-child dyad at lesson 5 or 6, 8 or 9 and 11 were used to construct the latent variables. We controlled for baseline of these measures and age in the models. We used the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) to assess fit (Browne & Cudeck, 1993; Hu & Bentler, 1999; Tabachnick & Fidell, 2007) and standard Wald-type tests to measure the strengths of predictive relationships.

Results

Inter-rater reliability

Table 2 shows the ICC coefficients for inter-rater reliability on Adherence, Competence and Child Challenge within lesson among raters. Lesson 9 was used for testing new coders' reliability; thus, a higher number were coded by two or more raters. ICCs were all within the good-excellent range (.66 – .92; Cicchetti, 1994). Inter-rater reliability was stable across lessons and measures. Overall, approximately 50% of observations were coded by more than one rater. Coder drift over time was assessed by surreptitiously reassigning 10% of observations to coders over the course of the study. The agreement between original and re-assigned ratings was high (intraclass $r = .80$ to $.90$ across Adherence, Competence and lesson numbers).

Competence items across implementer were found to create one coherent factor with a strong eigenvalue of 1.577 (other eigenvalues < 1) accounting for 80% of the variance on average by lesson. Internal consistency for the Competence scale across the 5 lessons was high, with Cronbach's alphas ranging from .92 to .95. The correlation between raw Adherence and Competence scores was relatively high, $r = .68$ ($p < .001$). Although the measures of these two constructs share 46% of the variance, there remains substantial

unique variance (54%). We further explored the relationship between the two measures by examining the correlation within implementer. The correlation of Adherence and Competence ranged from a significant .747 ($p < .001$) to negative and non-significant (–.174), with six of 10 mentors having a positive and significant correlation (.458 – .747). This variability underscores the importance of examining the two constructs separately and using a multilevel approach.

Descriptive data

Descriptive statistics for Competence and Adherence by lesson are shown in Table 2. The possible range for the Competence measure is 7 to 63, based on 7 items and a 9-point Likert rating scale. Ranges for raw scores of the Adherence measures differ slightly across lessons (raw scores from 0 to 21 for lessons 5 and 9; 0 to 23 for lessons 6, 8 and 11), and were recoded for a scale of 0 to 100 to reflect percent of lesson content delivered. We did not similarly rescale the Competence scores because it is not a ratio scale and because the range was consistent across lessons. There were no outliers. For the Competence measure, observed ranges were consistent across lessons and overall fell in the “adequate work” category based on the means/standard deviations. For Adherence measures, ranges were also consistent across lessons. Lesson 8 appeared to differ from the other lessons in terms of range; however, the sample is comparatively small for this lesson and a significance test showed no difference across lessons on the measure. Adherence scores indicated that on average 64% of program content was delivered by implementers to child participants (range: 56.7% to 69.4%).

Multi-level analyses of fidelity measures

Results of the multi-level analyses examining the variance in Competence and Adherence measures attributed to the implementer, and to the child within implementer, are summarized in Table 3. These analyses were conducted to validate that our measures reflect implementer behavior and not other variables such as child baseline aggression or implementer years of experience. For Competence, 67.7% of the variance was explained by implementer, and child within implementer did not contribute a significant amount of variance (est = .12, 0.1% variance explained). These findings were unchanged with the addition of fixed effects. Child gender, age and teacher-rated aggressive behavior were not related to implementer competence. Across implementers, years of experience in delivering the intervention did not impact competence scores. Implementers tended to have lower competence on lesson 11 than on lessons 5, 6, 8, or 9, (β range 1.55 – 4.71, SE range 1.13 – 1.75, p range .008 – .35) and were more competent with less challenging students ($\beta = 3.41$, $SE = 1.03$, $CI = 1.39, 5.43$, $p = .001$).

