



Published in final edited form as:

Prev Sci. 2011 September ; 12(3): 235–246. doi:10.1007/s11121-011-0225-6.

Sustaining Fidelity Following the Nationwide PMTO™ Implementation in Norway

Marion S. Forgatch^{1,2} and David S. DeGarmo¹

¹Oregon Social Learning Center

²Implementation Sciences International, Inc.

Abstract

This report describes three studies from the nationwide Norwegian implementation of Parent Management Training – Oregon Model (PMTO™), an empirically supported treatment for families of children with behavior problems (Forgatch and Patterson 2010). Separate stages of the implementation were evaluated using a fidelity measure based on direct observation of intervention sessions. Study 1 assessed growth in fidelity observed early, mid, and late in the training of a group of practitioners. We hypothesized increased fidelity and decreased variability in practice. Study 2 evaluated method fidelity over the course of three generations of practitioners trained in PMTO. Generation 1 (G1) was trained by the PMTO developer/purveyors; Generation 2 (G2) was trained by selected G1 Norwegian trainers; and Generation 3 (G3) was trained by G1 and G2 trainers. We hypothesized decrease in fidelity with each generation. Study 3 tested the predictive validity of fidelity in a cross-cultural replication, hypothesizing that higher fidelity scores would correlate with improved parenting practices observed in parent-child interactions before and after treatment. In Study 1, trainees' performance improved and became more homogeneous as predicted. In Study 2, a small decline in fidelity followed the transfer from the purveyor trainers to Norwegian trainers in G2, but G3 scores were equivalent to those attained by G1. Thus, the hypothesis was not fully supported. Finally, the FIMP validity model replicated; PMTO fidelity significantly contributed to improvements in parenting practices from pre- to post-treatment. The data indicate that PMTO was transferred successfully to Norwegian implementation with sustained fidelity and cross-cultural generalization.

Keywords

fidelity; implementation; PMTO; intervention; parenting practices

The National Institutes of Drug Abuse and Mental Health have provided substantial resources for the development and evaluation of programs to promote healthy development and to prevent and treat mental, emotional, and behavioral problems. The investment has created a golden age of programs evaluated as effective based on replicated randomized controlled trials (RCT) using Intent to Treat (ITT) analyses. The availability of such empirically supported treatments (EST) introduces a next great challenge to the field:

Correspondence should be directed to Marion S. Forgatch, ISII, 10 Shelton McMurphey Blvd., Eugene, OR 97401; marionf@oslc.org..

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The final publication is available at www.springerlink.com.

installing these programs in community settings with sustained method fidelity and positive family outcomes. The provision of an empirical base for large-scale implementations sets the stage for future studies of change in prevalence for a wide spectrum of adjustment outcomes.

Transferring ESTs from ivory-tower environs into community agencies is a formidable task given the difficulty of sustaining method fidelity in the field, and studies show that poor fidelity and poor treatment outcomes go hand in hand (see Eames 2008; Fixsen et al. 2005; Rohrbach 2006). Rigorous methods and measures are required to monitor fidelity and prevent fidelity decay (*aka* drift) during community implementations (Dumas et al. 2001; Follette and Beitz 2003; McHugh and Barlow 2010; Waller 2009). We examine three dimensions of an observation-based measure of fidelity: sensitivity to change in performance during training, capacity to detect method drift across generations of trainees, and predictive validity for the putative mechanisms of change. We evaluate these conditions by assessing community practice using the Fidelity of Implementation Rating Scale (FIMP; Knutson, et al. 2009). The implementation approach involved transferring program management from the purveyor (Fixsen et al. 2005) to community. The intervention is Parent Management Training – Oregon Model (PMTO™).

PMTO, a well-established intervention based on Social Interaction Learning theory, was first conceptualized by Gerald R. Patterson in the 1960s, with continuing refinement in the ensuing decades by colleagues at the Oregon Social Learning Center (OSLC; see Forgatch and Patterson 2010). PMTO provides preventive and clinical interventions for families of youngsters with behavioral problems in the externalizing spectrum (e.g., aggression, antisocial behavior, conduct problems, conduct disorder, oppositional defiance, delinquency, substance use). RCTs using ITT analyses have evaluated effects for specific problems in clinical samples (e.g., social aggression, stealing, delinquency, child abuse) and for selected prevention samples (e.g., divorce, remarriage, schools in high crime neighborhoods (Forgatch and Patterson 2010). The modern version of PMTO contains components that address positive and negative contingencies for behavior, emotional regulation, problem-solving skills, and academic achievement, all delivered with a strong emphasis on sophisticated teaching and process practices (e.g., Forgatch et al. 2005a, b, 2009). For decades, the OSLC group conducted program evaluations with primarily poor White samples. Since the 1990s, adapted versions of PMTO have been applied within diverse populations, and several large-scale implementations have been conducted, including a statewide program in Michigan and nationwide programs in Norway, Iceland, the Netherlands, and Denmark.

Current zeitgeist calls for assessing two dimensions of method fidelity: *adherence* to core program criteria as specified in manuals and *competent delivery* of the program (Dumas et al. 2001; Hogue et al. 2005; Perepletchikova et al. 2007; Waltz et al. 1993). Adherence quantifies coverage of specific intervention components. Many programs assess fidelity with self-reported measures completed by the practitioner or ratings provided by trained nonparticipant observers. In PMTO, certified practitioners trained to specified levels of inter-rater agreement score fidelity and never score their own sessions. The FIMP measure assesses PMTO competence and adherence in terms of five categories described in detail later: Knowledge, Structure, Teaching, Process, and Overall Development.

The Norwegian PMTO implementation began in 1999 when representatives from Norwegian ministries invited the PMTO developer/purveyor to participate in a nationwide program (Ogden et al. 2005). Implementations unfold in stages (e.g., Fixsen et al. 2007; Rogers 1995). In the first phase of the Norwegian implementation, we set goals, forged collaborative relationships, established agreements, and addressed logistical issues (Ogden

et al. 2005). The ultimate plan is provision of PMTO services to every municipality with full administration by Norwegians, a goal that required that leadership of the PMTO program be fully transferred from the program developer/purveyor to the Norwegians. The government established the National Implementation Team (NIT) to provide necessary leadership and infrastructure for the program from installation through full-scale nationwide practice. The NIT plans and oversees all PMTO activity, including training, coaching, certification, fidelity checks, and outcome evaluation. Two Norwegian ministries (Children and Family Affairs and Social and Health Affairs) initiated and funded the implementation (Ogden et al. 2005); NIDA funded the implementation study (Forgatch 2002–2007).