For Adherence, 41.1% of the variance was explained by implementer, and child within implementer did not contribute a significant amount of variance (est = .78, 6.5% variance explained). These findings were unchanged with the addition of fixed effects. Adherence was not affected by child age or teacher-rated aggressive behavior. Across implementers, years of experience delivering the intervention did not impact Adherence scores. Implementers were significantly less adherent on lesson 6 than on lessons 5, 8, 9, or 11. Implementers were more adherent with girls than with boys ($\beta = 1.18$, $SE = 0.47$, $CI = 0.26$,

2.10, $p = .014$), and with less challenging students ($\beta = 1.72$, $SE = 0.47$, $CI = 0.80, 2.64$, $p < .001$). We also examined implementer by lesson interactions for both Adherence and Competence and found the interaction was not significant for either construct ($p = .13$ and $p = .65$, respectively).

Measures' criterion-related validity: Predicting proximal outcomes

We examined criterion validity using structural equation modeling (SEM) by creating latent variables for Adherence and Competence within student to predict 6-month outcome measures from three sources: child, parent, and teacher reports. Seventy-five children (99%) had Adherence and Competence score data for three lessons: lesson 5 or 6, lesson 8 or 9, and lesson 11. These scores were used in structural equation modeling as manifest indicators of the latent Adherence and Competence measures. We controlled for baseline differences as well as age in the models. Child challenge was not used in these analyses because very few children were challenging in more than one lesson. We therefore considered the Challenge score to reflect a transient state rather than an individual characteristic of the child. We did not include additional covariates because the small sample necessitated parsimonious, yet meaningful, models.² We also examined ICCs for implementer on outcomes: All were low ($< .09$) indicating that it is not the individual implementers themselves who directly affected outcomes but rather the implementers' behavior with the child. Finally, given our small sample size and the moderate correlation between Competence and Adherence, these analyses were run separately.

In all models, lesson values for Adherence and Competence significantly loaded onto their respective latent constructs, with similar loading values across models. SEM results are shown in Table 4 (values in columns three to five are regression weights and values in columns six to nine are model fit statistics). Higher Adherence and Competence both predicted reduced child-reported externalizing behaviors with the effects being strongly significant. Neither predicted child-reported internalizing behaviors. Within the parent-reported measures, higher Adherence and higher Competence predicted reduced conduct problem behaviors, and higher Adherence predicted reduced internalizing and behavioral dysfunction problems. Therefore, the actual amount of content delivered (Adherence) as well as quality with which the program was delivered (Competence) predicted child participants' response to the intervention at 6 months by both child and parent report. Neither Adherence nor Competence predicted teacher-rated externalizing behaviors (Authority Acceptance). Model fit varied; however, with the exception of parent-reported externalizing and Competence, at least one fit index was acceptable for all models with significant effects. Given the small sample size and the fact that we were using SEM for criterion-related validity and not for hypothesis testing, these lower acceptable or marginal fit statistics were expected and acceptable.

²We explored including gender in the SEM models but it did not improve the model fit or affect results, and there was no change when we removed gender from the model.

Discussion

In this study, implementer fidelity was conceptualized as adherence to the manual and competence in program delivery, and measures were developed for each. Because we had not conducted prior fidelity research on the intervention, and had empirical questions about the association between the fidelity measures and the relationship of each with child outcomes, we chose observational methods to assess implementer behaviors with child participants. Observational measures of adherence have been used frequently in the literature; competence has been examined less often (Schoenwald & Garland, 2013). Although Adherence and Competence measures were correlated in this study, the importance of creating and studying the two components of fidelity for purposes of training, implementation monitoring, and for outcome studies remains relevant. In particular, we found that although most implementers were comparable on both measures, some were quite discrepant. Thus, examination of multiple observations per Mentor revealed important variability at the individual level. Moreover, in another study using similar methods and measures, adherence and competence shared less variance (25%, Cross et al., 2014; Gould et al., 2013).

A primary finding of our study is that rigorously developed implementer fidelity measures of Adherence and Competence were associated with enhanced outcomes at the end of the first intervention year, which demonstrated the measures' criterion-related validity. As expected, higher Adherence and Competence scores were associated with lower levels of child-reported externalizing behaviors and lower parent-reported conduct problems. Higher Adherence was also associated with lower internalizing and behavioral dysfunction per parent report. That is, there were clear variations in how well implementers delivered the intervention, and higher fidelity was associated with more positive proximal outcomes reported by children and parents. Neither measure of fidelity was related to proximal child outcomes on the teacher measure of aggressive-disruptive behavior. Our primary outcome study assessing impact of the RRP intervention over three school years found intervention impact on reduced classroom behavior problems, measured by teacher report, only for older children in the sample and under conditions of high parent participation (Wyman et al., 2013). Reducing aggressive-disruptive behaviors that generalize to improved classroom behavior may require more intensive involvement of parents, which is not captured by our fidelity measures.

Our implementer fidelity measures had high inter-rater reliability and high internal consistency for the Competence measure. There were generally higher ratings and less variability in Adherence than Competence. While competence in this effectiveness trial was generally lower than fidelity reported in efficacy trials (e.g., 50–60% versus 70–80%), this may be due to a number of factors. First, fidelity in other studies focuses primarily on adherence, which was higher than competence scores. Secondly, our ordinal coding scheme may be more demanding than other rating systems. Thirdly, the implementers in this study, who were paraprofessionals, may not have delivered the intervention as well as those in some other prevention studies (e.g., teachers, professional staff).

We examined our measures to answer the question: Are we measuring implementer fidelity or are other variables contributing to the fidelity scores? The most important finding from multilevel models is that the implementer accounted for the greatest amount of variance in the scores on each of the measures. In fact, for Competence, implementer accounted for over two-thirds of the variance in scores, and, for Adherence, over forty percent of the variance was accounted for by implementer. Because child within implementer was not significant, we can conclude that fidelity scores reflect the quality of the delivery of the program content by the Resilience Mentor, not attributes of the child.

There were, however, some differences in implementer fidelity associated with lesson number, gender, and child behavioral challenges. Specifically, compared to the other four rated lessons, implementers were less competent in their delivery of a lesson (11) that involved abstract concepts and associated behavioral skills (i.e., developing self-soothing strategies to manage problems or situations that are not wholly in the child's control). Increased training and clarity in the manual may be required for interventionists to deliver the lesson in a manner that is individualized to the child without sacrificing core competencies. We also found that implementers were less adherent with a different lesson (6), which focuses on engaging the child in learning and practicing labeling feelings using a "charades" type of game. It appears that the implementers' familiarity with the general game of charades may have interfered with carrying out the activities as prescribed. In other words, implementers "stuck to what they knew" about charades and missed several important content points and activities as a result.

These findings point to the strength of multiple assessments of fidelity. Had we relied on a single lesson to assess fidelity, we would have risked making erroneous conclusions about variability in fidelity and, potentially, the association with child outcomes. Importantly, our analysis indicates the feasibility of making adjustments to which lesson is assessed, as we found no indication of important interaction effects of lesson by facilitator within implementer-child dyad. This finding also highlights how information about implementation fidelity can be used to inform and enhance training. When training new implementers, we now attend closely to these lessons since our data indicate that they are more challenging to implement fully with high quality.

Based on our multi-level analyses, we also found that implementer fidelity varied in terms of higher Adherence but not Competence with girls. That is, our implementers delivered more of the manualized content to girls but were equally competent in their delivery (i.e., emotionally responsive, used active learning strategies, tailored activities) with boys and girls. Another way to view this finding is that implementers went "off manual" more with boys than with girls and had lower Adherence. As others have noted (Aarons et al., 2012), adaptations of programs during implementation in real world settings are common and can reflect positive "edits" to the intervention. Our Adherence measure coded prescribed elements only and scores reflect only what occurred; we did not note 'non-manualized' but effective adaptations of the content. One interpretation of the finding that implementers demonstrated lower adherence with boys but with competence equal to their work with girls, is they changed some of the content to be competent as they individualized the program for boys. We plan to examine this hypothesis by rating 'off manual' content as consistent or

inconsistent with the program and may integrate positive adaptations into the intervention and training program.

In terms of examining variations on fidelity across lessons, we found that when children demonstrated behaviors that coders deemed challenging, implementers demonstrated lower levels of Adherence and Competence. It is important to note that the challenges were not systematic or related to specific children – that is, over three coded lessons only seven students (9.2 %) demonstrated challenging behaviors across each observation. Thus, it was not the individual child but specific behaviors that challenged implementer fidelity. Because our measures rely on observational data, it is possible to examine child behavior at a more microlevel to understand what those specific behaviors are and modify our training and supervision processes to increase implementers' skill in addressing these behaviors such that fidelity is not compromised. The finding that challenging behaviors in child participants negatively impacted implementer fidelity is likely to be generalizable to other similar interventions.