In Study 1, we evaluate the PMTO training program by scoring candidates for certification over the course of training using the FIMP measure and asking the question, “Does training improve performance?” The adult training research literature has identified best practices that promote high levels of program adoption and retention (e.g., reviews by Arthur et al. 2002; Fixsen et al. 2005; Salas and Cannon-Bowers 2001). Few studies, however, evaluate such skill acquisition in the field (McHugh and Barlow 2010). In the Norwegian PMTO implementation, the first group of trainees consisted of community practitioners earmarked to become frontline therapists and/or trainers for future generations of professionals. The training syllabus combined didactic instruction, role play in workshop settings, simulated practice with fictional families, practice with cases in community agencies, and extensive technical assistance (e.g., coaching based on observation of therapy sessions). Trainees also learned about PMTO efficacy and process research and the underlying theoretical foundations. The training program is described in more detail later. We hypothesize that the training program increased the mean FIMP score for the group at certification and that practitioner skill became more homogeneous as evidenced with a significant decrease in the variability of FIMP scores at certification.

Given a plan, a collaboration, an infrastructure, and a trained progenitor generation (G1), Norway was ready to extend the method to new generations. In Study 2, we evaluate the extent to which fidelity achieved by G1 was sustained following the transfer of PMTO to Norwegian administration. From the G1 certified practitioners, the Norwegian leadership selected national, regional and local leaders; trainers; coaches; and fidelity raters to roll out the program nationwide and train future PMTO professionals. The problem that commonly thwarts successful transfer from purveyor to community practice is decline in model fidelity (Fixsen et al. 2005; McHugh and Barlow 2010). Studies have identified several suspicious contributors to drift. These include: family factors (e.g., psychosocial problems, failure to complete intervention tasks, poor attendance); practitioner factors (e.g., external attributions about clients, negative emotions, abandoning protocols for competing methods when challenged, failure to push for behavioral change); organizational factors (e.g., unsupportive climate in agencies, infrastructure shortcomings, lack of resources); and insufficient training (Waller 2009). In Study 2, we hypothesized a decline in model fidelity from G1, trained by the PMTO purveyors to G3, trained by Norwegian trainers.

The third question asks whether the model validating the FIMP measure using data from an efficacy trial will replicate within a large-scale implementation. The initial FIMP validation study was based on a small sample of practitioners and families under highly controlled conditions (Forgatch et al. 2005). The test of the model yielded significant paths from fidelity observed during intervention sessions to changes in parenting practices observed in parent-child interactions before and after intervention. In this paper, we examine FIMP validity during the course of a nationwide implementation in which community practitioners provided services to families within child mental health and child welfare systems. We hypothesize that the findings from the controlled efficacy trial will replicate during the Norwegian nationwide implementation. Replication within a different country, another

language, and in differing systems of care in community agencies moves the question of modeling the measure's validity into one of model generalizability.

We present FIMP data based on three generations of Norwegian practitioners: G1 trained by the program purveyors, and G2 and G3 trained by PMTO certified Norwegians. We test three hypotheses.

Hypothesis Study 1: Training increases fidelity and decreases variability in practice.

Hypothesis Study 2: Fidelity decays from one generation of trainees to the next.

Hypothesis Study 3: High fidelity scores predict greater improvements in parenting.

Study 1: Evaluating the PMTO Training Program

The training goal was to instill a deep level of PMTO knowledge and skill acquisition that would generalize from the training environment to practice in the field. The syllabus incorporated best practices identified from training research: a) didactic presentation of relevant information; b) modeling of key components and procedures; c) practice in situations of increasing difficulty; d) coaching with regular feedback; and e) certification based on demonstration of competent application in the field. Candidates who completed the multi-stage curriculum submitted video recordings of their work with two certification families. PMTO experts reliable on the FIMP rated these sessions and certified candidates who achieved passing scores. To study the process and outcome of the G1 training, we FIMP scored videotapes of therapy sessions conducted at three points during training: early, mid training, and at certification. By certification, we expected increased FIMP scores and decreased variance.

Method

Trainees represented Norway's five health regions, which span the sparsely populated northeastern tip of the country to the more densely populated areas in the southern and western sections of the country. Agencies represented two systems of care: child mental health and child welfare. To be eligible to participate, commitments were required from the county health directors, agency leaders, and the trainees. Each agency agreed to provide resources, such as money and time to engage in the training activities. Trainees agreed to fulfill the demands of the training program specified later. The specifics of the commitments and details about the trainees are described elsewhere (Ogden et al. 2005).

Of the 35 professionals who began training, 29 completed with certification. Some candidates were selected to serve as practitioners; others were chosen as potential leaders for subsequent PMTO generations. The training began in the fall of 1999 and finished in the spring of 2001 with certification.

Procedures—Video recordings of trainees' practice were rated for three points during the training. Early samples of fidelity were based on sessions with fictional families. Midpoint data came from sessions with training families approximately halfway between the time candidates began applying PMTO with cases and certification sessions. The final assessment was based on certification families.

PMTO Training Program—The program included 21 workshop days conducted in Oslo, Norway, in six sets of 3 days spaced over 18 months. Workshops included didactic presentations supplemented with written materials (e.g., workbooks, journal articles, chapters, parent materials); modeling procedures through video and trainer role plays; and

participant role play. Simulation practice enhances skill acquisition when added to lectures and demonstration (Salas and Cannon-Bowers 2001). Participants engaged in simulation training by establishing fictional families using colleagues and video recording their role plays of sessions based on core components. Candidates received written or verbal feedback.

Trainees received coaching based on video recordings of their practice in the field, a procedure that significantly strengthens learning transfer from training to application (Fixsen et al. 2005; Salas and Cannon-Bowers 2001). Candidates treated a minimum of three cases within their agency. Sessions were translated from Norwegian into written transcripts and the trainers provided written feedback embedded within the transcripts, live group coaching, or individual telephone consultation. Advanced Norwegian candidates who received additional training also provided monthly group coaching.

When trainees achieved a certain standard of proficiency, the trainers invited them to begin two new certification cases, from which they submitted four videos on required topics, two sessions from each family. Candidates were certified when they attained a mean score of 6 (of a possible 9 points) on each session. Trainees who failed to achieve minimal scores resubmitted another session on the same topic. Certification scores included some failed sessions.

Fidelity Measure—PMTO fidelity is assessed with FIMP (Knutson et al. 2009), a rating system based on two prior OSLC systems designed to evaluate intervention process: Therapist Performance Observational System (Reid et al. 1979) and Therapy Process Code (Chamberlain et al. 1986). FIMP evaluates two aspects of practice: 1) therapists' adherence to practices and procedures spelled out in the manuals; and 2) therapists' application of clinical and teaching skills. The FIMP manual defines each category's key features, provides rating examples and guidelines, and details scoring procedures. Each category uses a 9-point scale, in which 1–3 indicates “needs work” (unacceptable performance), 4–6 is “acceptable,” and 7–9 is “good work.” The FIMP manual is available upon request.