We recognize that measuring fidelity with observational measures in community-based, non-research settings, has numerous practical challenges (Schoenwald & Garland, 2013). Methodological rigor cannot trump relevance in real world contexts. We have begun to address the challenge by developing 'user friendly' supervision tools based on our research measures of adherence and competence. These measures, developed with extensive input by community-based supervisors, have been shown to correlate highly with fidelity measures used in research. Clinical supervisors are trained to observe and reliably rate Mentors delivering the intervention (either *in vivo* or via video recordings) to inform supervision. Reliability checks are conducted with the research team. Monitoring implementer fidelity is, therefore, integrated into the community-based program supervision model. The translation of research measures to clinical tools is one strategy that may be useful for implementation in real world community settings.

There are several limitations to this study. Although the number of individual observations is high and the number of students substantial, the number of implementers is smaller and therefore the overall N is small. To account for the nested nature of the data, we conducted multi-level modeling to uncover variations in implementer behaviors across lessons and in response to extreme challenges by child participants. Thus, our thorough examination of the measures and outcomes balances the limited number of implementers in the study. In addition, the generalizability of our findings may be limited. Ours is a school-based preventive intervention delivered by paraprofessionals; findings may be relevant to programs with similar foci, setting or implementers. Although adherence ratings of implementer fidelity are necessarily intervention-specific, the approaches presented here are widely applicable. Moreover, the competencies we identified and reliably measured may be applicable for other programs. In fact, we have received requests for our Competence measure for use in other interventions. The fit statistics for the criterion related validity structural equation models varied – some were acceptable but others were not as strong. Given the nature of our research question they are acceptable. Nevertheless, replication of these findings in a larger sample is an important next step. It is also possible that the correlation between adherence and competence in this study may be due in part to coder bias

because raters sometimes completed both measures for a single observation. Relatedly, the correlation, coupled with our sample size, did not allow us to examine criterion-related validity in models including both measures.

Finally, we measured implementer fidelity in only one component of the intervention. Wyman et al.'s (2013) study of overall program outcomes found a very strong effect for parent involvement on child outcomes. We do not have recordings of the parent component, however, and cannot measure fidelity of program delivery in these meetings. It is possible that implementer success engaging parents in this component of the intervention is related to their competencies delivering the child component (i.e., emotional responsiveness and tailoring). Wyman et al. (2013) also found that age-related factors were associated with child outcomes indicating that younger children were not cognitively ready for participation in the intervention. In these cases, implementer fidelity would not likely be a factor in outcomes. We focused our study on implementer delivery of the child component on proximal outcomes of Year 1 of the intervention only. In a subsequent study, we will extend our analyses to include fidelity ratings for Year 2 of the intervention and will examine implementer fidelity in the context of long-term outcomes for those children who received the intervention. To complete these analyses, we will include estimates of implementer fidelity for children for whom we do not have observed ratings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the National Institute of Mental Health to Dr. Cross (K23MH73615; K23MH073615-03S1) and to Dr. Wyman (R01MH068423). Dr. Cross also received support from the Implementation Research Institute (IRI) at the George Warren Brown School of Social Work, Washington University in St. Louis through an award from the National Institute of Mental Health (R25 MH080916-01A2). We also acknowledge support from the National Institute on Drug Abuse (P30 DA027828) and the National Institute of Mental Health (R01MH040859). We gratefully acknowledge the coding efforts of Emma Forbes-Jones, Ph.D. and Holly Watkins, MA. We are very appreciative of the Rochester (New York) City School District, participating children and families in the intervention, and the Resilience Mentors who implemented the intervention.

References

- American Psychological Association. Diagnostic and statistical manual of mental disorders. 5. Washington, DC: Author; 1994.
- Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*. 2010; 38:4–23. [10.1007/s10488-010-0327-7](https://doi.org/10.1007/s10488-010-0327-7) [PubMed: 21197565]
- Aarons GA, Green AE, Palinkas LA, Self-Brown S, Whitaker DJ, Lutzker JR, Silovsky JF, Hecht DB, Chaffin MJ. Dynamic adaptation process to implement an evidence-based child maltreatment intervention. *Implementation Science*. 2012; 7:32. [10.1186/1748-5908-7-32](https://doi.org/10.1186/1748-5908-7-32) [PubMed: 22512914]
- Breitenstein SM, Fogg L, Garvey C, Hill C, Resnick B, Gross D. Measuring implementation fidelity in a community-based parenting intervention. *Nursing Research*. 2010; 59:158–165. [http://dx.doi.org/10.1097/NNR.0b013e3181dbb2e2](https://doi.org/10.1097/NNR.0b013e3181dbb2e2). [PubMed: 20404777]
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. *Testing structural models*. Newbury Park, CA: Sage Publications; 1993.

- Brownson, RC.; Colditz, GA.; Proctor, EK. Dissemination and implementation research in health: Translating science to practice. New York: Oxford University Press; 2012.
- Byrnes HF, Miller BA, Aalborg AE, Plasencia AV, Keagy CD. Implementation fidelity in adolescent family-based prevention programs: Relationship to family engagement. *Health Education Research*. 2010; 25:531–541.10.1093/her/cyq006 [PubMed: 20142414]
- Carroll KM, Nich C, Sifry RL, Nuro KF, Frankforter TL, Ball SA, Fenton L, Rounsaville BJ. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*. 2000; 57:225–238. [http://dx.doi.org/10.1016/S0376-8716\(99\)00049-6](http://dx.doi.org/10.1016/S0376-8716(99)00049-6). [PubMed: 10661673]
- Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994; 6:284–290.10.1037/1040-3590.6.4.284
- Cross W, Pisani A, Schmeelk-Cone K, Xia Y, Tu X, Funkfah J, Gould. Fidelity assessment of the transfer of training in a suicide prevention program for crisis hotlines. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*. 2014 In press.
- Cross W, West J. Examining implementer fidelity: Conceptualising and measuring adherence and competence. *Journal of Children's Services*. 2011; 6:18–33.10.5042/jcs.2011.0123
- Cross, W.; Wyman, PA. Promoting resilient children initiative (PRCI): Mentor's manual. University of Rochester: Authors; 2004. Unpublished manuscript
- Curran GM, Mukherjee S, Allee E, Owen RR. A process for developing an implementation intervention: QUERI Series. *Implementation Science*. 2008; 3:17.10.1186/1748-5908-3-17 [PubMed: 18353186]
- Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation. *Implementation Science*. 2009; 4:50.10.1186/1748-5908-4-50 [PubMed: 19664226]
- Dumas JE, Lynch AM, Laughlin JE, Smith EP, Prinz RJ. Promoting intervention fidelity: conceptual issues, methods, and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventative Medicine*. 2001; 20:38–47. [http://dx.doi.org/10.1016/S0749-3797\(00\)00272-5](http://dx.doi.org/10.1016/S0749-3797(00)00272-5).
- Durlak JA, DuPre EP. Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*. 2008; 41:327–350.10.1007/s10464-008-9165-0 [PubMed: 18322790]
- Fixsen, DL.; Naoom, SF.; Blase, KA.; Freidman, RM.; Wallace, F. Implementation Research: A Synthesis of the Literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network; 2005.
- Floyd F, Widaman K. Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*. 1995; 7:286–299.10.1037/1040-3590.7.3.286
- Forgatch MS, DeGarmo DS. Sustaining fidelity following the nationwide PMTO implementation in Norway. *Prevention Science*. 2011; 12:235–246. <http://dx.doi.org/10.1007/s11121-011-0225-6>. [PubMed: 21671090]
- Forgatch MS, DeGarmo DS, Beldavs ZG. An efficacious theory-based intervention for stepfamilies. *Behavior Therapy*. 2005; 36:357–365. [http://dx.doi.org/10.1016/S0005-7894\(05\)80117-0](http://dx.doi.org/10.1016/S0005-7894(05)80117-0). [PubMed: 16718303]
- Glasgow RE, Vinson C, Chambers D, Khoury MJ, Kaplan RM, Hunter C. National Institutes of Health approaches to dissemination and implementation science: Current and future directions. *American Journal of Public Health*. 2012; 102:1274–81.10.2105/AJPH.2012.300755 [PubMed: 22594758]
- Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health*. 1999; 89:1322–1327.10.2105/AJPH.89.9.1322 [PubMed: 10474547]
- Gould MS, Cross W, Pisani AR, Munfakh JL, Kleinman M. Impact of Applied Suicide Intervention Skills Training (ASIST) on national suicide prevention Lifeline counselor interventions and suicidal caller outcomes. *Suicide and Life-Threatening Behaviors*. 2013; 43:676–91.10.1111/sltb.12049