The five FIMP categories are defined briefly next. *Knowledge*: Demonstrated understanding of PMTO content and theoretical principles. *Structure*: Ability to accomplish agenda activities and goals while addressing family issues. Includes maintaining orderly flow, leading without dominating, responsiveness to family, good transitions, and sensitive timing and pacing. *Teaching*: Proficiency in strategies that promote parents' mastery and use of PMTO practices. Verbal teach includes standard pedagogical tactics (give information, make suggestions); active teach engages families in the learning process by brainstorming, role-playing, and eliciting solutions. *Process*: Provides support that promotes a safe and supportive learning context. Includes questioning that leads to insight, maintaining balance among participants, encouraging skill development, joining family's storyline. *Overall Development*: Promotes family's growth in PMTO use. Includes likelihood that family can/will use procedures, family's apparent satisfaction, likelihood of continuing, managing unique/difficult aspects of contexts/issues.

FIMP ratings are based on time samples of therapy sessions in which two core parenting practices are delivered, skill encouragement and limit setting. Two sessions are rated for each component, one introducing the component and another troubleshooting that component. For certification purposes, full sessions are rated. For research and reliability assessment, segments of approximately 10 minutes are sampled from video-recorded family intervention sessions. To identify segments for rating, trained assistants spot-check tapes labeled with topics, seeking segments of approximately 10 minutes with content on a relevant component (i.e., skill encouragement or limit setting) and teaching activity (e.g.,

debriefing home practice, role playing, brainstorming for incentives or negative consequences).

FIMP raters are required to be certified PMTO practitioners and are familiar with PMTO manuals and practices. During training, coders learn the coding manual, view and score video recordings, discuss agreements/disagreements with the trainer, and take the reliability test. In the present study, the FIMP raters were the PMTO trainers; sessions were randomly assigned to raters without regard to time point in training (i.e., early, mid, late). A principal components factor analysis of the five items obtained a single factor solution at early, mid, and certification time points explaining 93%, 93%, and 91% of the variance, respectively across time. Eigen values were 4.67, 4.66, and 4.58, respectively with Cronbach's alphas of .98, .98, and .97. The final therapist fidelity score was a mean of the 5-item scale. Blind interrater agreement was assessed for 18 calibrator and reliability pairs, who were required to achieve 70% agreement for each item; 91% were in agreement within 1 unit on the 9-point scale for all 5 items.

Study 1 Results

We hypothesized significant improvements in therapist fidelity from early training to certification with reduced variation in scores by endpoint. Means and standard deviations for the 29 completers are presented in Table 1 for early training (T1), mid (T2), and certification (T3). Data include scores for each fidelity dimension and the FIMP scale score averaging fidelity across all categories. We employed general linear modeling (GLM) repeated measures to test for improvements in fidelity evidenced by a significant within-subjects linear time factor or a nonlinear positive quadratic. The F tests for time shown in Table 1 supported the key hypothesis. The mean FIMP scale score showed significant linear change ($F(1,28) = 6.71, p < .05$) and significant quadratic change ($F(1,28) = 8.88, p < .01$) indicating a greater improvement in fidelity at certification relative to early training. In fact, each separate FIMP dimension showed significant positive quadratic time effects by certification relative to early training scores.

To examine variance components, we plotted the distributions and specified *post-hoc* tests for homogeneity of variance using structural equation modeling (SEM). The SEM was a simple three variable covariance model specifying equality constraints for the variance components. Nested model comparisons indicated that the variances were equal when comparing all time periods, except for the change in variance from T2 to T3 fidelity scores. The nested model change in chi-square test indicated a significant improvement in model fit ($\Delta\chi^2(1) = 5.41, p < .05$).

Study 1 Discussion

The data indicate that the training program produced practitioners demonstrating competent adherence to the PMTO method. The findings support both hypotheses: 1) trainees demonstrated significant increases in PMTO performance; and 2) at certification, trainees' scores were more homogeneous. At first, the rather high scores achieved by most trainees at T1 surprised us. In retrospect, we think their performance reflects the fact that the sessions were role plays conducted with fictional families rather than actual treatment sessions. This simulation task was designed to promote success by reducing the demand characteristics of the job; the data suggest that the task functioned as planned. One way to test the value of the task would be to randomly assign trainees to engage or not in the simulated practice and compare their progress with a group using treatment families only.

One may well ask why PMTO requires such an intensive training program. Candidates were groomed not only to be therapists, but leaders for future generations. As leaders, they would

be expected to train, coach, evaluate method fidelity, and provide leadership within the nationwide implementation infrastructure. This required them to be well steeped in the theory and research and skillful in its application with families struggling with multiple problems and a wide range of adversities.

This pilot study investigating growth during a long and complex training program has a number of limitations. First, the FIMP raters were not fully blind to the trainees' stage in training (i.e., early, mid, late training). Furthermore, the raters were the trainers. These factors may have contributed to bias to see improvement when there was none. A better-designed study would have fully blind raters who were unfamiliar with the trainees. A second problem with this study was the failure to assess factors that may have contributed to the varying trajectories in growth during training. Measures of training quality, trainee characteristics, and agency support should be included in future studies.

Study 2: Evaluating Drift across Generations

Sustaining model fidelity is a major challenge when transferring an EST from purveyors to communities (Fixsen et al. 2005; McHugh and Barlow 2010). Using separate cohorts of trainees, we evaluated PMTO fidelity across three generations of community practitioners in Norway. The challenge in transfer of a complex intervention is similar to the problem in relay races in which the risk for failure is greatest when passing the baton from one runner to the next. In the transfer of PMTO from purveyors to the Norwegian team, we were concerned that the baton would drop in the transition of training by purveyors to training led by their trainees. If that happened, we presumed that the next generation would suffer as well. Other issues generated risk for drift. Would PMTO, a program developed and tested in one culture in highly controlled settings, survive in another culture in diverse agencies spread throughout an entire nation? Would the training program provided for G1 prove sufficient for maintaining fidelity? Could the relatively newly established NIT team handle the many challenges to the infrastructure? Would Norwegian families respond favorably and benefit from the program? Could the NIT provide the structure, monitoring, and discipline necessary to sustain fidelity? Would adaptations incongruent with PMTO principles be required? We hypothesized these issues would lead to drift across generations.