- Hamre BK, Justice LM, Pianta RC, Kilday C, Sweeney B, Downer JT, Leach A. Implementation fidelity of MyTeachingPartner literacy and language activities: Association with preschoolers' language and literacy growth. *Early Childhood Research Quarterly*. 2010; 25:329–347. <http://dx.doi.org/10.1016/j.ecresq.2009.07.002>.
- Hepner KA, Hunter SB, Paddock SM, Zhou AJ, Watkins KE. Training addiction counselors to implement CBT for depression. *Administration and Policy in Mental Health*. 2011; 38:313–23.10.1007/s10488-011-0359-7 [PubMed: 21626444]
- Hogue A, Henderson CE, Dauber S, Barajas PC, Fried A, Liddle HA. Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting & Clinical Psychology*. 2008; 76:544–555.10.1037/0022-006X.76.4.544 [PubMed: 18665684]
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariate structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55.10.1080/10705519909540118
- Kellam SG, Ling X, Merisca R, Brown CH, Ialongo N. The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology*. 1998; 10:165–186.10.1017/S0954579498001564 [PubMed: 9635220]
- Liber JM, van Widenfelt BM, van der Leeden AJ, Goedhart AW, Utens EM, Treffers PD. The relation of severity and comorbidity to treatment outcome with Cognitive Behavioral Therapy for childhood anxiety disorders. *Journal of Abnormal Child Psychology*. 2010; 38:683–694.10.1007/s10802-010-9394-1 [PubMed: 20180011]
- Lillehoj CJ, Griffin KW, Spoth R. Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education & Behavior*. 2004; 31:242–257.10.1177/1090198103260514 [PubMed: 15090124]
- Moore, JE.; Beck, TC.; Sylvertsen, A.; Domitrovich, C. Making sense of implementation: multiple dimensions, multiple sources, multiple methods. Paper presented to the 17th Annual Meeting of the Society for Prevention Research; Washington, DC, USA. 2009.
- Muthén, LK.; Muthén, BO. Mplus Version 6. Los Angeles, CA: Muthén & Muthén; 1998–2010.
- Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, Bunger AC, Glass JE, York JL. A compilation of strategies for implementing clinical innovations in health and mental health. *Medical Care Research and Review*. 2012; 69:123–157.10.1177/1077558711430690 [PubMed: 22203646]
- Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, Griffey R, Hensley M. Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health*. 2011; 38:65–76.10.1007/s10488-010-0319-7 [PubMed: 20957426]
- SAS. SAS, Version 9.2. Cary, NC: SAS Institute Inc; 2008.
- Schoenwald SK. It's a bird, it's a plane, it's fidelity measurement in the real world. *Clinical Psychology: Science and Practice*. 2011; 18:142–147. <http://dx.doi.org/10.1111%2Fj.1468-2850.2011.01245.x>. [PubMed: 21691439]
- Schoenwald SK, Garland AF. A review of treatment adherence measurement methods. *Psychological Assessment*. 2013; 25:146–156.10.1037/a0029715 [PubMed: 22888981]
- Schoenwald SK, Garland AF, Chapman JE, Frazier SL, Sheidow AJ, Southam-Gerow MA. Toward effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*. 2011; 38:32–43.10.1007/s10488-010-0321-0 [PubMed: 20957425]
- Smith GT, McCarthy DM. Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment*. 1995; 7:300–308.10.1037/1040-3590.7.3.300
- Snyder J, Reid J, Stoolmiller M, Howe G, Brown H, Dagne G, Cross W. The role of behavior observation in measurement systems for randomized prevention trials. *Prevention Science*. 2006; 7:43–56.10.1007/s11121-005-0020-3 [PubMed: 16572301]

- Stetler CB, McQueen L, Demakis J, Mittman BS. An organizational framework and strategic implementation for systems-level change to enhance research-based practice: QUERI series. *Implementation Science*. 2008; 3:30.10.1186/1748-5908-3-30 [PubMed: 18510750]
- Streiner DL. Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment*. 2003; 80:217–222.10.1207/S15327752JPA8003_01 [PubMed: 12763696]
- Tabachnick, BG.; Fidell, LS. Using multivariate statistics. Vol. 5. Boston, MA: Pearson Education, Inc; 2007.
- Valla JP, Bergeron L, Smolla N. The Dominic-R: A pictorial interview for 6-to11-year old children. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2000; 39:85–93.10.1097/00004583-200001000-00020 [PubMed: 10638071]
- Waltz J, Addis ME, Koerner K, Jacobson NS. Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting & Clinical Psychology*. 1993; 61:620–630.10.1037/0022-006X.61.4.620 [PubMed: 8370857]
- Webb CA, DeRubeis RJ, Barber JP. Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*. 2010; 78:200–211.10.1037/a0018912 [PubMed: 20350031]
- Wells MG, Burlingame GM, Lambert MJ. Conceptualization and measurement of patient change during psychotherapy: Development of the Outcome Questionnaire and Youth Outcome Questionnaire. *Psychotherapy: Theory, Research, Practice, Training*. 1996; 33:275–283.10.1037/0033-3204.33.2.275
- Werthamer-Larsson L, Kellam S, Wheeler L. Effect of first-grade classroom environment on shy behavior, aggressive problems, and concentration problems. *American Journal of Community Psychology*. 1991; 19:585–601.10.1007/BF00937993 [PubMed: 1755437]
- Wyman, PA.; Cross, W.; Wu, P.; Schmeelk-Cone, K.; West, J.; Tu, X.; Lu, Naiji; Brown, CH. Intervention to strengthen emotion self-regulation in young urban children: Impact on internalizing and externalizing problems and mediating effects of self-regulation skills. 2013. (in review)
- Wyman PA, Cross W, Hendricks Brown C, Yu Q, Tu X, Eberly S. Intervention to strengthen emotional self-regulation in children with emerging mental health problems: Proximal impact on school behavior. *Journal of Abnormal Child Psychology*. 2010; 38:707–720. <http://dx.doi.org/10.1007/s10802-010-9398-xs>. [PubMed: 20180009]

Table 1**Implementer and Participant Demographics**

Implementers		
Sex	Male	10% (1)
	Female	90% (9)
Age (yrs)	Mean (<i>SD</i>)	42 (6.2)
	Range	27 – 49
Race/ethnicity	White	20% (2)
	African American	50% (5)
	Hispanic	30% (3)
Education	High school diploma	20% (2)
	Some college	60% (6)
	Bachelor's or Equivalent degree	20% (2)

Child Participants *		
Sex	Male	64.5% (49)
	Female	35.5% (27)
Age (yrs)	Mean (<i>SD</i>)	7.3 (1.000)
	Range	6 – 9.2
Race/ethnicity	White	6.6% (5)
	African American	60.5% (46)
	Hispanic	30.3% (23)
	Other	2.6% (2)
Grade at enrollment	1 st grade	50.0% (38)
	2 nd grade	30.3% (23)
	3 rd grade	19.7% (15)

* Note. There is no difference on any demographic variable for child participants in the study ($n = 76$) compared with the overall RCT sample ($N = 403$).