Method

Fidelity data were available for the 29 G1s, 51 G2s, and 51 G3s. Fidelity was scored using the same procedures and measures described in Study 1. The team of FIMP raters for this study included the purveyor team who scored G1 and FIMP reliable certified Norwegian PMTO practitioners who scored G2 and G3. Coders never rated their own sessions. Training required approximately 40 hours before achieving reliability. Since the first FIMP training, the process has become more efficient, with current training requiring approximately 25 hours in Norway, Iceland, the Netherlands, and Michigan. Inter-rater reliability scores were computed from randomly selected FIMP sessions for at least 28% of the cases in G2 and G3. Intra-class correlation coefficients (ICC) for inter-rater agreement were .76 and .82 respectively for G2 and G3. Cronbach's alpha was .98 and .97, respectively.

Study 2 Results

We hypothesized decay in PMTO fidelity with each successive generation using GLM analysis of variance (ANOVA). Means, standard deviations, and F tests are shown in Table 2. Fidelity scores differed by generation, however, not as expected. *Post-hoc* Bonferroni contrasts indicated that although G2 therapists scored significantly lower compared to G1, G1 and G3 scores did not differ. Thus, despite our hypothesis, decay in competent adherence was not incremental over the course of three generations. Rather, fidelity temporarily

declined from G1 to G2 when the leadership changed hands from the purveyor to G1 trainers. However, at G3, the fidelity scores improved and did not differ significantly from scores achieved by candidates in G1.

Study 2 Discussion

The concern about dropping the baton between runners proved true, but only in the transition from purveyor to the Norwegian community. Once the new leaders had time to ripen, their trainees recovered the high level of competent adherence to the PMTO program that had been attained by G1. The transfer required considerable work by the Norwegian implementation team. The Norwegian trainers had to translate the purveyors' materials, make language adjustments, adapt parent materials to fit cultural metaphors and perspectives, develop training programs, train trainers, develop communication systems, and strengthen and expand the infrastructure to monitor all stages. Future studies will want to evaluate the quality of training provided to each generation of trainees and assess important characteristics such as agency and practitioner characteristics. That the recovery in fidelity from G2 to G3 was so rapid is quite remarkable. An important question, of course, is whether this pattern of sustained fidelity from G1 to G3 can be replicated.

Study 3: Predictive Validity of FIMP

The theoretical model underlying PMTO specifies parents as the most proximal and enduring agents of change for their children. PMTO practitioners coach parents to remediate their children's behavior problems and promote prosocial behaviors. The hypothesis is that intervention effects on parenting practices mediate intervention effects on the child, a finding supported by in several experimental tests (Beauchaine et al. 2005; DeGarmo et al. 2004; Eddy and Chamberlain 2000; Ogden and Amlund-Hagen 2008; Tremblay et al. 1995). Thus, a test of predictive validity would require FIMP scores to predict improvements to parenting practices. The original FIMP test was conducted within a RCT in Oregon with 110 families of boys and girls at risk for conduct problems. The ITT analysis revealed large effect sizes for improved parenting practices in the PMTO condition. Improvements in parenting in turn were significantly associated with reductions in child noncompliance, externalizing and internalizing behavior problems (DeGarmo and Forgatch 2007; Forgatch et al. 2005), and improvements in the marital relationship (Bullard et al. 2010).

In the Oregon FIMP validity study, 20 PMTO-treated families were randomly selected from the caseload of the four PMTO practitioners and segments of their sessions were scored with the FIMP measure. Structural equation path models specified fidelity as a predictor of 12-month pre-post change in observed parenting practices for mothers and stepfathers. Higher PMTO fidelity was associated with significant increases in mothers' and stepfathers' effective parenting controlling for initial levels. Fidelity accounted for 30% of the variance in change in mother and stepfather parenting (Forgatch et al. 2005b).

Evaluating effects of fidelity within an RCT with a small subsample was a good first step in establishing the instrument's internal and predictive validity. To validate the measure for use in implementation studies, findings must generalize to community settings and be tested with varying cultures with sizeable samples of practitioners and families. We hypothesized that FIMP predictive validity established in an efficacy trial would generalize to the nationwide sample in Norway.

Method

We recruited participants throughout Norway. In eligible participants, families received PMTO therapy, engaged in family interaction task (FIT) assessments, and sessions for core PMTO components were FIMP rated. This data pool included certification therapy sessions

from trainees in G1, G2, and G3 and therapy sessions from practicing certified PMTO clinicians. There were 242 families meeting these eligibility requirements at baseline, of which 237 were mothers and 183 were fathers. Children's mean age was 8.13 ($SD = 2.24$) and 72% were boys. Parents' mean age was 38.07 ($SD = 6.59$). The mean gross annual income for participants was 436,843.42 Norwegian Kroner ($SD = 224,518.34$), which is approximately \$84,000. Nearly all parents were White ethnic Norwegians (96.8%) with 1.1% Danish, 1.6% White other, and .5% Asian or Pacific Islander. This ethnic homogeneity reflects the general makeup of families in Norway at the time of the study. All families and therapists completed informed consent procedures in accordance with IRB procedures approved in both Norway and at OSLC. The present analysis includes the 110 therapists treating the 242 families. Therapists treated 2.80 families on average ($SD = 1.31$) ranging from 1 to 6 families. Therefore, the present data represent a multi-level nested structure due to clustering of families within therapists. To address nonindependence, we conducted multi-level modeling described below in the analysis plan.

Procedures—Direct observation in separate settings was used to assess method fidelity and parenting practices. Fidelity was scored from videos of PMTO therapy sessions. Parenting practices were scored from videos of FIT assessments with parents and children at baseline (BL) and 9 months later (post intervention). Direct observation methods were new to clinical research in Norway requiring time to develop infrastructure, methods, and train coders to reliability in each observational system. FIMP raters were certified PMTO therapists. Coders of parent-child interactions were students at the University of Oslo or paid employees. Families received small sums for participation in each assessment. Procedures to score fidelity were the same in original Oregon validity study and the present Norwegian sample. Procedures to assess the FIT in Norway were streamlined.

The assessment procedures for the FIT have been developed and validated in passive and experimental longitudinal studies and found to have convergent, discriminant and predictive validity and sensitivity to change (e.g., Forgatch and DeGarmo 1999; Patterson et al. 1992; Reid et al. 2002). The FIT totaled 48 minutes in the original FIMP study and included problem-solving discussions, a teaching task, a cooperation/play task, and a refreshment period (Forgatch et al. 2005). The FIT was tailored for Norwegian implementation with tasks adjusted for two age ranges of children: over 8 (30 minutes) and under 8 (25 minutes). Both groups included a problem solving task and a task evaluating family cooperation during the FIT. The under-8 group also included free-play, clean-up, and waiting tasks. The over-8 group had an additional problem-solving task and planned a fun family activity.

The FIT was scored with Family and Peer Process Code (FPPC: Stubbs et al. 1998) and Coder Impressions of Lab Tasks (Forgatch et al. 1992). Coder training requires approximately 16 to 20 weeks at 20 hours/week to achieve reliability at OSLC. In Norway, coder training initially required 40 weeks, 20 weeks to achieve reliability and 20 additional weeks for stability. With experience in use of observational procedures, the training time has reduced to approximately 20 weeks at 15–20 hours/week. To achieve reliability, coders must score two tapes in a row with 75% event-by-event agreement and a Cohen's Kappa coefficient of .65. Reliability checks take place biweekly. Coders are blind to families' group assignment and pre- or post condition and never score the same family for both assessments. Twenty percent of the assessments were randomly selected for blind reliability checks. The average Kappa coefficient for separate dimensions was .67 and the inter-rater agreement ICC was .78.

Observed Parenting Practices—The construct for parenting practices included three scales: skill encouragement, monitoring, and inept discipline. Trained coders provided Likert-type ratings after viewing the interaction tasks. Parenting construct scores were

created using procedures outlined by Stoolmiller and colleagues employed in the original fidelity test. These procedures provide a method for combining indicators of differing response scales that retain the mean information for assessing change (Stoolmiller and Bank 1995; Stoolmiller et al. 1993). In the present report, all indicators were rescaled to a common continuous ratio level metric of 0 to 1 at the item level before combining the composite score to assess change in parenting.

Inept discipline was a 13-item scale score rated on a 5-point scale. Items included overly strict, authoritarian, oppressive, inconsistent, erratic, used nagging. Cronbach's alphas (α) for mothers and fathers respectively were .82 and .79 at BL and .82 and .83 at post-intervention.

Skill encouragement was an 11-item scale. Sample items included breaks task into manageable steps, reinforces success, prompts, provides reinforcement for correct responses ($\alpha = .87$ and $.89$ for mothers and fathers respectively at BL and $.88$ and $.90$ at post-intervention).

Monitoring was rated by two reporters. For under 8, parent interviewers rated three Likert-scale items: supervision during assessment, tracking outside lab, skillful at obtaining information. Coders rated two items: apparent knowledge of child's activities, tolerance of negative behavior. For over-8, interviewers also rated two items involving activities away from home. Items were rescaled across age groups before averaging (under-8: $\alpha = .74$ and $.74$ for mothers and fathers respectively at BL and $.72$ and $.69$ at 9 months; over-8: $\alpha = .73$ and $.80$ at BL for mothers and fathers and $.77$ and $.72$ at 9 months).

Analysis Plan—The predictive validity hypothesis was tested with structural equation modeling (SEM) to estimate effects of FIMP ratings on pre-post change in parenting controlling for children's gender and age. Maximum likelihood SEM parameters were estimated with MPlus6.0 (Muthén and Muthén 2010). Following recommendations for missingness, data were modeled using full-information maximum likelihood (FIML), which uses all available information to handle missing data. FIML estimates are computed by maximizing the likelihood of a missing value based on observed values in the data (Jeličić et al. 2009). Compared to mean-imputation, list-wise, or pair-wise models, FIML provides more statistically reliable standard errors. Individuals who have baseline data and no follow up data contribute nothing to the likelihood of estimates and are effectively excluded from change analyses (Brown et al. 2008). Although each family was uniquely scored for therapist fidelity, we also employed multilevel SEM to address nonindependence and potential bias due to clustering of families within therapists. Although the amount of clustering within the sample is relatively small, violation of non-independent observations may lead to increased Type I error (Clarke 2008). Therefore we examined findings with estimates robust to nonnormal standard errors and with estimates adjusting for clustering.

Study 3 Results

We first examined attrition. For mothers, 148 of 237 assessed at baseline participated in the follow-up assessment (62%); for fathers the equivalent numbers were 84 of 183 (46%). Attrition analyses indicated that mothers with post assessment parenting scores did not differ at baseline compared to those mothers lost to follow up. Fathers lost to follow up, however, scored lower on baseline parenting compared to those remaining in the study (completers $M = .65$, $SD = .12$, and attriters $M = .58$, $SD = .17$, respectively, $t = 3.28$, $p < .01$). Clinical studies have shown that less skilled and antisocial parents are at risk for drop out and fathers are at greater risk for lack of engagement and drop out relative to mothers (Bagner & Eyberg, 2003). This suggests that the contribution of change variance by fathers may be underestimated. Baseline fathers' parenting is covaried.

We next specified an SEM path model with three latent variables: intervention fidelity measured by ratings of skill encouragement and limit setting sessions; pre-treatment parenting measured by mother and father parenting construct scores; and a post-treatment factor measured by change in mothers' and fathers' parenting. Thus, analogous to a multi-wave growth model, effects represented prediction of change controlling for initial status. Latent variables obtained reliable factor loadings and the model obtained excellent fit to the data [$\chi^2(4) = .63, p = .96; CFI = 1.00; RMSEA = .00$]. Results shown in Figure 1 supported the hypothesis. Higher levels of fidelity predicted increases in effective parenting. Paths are standardized coefficients with estimates for the multilevel model in parentheses. Table 3 presents means, standard deviations, and bivariate correlations for the study variables along with the intra-therapist intraclass correlation coefficients (ICC). Age and child gender were not significant covariates.

Study 3 Discussion—Results replicated the earlier findings; the larger dataset provided from the nationwide Norwegian implementation obtained similar substantive findings to the original test of fidelity in the smaller RCT. Replication of the model is generalized within a cross-cultural context. Although some details of procedures in the replication study differed from the original, the general tenor of the findings support the consistency and generalizability of effects. The replicated model specifies how change in parenting comes about during therapy. Both studies speak to the fact the therapist's competent adherence in their application of the PMTO model predicted the degree of change in parenting. It is noteworthy that the findings were based on two observational systems scored from two settings (therapy sessions and parent/child interactions) in two systems of care (child mental health and child welfare) with sampling at three points in time (before, during and after therapy).

General Discussion

This nationwide implementation began as a top-down approach and became more bottom up as Norwegians trained Norwegians in subsequent generations (Fixsen et al. 2005). In Study 1, we examined growth in fidelity during the G1 training by the PMTO purveyors. As hypothesized, trainees' performance improved and became more homogeneous. In Study 2, we found a small but significant decline in fidelity following program transfer from purveyor to community. By G3, however, fidelity scores were equivalent to those attained by G1, indicating recovery. Finally, we tested a replication model for FIMP validity. The model specified that competent adherence to the PMTO method observed during delivery of the intervention predicted improvements in parenting practices from pre to post treatment. The model replicated, providing strong support for cross-cultural generalization.