Table 2

Sample Descriptives and ICC for Adherence and Competence Measures

Lesson	n	Adherence			Competence			Child Challenge		
		M (SD)	range	% range	ICC	M (SD)	range	ICC	M (SD)	range
5 ^a	17	13.88 (3.64)	4–19	19–90%	0.92	39.35 (9.69)	21–52	0.84	1.11 (0.32)	1–2
6 ^b	57	13.04 (3.84)	5–20	21.7–87%	0.78	36.49 (10.16)	18–55	0.82	1.29 (0.46)	1–2
8 ^b	18	15.11 (2.63)	10–19	43.5–82.6%	0.68	34.67 (9.31)	20–49	0.79	1.44 (0.60)	1–3
9 ^a	60	14.58 (3.44)	5–21	23.8–100%	0.66	36.20 (10.52)	15–59	0.77	1.40 (0.52)	1–3
11 ^b	54	15.12 (3.73)	4–23	17.4–100%	0.81	33.28 (10.47)	16–51	0.76	1.33 (0.51)	1–3

Note.

^a = Adherence range = 0 – 21;

^b Adherence range = 0 – 23;

Competence range = 7–63. - Not enough separate ratings of Lesson 5 Child Challenge to calculate ICC.

Table 3

Multilevel Analyses of Measures' Psychometrics

	Competence	Adherence
Fixed part – Level 1	<i>F</i> (<i>df</i> = num, den)	<i>F</i> (<i>df</i> = num, den)
Lesson number	2.78 (4,159) *	5.04 (4,157) **
Implementer years experience	0.19 (2,70.5)	0.92 (2,74.9)
Child gender	0.05 (1,51.2)	6.42 (1,57.7) *
Authority Acceptance	0.03 (1,50.5)	0.19 (1,57.1)
Child age in months	0.85 (1,54.2)	0.17 (1,62)
Child Challenge	11.06 (1,151) **	13.21 (1,171) **
Random part – Level 2	Percent of Variance (coeff, <i>SE</i>)	Percent of Variance (coeff, <i>SE</i>)
Implementer	67.7% (71.32, 34.56)	41.1% (4.97, 2.59)
Child within implementer	0.1% (0.12, 3.08)	6.5% (0.78, 0.68)

* $p < .05$;

** $p < .01$;

Percent of Variance = ICC*100; Lesson number was dummy coded using a contrast statement.

Table 4

Criterion Related Validity Using SEM to Predict Outcomes From Fidelity Scores

Source	Outcome	Variables predicting outcome					Model fit		
		Adherence	Baseline	Age	χ^2	CFI	TLI	RMSEA	
Child	Externalizing	-.22 ^{***}	.54 ^{***}	.18 ^{***}	29.57 ^{***}	.91	.86	.10	
	Internalizing	<.001	.03+	.11 ^{***}	17.24 [*]	.97	.96	.06	
Teacher	Authority Acceptance	.12	.58 ^{***}	.01	19.69 [*]	.83	.73	.15	
Parent	Conduct	-.284 ^{***}	.71 ^{***}	-.06	18.17 [*]	.89	.82	.14	
	Internalizing	-.3.56 ^{***}	.76 ^{***}	-.08	17.56 [*]	.89	.83	.13	
	Behavior Dysfunction	-.2.88 ^{**}	.65 ^{***}	-.02	12.89	.95	.92	.08	
Child	Externalizing	-.1.49 ^{**}	.57 ^{***}	.16 ^{***}	34.70 ^{***}	.93	.89	.11	
	Internalizing	<.001	.03+	.11 ^{***}	17.24 [*]	.97	.96	.06	
Teacher	Authority Acceptance	-.01	.57 ^{***}	.01	8.68	1.00	1.00	.00	
Parent	Conduct	-.1.82 ^{**}	.67 ^{***}	-.08	23.30 ^{**}	.87	.80	.17	
	Internalizing	-.92	.76 ^{***}	-.13	19.26 [*]	.90	.85	.15	
	Behavior Dysfunction	-.73	.66 ^{***}	-.06	14.75 ⁺	.95	.93	.09	

Note. Child $N = 75$, Implementer $N = 10$; df for $\chi^2 = 9$ for all analyses;

⁺ $p < .10$;

^{*} $p < .05$;

^{**} $p < .01$;

^{***} $p < .001$.