The Norwegian research team provided icing on the cake for this implementation by conducting their own PMTO effectiveness trial in Norway (Ogden and Amlund-Hagen 2008). The RCT compared PMTO with community treatment as usual in child mental health and child welfare agencies nationwide. Participants were 112 families referred for conduct problems for boys and girls. The RCT replicated and extended findings reported in other PMTO studies (Forgatch and Patterson 2010). Based on multi-method and -agent assessment and ITT analysis, PMTO enhanced parental discipline, reduced children's externalizing problems, and improved compliance and child social competence. High scores on FIMP ratings for families in the PMTO condition were associated with benefits to the parenting practices of effective discipline and positive involvement and higher ratings of parents' satisfaction with treatment.

The implementation goal in Norway has been to maintain basic PMTO principles with minimal adaptation. When cultural concerns emerged, teams were established to address

them. One group focused on language issues, adjusting words and ideas to fit Norwegian perspectives. For example, the construct of punishment/discipline has a history in Norway that makes professionals and parents alike uncomfortable. The language team decided to use 'negative consequences' to label practices involving discipline. Another phrase, 'time out,' was considered too American. One suggestion was a phrase that translated into 'thinking time,' a principle incongruent with the PMTO principle of simple disengagement from escalating conflicts. The term that proved acceptable to all was 'break time.' Call the practices discipline or negative consequences, time out or break time, the underlying principles remain the same— parents use contingent short negative sanctions for specific misbehaviors. Another arena for cultural adaptation involved the appearance of the materials. Some of the training candidates with skills in graphic design created materials that were pleasing to Norwegian families. Thus, the adaptations were primarily topographical, culturally relevant, and carefully negotiated between purveyor and adopters.

Rogers (1995) discusses the general debate between reinvention on the one hand and fidelity on the other. One assumes that this general tension pervades most efforts to disseminate innovations such as PMTO. The details of what adjustments may be required are likely to vary from one context to another. For example, accommodations made in the JOBS program for unemployed workers in Finland or China (Price 2000) may require different adaptations from those confronting a mental health prevention program tailored for Latinos (Bernal et al. 1995; Domenech Rodríguez et al., 2011). Furthermore, some adaptations can prove to be perilous undertakings. For example, as shown in a careful review of efforts to adapt Olweus' prestigious BULLYING program to three different cultures, the data showed that the outcomes were noticeably less in all three cases than findings for the original model (Stevens et al. 2001).

The limitations inherent in this study leave considerable room for future work in the field of implementation science. Regarding the G1 training in Study 1, the sample of therapists was small and without assessment of potentially important covariates that may explain individual differences in training trajectories. With replication in larger samples, factors at the therapist level, characteristics of clients, organizational structures, and support for training and supervision can be considered (Glisson 2007). These same factors may also explain variation in the generational findings reported in Study 2. Finally, in Study 3, the predictive validity of the FIMP model represents a single replication. Will implementations in Iceland, the Netherlands, and Michigan produce similar models? A study in Iceland yields promise for prevalence effects showing reduction in referrals to specialist services for behavior problems (Björnsdóttir and Sigmarsdóttir 2009). Will other PMTO implementation sites yield clinically significant prevalence reductions for societal problems like drug abuse, crime, and school failure?

The transfer of empirically supported treatments from developer to community delivery systems is a dynamic process requiring extensive collaboration and long-term commitment (Herschell et al. 2004). Although providing families with effective programs has become a priority, many programs are installed without evaluation during or following training. Furthermore, method fidelity, if attained in the first place, quickly decays. We lack clarity about factors that produce effective implementations and how we can achieve them. When implementations fail, we cannot say why. Our definitions of effective training are unspecified and therefore not measured. When drift occurs, we do not know if decay is the result of poor training, lack of theory, unregulated adaptations, failure to monitor, or some combination of these. What kind of infrastructure is required to sustain fidelity? How are practices and outcomes evaluated and do these evaluations rest on RCTs with ITT analysis? Without rigorous methodology, we can expect failures of performance for ESTs in the field and will not understand why the failures occur or how to prevent them.

Community agencies are now required to provide families with ESTs. Agencies send their clinicians from one workshop to another to enhance their repertoires. Little monitoring is provided concerning later usage of programs. Adaptations take place with neither rhyme nor reason. Soon the principles that produced positive outcomes in the first place are lost. Until EST training programs provide fidelity data for their trainees and systematically monitor fidelity during follow-up practice in the field, the families served may not actually be receiving evidence-based practice. To retain effectiveness in the field, high standards of monitoring and evaluating fidelity and treatment outcomes must be upheld.

The studies summarized in this report contribute to the developing science of implementation. Our data indicate that an EST can be transferred from developer to an adopting community with sustained fidelity when certain conditions are met. In the PMTO implementation, a strong collaboration was established between the purveyor and community and an effective infrastructure was forged. The Norwegian leadership made a long-term commitment to provide sufficient resources to conduct the many procedures required to sustain fidelity. They conducted a nationwide RCT demonstrating that the expected outcomes were achieved when PMTO was applied in community agencies throughout the country. The Norwegian NIT team regularly monitors fidelity through certification and recertification procedures using FIMP, conducts regular FIMP reliability checks with the PMTO purveyor team, and carefully continues to evaluate adaptations and continued effectiveness by conducting RCTs.

We presented an observation-based measure of intervention fidelity and a set of methods that can increase rigor in the implementation field for mental health, drug abuse, and child welfare issues. We described our implementation approach and evaluated fidelity during and following the transfer of PMTO into an adoptive community that encompassed an entire nation in two systems of care. This focus enabled us to evaluate implementation through the lens of performance of those trained in the method by the program developers and trained leaders from the adopting community. The program is arduous, a strong infrastructure is required, and constant coaching and monitoring are essential. Replication studies are currently underway in Iceland and the Netherlands. Will the findings prove to be robust?

Acknowledgments

The project described was supported by Award Number R01DA 16097 from the Prevention Research Branch, NIDA, U.S. PHS; and Award Number P30 DA023920 from the Division of Epidemiology, Services and Prevention Research, Prevention Research Branch, NIDA, U.S. PHS. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIDA.

References

- Arthur MW, Hawkins JD, Pollard JA, Catalano RF, Baglioni AJ Jr. Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors: The Communities That Care Youth Survey. *Evaluation Review*. 2002; 26:575–601. [PubMed: 12465571]
- Bagner DM, Eyberg SM. Father involvement in parent training: When does it matter? *Journal of Clinical Child and Adolescent Psychology*. 2003; 32:599–605. [PubMed: 14710469]
- Beauchaine TP, Webster-Stratton C, Reid MJ. Mediators, moderators, and predictors of 1-year outcomes among children treated for early-onset conduct problems: A latent growth curve analysis. *Journal of Consulting and Clinical Psychology*. 2005; 73:371–388. [PubMed: 15982136]
- Bernal G, Bonilla J, Bellido J. Ecological validity and cultural sensitivity for outcome research: Issues for the cultural adaptation and development of psychosocial treatments with Hispanics. *Journal of Abnormal Child Psychology*. 1995; 23:67–82. [PubMed: 7759675]

- Björnsdóttir A, Simarsdóttir M. Parent Management Training – The Oregon Model (PMTO™): Effect of a prevention and treatment program for behavioral problems among kindergarten and elementary school children in Hafnarfjörður. *Icelandic Journal of Education*. 2009; 18:9–28.
- Brown HC, Wang W, Kellam SG, Muthén BO, Petras H, Toyinbo P, et al. Methods for testing theory and evaluating impact in randomized field trials: Intent-to-treat analyses for integrating the perspectives of person, place, and time. *Drug and Alcohol Dependence*. 2008; 95:S74–S104. [PubMed: 18215473]
- Bullard L, Wachlarowicz M, DeLeeuw J, Snyder J, Low S, Forgatch MS, et al. Effects of the Oregon Model of Parent Management Training (PMTO) on marital adjustment in new stepfamilies: A randomized trial. *Journal of Family Psychology*. 2010; 24:485–496. [PubMed: 20731495]
- Chamberlain, P.; Davis, B.; Forgatch, MS.; Frey, S.; Patterson, GR.; Ray, JR.; Rothschild, R.; Trombly, T. *The Therapy Process Code: An observational system*. Oregon Social Learning Center; Eugene, OR: 1986.
- Clarke P. When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. *Journal of Epidemiological Community Health*. 2008; 62:752–758.
- DeGarmo DS, Forgatch MS. Efficacy of parent training for stepfathers: From playful spectator and polite stranger to effective stepfathering. *Parenting: Science and Practice*. 2007; 7:1–25.
- DeGarmo DS, Patterson GR, Forgatch MS. How do outcomes in a specified parent training intervention maintain or wane over time? *Prevention Science*. 2004; 5:73–89. [PubMed: 15134313]
- Domenech Rodríguez MM, Baumann AA, Schwartz AL. Cultural adaptation of an evidence based intervention: From theory to practice in a Latino/a community context. *American Journal of Community Psychology*. 2011; 47:170–186. [PubMed: 21116707]
- 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 Preventive Medicine*. 2001; 20:38–47. [PubMed: 11146259]
- Eames C. The Leader Observation Tool: A process skills treatment fidelity measure for the Incredible Years parenting programme. *Child Care, Health and Development*. 2008; 34:391–400.
- Eddy JM, Chamberlain P. Family management and deviant peer association as mediators of the impact of treatment condition on youth antisocial behavior. *Journal of Consulting and Clinical Psychology*. 2000; 68:857–863. [PubMed: 11068971]
- Fixsen, DL.; Naoom, SF.; Blase, KA.; Friedman, RM.; Wallace, F. *Implementation research: A synthesis of the literature*. University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network; Tampa, FL: 2005.
- Fixsen DL, Naoom SF, Blase KA, Wallace F. Implementation: The missing link between research and practice. *The APSAC Advisor*, Winter/Spring. 2007:4–11.
- Follette WC, Beitz K. Adding a more rigorous scientific agenda to the empirically supported treatment movement. *Behavior Modification*. 2003; 27:369–386. [PubMed: 12841589]
- Forgatch, MS.; Principal Investigator. *Implementing parent management training in Norway* (Grant No. R0116097). NIDA NNPRI: Community Multi-Site Prevention Trials (CMPT) Oregon Social Learning Center; Eugene: 2002–2007.
- Forgatch MS, DeGarmo DS. Parenting through change: An effective prevention program for single mothers. *Journal of Consulting and Clinical Psychology*. 1999; 67:711–724. [PubMed: 10535238]
- Forgatch MS, DeGarmo DS, Beldavs Z. An efficacious theory-based intervention for stepfamilies. *Behavior Therapy*. 2005; 36:357–365. [PubMed: 16718303]
- Forgatch, MS.; Knutson, NM.; Mayne, T. *Coder impressions of ODS lab tasks*. Oregon Social Learning Center; Eugene: 1992.
- Forgatch, MS.; Patterson, GR. Parent Management Training – Oregon Model: An intervention for antisocial behavior in children and adolescents. In: Weisz, JR.; Kazdin, AE., editors. *Evidence-based psychotherapies for children and adolescents*. 2nd ed.. Guilford; New York: 2010. p. 159-178.
- Forgatch MS, Patterson GR, DeGarmo DS. Evaluating fidelity: Predictive validity for a measure of competent adherence to the Oregon model of parent management training (PMTO). *Behavior Therapy*. 2005a; 36:3–13. [PubMed: 16718302]

- Forgatch MS, Patterson GR, DeGarmo DS, Beldavs ZG. Testing the Oregon delinquency model with nine-year follow-up of the Oregon Divorce Study. *Development and Psychopathology*. 2009b; 21:637–660. [PubMed: 19338702]
- Glisson C. Assessing and changing organizational culture and climate for effective services. *Research on Social Work Practice*. 2007; 17:736–747.
- Herschell AD, McNeil CB, McNeil DW. Clinical child psychology's progress in disseminating empirically supported treatments. *Clinical Psychology: Science and Practice*. 2004; 11:267–288.
- Hogue A, Liddle HA, Singer A, Leckrone J. Intervention fidelity in family-based prevention counseling for adolescent problem behaviors. *Journal of Community Psychology*. 2005; 33:191–211.
- Jeličić H, Phelps E, Lerner RM. Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychobiology*. 2009; 45:1195–1199.
- Knutson, NM.; Forgatch, MS.; Rains, LA.; Sigmarsdóttir, M. Fidelity of Implementation Rating System (FIMP): The manual for PMTO™. Revised ed.. Implementation Sciences International, Inc.; Eugene, OR: 2009.
- McHugh RK, Barlow DH. The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist*. 2010; 65:73–84. [PubMed: 20141263]
- Muthén, LK.; Muthén, BO. MPlus: Statistical analysis with latent variables user's guide. Sixth ed.. StatModel; Los Angeles, CA: 2010.
- Ogden T, Amlund-Hagen K. Treatment effectiveness of Parent Management Training in Norway: A randomized controlled trial of children with conduct problems. *Journal of Consulting and Clinical Psychology*. 2008; 76:607–621. [PubMed: 18665689]
- Ogden T, Forgatch MS, Askeland E, Patterson GR, Bullock BM. Implementation of parent management training at the national level: The case of Norway. *Journal of Social Work Practice*. 2005; 19:317–329.
- Patterson, GR.; Reid, JB.; Dishion, TJ. Antisocial boys. Vol. Vol. 4. Castalia; Eugene, OR: 1992.
- Perepletchikova F, Treat TA, Kazdin AE. Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*. 2007; 75:829–841. [PubMed: 18085901]
- Price, RH. Mobilization, reinvention, and scaling up: Three core processes in knowledge exchange, adaptation, and implementation. Inaugural World Conference, “The Promotion of Mental Health and Prevention of Mental and Behavioral Disorders”; Atlanta, GA. 2000, December;
- Reid, JB.; Fleischman, MJ.; Arthur, J.; Toobert, DJ.; Stern, S.; Patterson, GR. Therapist performance observational system. Association for the Advancement of Behavior Therapy; San Francisco. 1979, December;
- Reid, JB.; Patterson, GR.; Snyder, J. Antisocial behavior in children and adolescents: A developmental analysis and model for intervention. American Psychological Association; Washington, DC: 2002.
- Rogers, EM. Diffusion of innovations. Fourth ed.. Free Press; New York: 1995.
- Rohrbach L. Type II translation: Transporting prevention interventions from research to real-world settings. *Evaluation & the Health Professions*. 2006; 29:302–333.
- Salas E, Cannon-Bowers JA. The science of training: A decade of progress. *Annual Review of Psychology*. 2001; 52:471–499.
- Stevens V, Bourdeaudhuij ID, Oost PV. Anti-bullying interventions at school: Aspects of programme adaptation and critical issues for further programme development. *Health Promotion International*. 2001; 16:155–167. [PubMed: 11356754]
- Stoolmiller, M.; Bank, L. Autoregressive effects in structural equation models: We see some problems. In: Gottman, JM.; Sackett, G., editors. *The analysis of change*. Erlbaum; Hillsdale, NJ: 1995. p. 263–276.
- Stoolmiller M, Duncan TE, Bank L, Patterson GR. Some problems and solutions in the study of change: Significant patterns of client resistance. *Journal of Consulting and Clinical Psychology*. 1993; 61:920–928. [PubMed: 8113493]

- Stubbs, J.; Crosby, L.; Forgatch, MS.; Capaldi, DM. Family and peer process code: A synthesis of three Oregon Social Learning Center behavior codes. Training manual. 1998. Available at www.oslc.org/resources/codematerials/familypeerprocesscode.pdf
- Tremblay RE, Pagani-Kurtz L, Mâsse LC, Vitaro F, Pihl RO. A bimodal preventive intervention for disruptive kindergarten boys: Its impact through mid-adolescence. *Journal of Consulting and Clinical Psychology*. 1995; 63:560–568. [PubMed: 7673533]
- Waller G. Evidence-based treatment and therapist drift. *Behaviour Research and Therapy*. 2009; 47:119–127. [PubMed: 19036354]
- Waltz J, Addis ME, Koerner K, Jacobson NS. Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*. 1993; 61:620–630. [PubMed: 8370857]

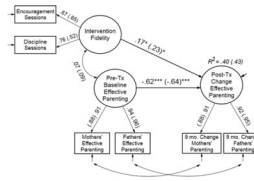


Figure 1. Structural equation model for effects of PMTO intervention fidelity on nine-month pre-post intervention change in effective parenting. Paths are standardized beta coefficients. Multilevel parameters adjusting for clustering in parentheses. $\chi^2(4) = .63, p = .96$; comparative fit index (CFI) = 1.00; root mean square error of approximation (RMSEA) = .00; *** $p < .001$; * $p < .05$.

Table 1

Means, standard deviations, and repeated measures F tests for fidelity across time for Generation 1 Therapists (n = 29)

Fidelity Measure	Early		Mid		Certification		F (1,28) Linear Time Effect	F (1,28) Quadratic Time Effect
	M	SD	M	SD	M	SD		
Knowledge	6.37	1.44	5.83	1.61	7.00	0.99	3.69	11.25**
Structure	6.19	1.41	5.86	1.69	7.04	1.09	6.92*	6.81*
Teach	5.91	1.46	5.54	1.77	6.82	1.09	7.41*	7.95**
Process	5.95	1.56	5.70	1.67	6.89	1.18	7.25*	5.91*
Overall	6.09	1.37	5.66	1.61	6.92	1.08	6.26*	10.11**
FIMP Scale	6.10	1.41	5.72	1.62	6.94	1.04	6.71*	8.88**

Note:

* $p < .05$;

** $p < .01$;

Table 2

Means, standard deviations, and mean F tests for therapist fidelity across three generations

Fidelity Measure	Generation 1 (n = 29)		Generation 2 (n = 51)		Generation 3 (n = 51)		F (2,128)	Significant Contrast
	M	SD	M	SD	M	SD		
Knowledge	7.00	0.99	6.58	0.70	7.13	0.60	7.56 ^{***}	2 < 1, 3
Structure	7.04	1.09	6.42	0.75	6.96	0.66	7.86 ^{**}	2 < 1, 3
Teach	6.82	1.09	6.19	0.71	6.80	0.72	9.01 ^{***}	2 < 1, 3
Process	6.89	1.18	6.17	0.75	6.87	0.71	10.95 ^{***}	2 < 1, 3
Overall	6.92	1.08	6.33	0.75	6.95	0.64	9.16 ^{***}	2 < 1, 3
FIMP Scale	6.94	1.04	6.34	0.71	6.94	0.63	9.73 ^{***}	2 < 1, 3

Note:

* $p < .05$;

** $p < .01$;

*** $p < .001$;

Table 3

Means, standard deviations and bivariate correlations for Study 3 variables

Variable	1	2	3	4	5	6	7	8
1. FIMP encouragement	---							
2. FIMP discipline	.49***	---						
3. Mother's parenting-Pre	.05	.08	---					
4. Father's parenting-Pre	.13	.07	.84***	---				
5. Mother's Parenting-Post	.13	.12	.33***	.20*	---			
6. Father's parenting-Post	.20†	.27**	.29**	.26**	.88***	---		
7. Δ Mother's parenting	.05	.07	-.59***	-.45***	.56***	.56***	---	
8. Δ Father's parenting	.07	.10	-.48***	-.16***	.58***	.61***	.87***	---
Mean	6.60	6.58	.62	.60	.66	.64	.02	.00
Standard deviation	.92	.87	.15	.16	.14	.15	.17	.18
SEM Model ICCs	.43	.43	.07	.04	---	---	.01	.02

Note. FIMP = Fidelity of Implementation rating system; Pre = pre intervention; Post = post intervention;

† $p < .10$;

* $p < .05$;

** $p < .01$;

*** $p < .001$;