

Program Evaluation Challenges: Is Aggression Replacement Training (ART) Effective?

Eva Feindler, Ph.D.¹ Emily Engel PsyD., & Meghann Gerber PsyD.

Abstract

Aggression Replacement Training (ART), a multi component intervention based upon CBT principles, is widely known as one of the only effective programs for aggressive adolescent offenders. This review of outcomes and methodological limitations is followed by a detailed program evaluation of ART across a set of group homes using standardized psychological measures. Outcomes were compromised by problems with treatment fidelity and adherence to the ART protocol. Suggestions for future research and practice in real world settings are included. At this time, it does not appear that ART has enough of an evidence base to be considered an efficacious treatment for adolescents with disruptive behavior.

ART Effectiveness Research

Research conducted over the past 30 years demonstrates that ART (Aggression Replacement Training) produces significant increases in constructive social behaviors and moral reasoning, and decreases in impulsivity and antisocial behavior (Goldstein, Glick & Gibbs, 1998; Feindler & Weisner, 2005). Versions of ART have been implemented in a variety of school and residential contexts and with a variety of populations across the United States and in Scandinavian and Russian communities abroad. Goldstein and Glick (1987) first evaluated ART at a limited- security institution for youth convicted of crimes, such as burglary, armed robbery, criminal mischief, and various drug offenses. Of the 60 youth who participated in the study, 24 received a 10-week ART program, which met three times per week. A second group of 24 participants were assigned to a non-ART, brief instruction control group. A third group of 12 youth not receiving ART or the brief instruction intervention was included as a non-treatment control group. Analysis revealed that ART participants acquired and transferred 4 of the 10 Skill streaming skills (Making a Complaint, Getting Ready for a Difficult Conversation, Dealing with Someone Else's Anger, and Dealing with Group Pressure) compared with both control groups (Goldstein & Glick, 1987). Significant differences were also found between ART and both control groups on the number and intensity of in-facility acting-out behaviors and impulsiveness. No between-group differences were evident on the moral reasoning measure, however.

A challenge to programs conducted within the safety of a residential center is that treatment effects may not be maintained once the participant is back in natural environments that trigger and reinforce antisocial behavior (Goldstein, Nensen, Daleflod & Kalt., 2004). During a 1-year period, 54 youth were discharged from this facility, 17 of whom had received ART (Goldstein & Glick, 1987). Researchers contacted the Division of Youth team members (parole officers), and collected follow-up data, without informing the workers of what treatment the participants had received.

¹ Director, Clinical Psychology Doctoral Program, LIU Post, 720 Northern Blvd., Lodge B, Room 101, Brookville, NY 11548, USA. 516-299-3212

ART participants received higher ratings on four of the six areas of community functioning (home and family, peer, legal, and overall) than control participants. Goldstein and Glick (1987) conducted their second ART effectiveness evaluation at a NY Division of Youth maximum security center for males age 13 to 21 years convicted of more serious crimes, such as murder, manslaughter, rape, sodomy, attempted murder, assault, and armed robbery. Compared to control groups, ART participants significantly acquired and transferred 5 of 10 Skill streaming skills, however, in this second study, these ART youth also demonstrated significant improvement in moral reasoning stages as measured by the Sociomoral Reflection Measure (Gibbs, Bassinger & Fuller, 1992). ART participants' utilization of constructive, prosocial behaviors also increased significantly, and their rated impulsiveness decreased. Unfortunately, no significant differences were found between these groups in the number or intensity of acting-out behaviors. This difference was attributed to the fact that this was a higher security center, and stronger controls were in place over residents, suggesting that youth at in this secure context are less likely to engage in acting-out behaviors, regardless of their treatment group (floor effect).

In order to discern whether ART could also be effective with delinquents reintegrating back into the community, Goldstein, Glick, Irwin, Pask-McCartney, & Rubama (1989) studied 84 adolescents following their release from detention. Youth were randomly assigned to three treatment groups: (1) ART provided to both youths and their families, (2) ART provided to youths only, and (3) a non-ART control group. Sixty-five participants attended two 1½-2 hour sessions per week for approximately three months. Analyses revealed that participants in both ART groups made significantly more improvements in interpersonal skill competence than non-ART adolescents as measured by The Skills Checklist (Goldstein, Sprafkin, Gershaw, & Klein, 1979 as cited in Goldstein et al., 1989). The same pattern was found in self-reported anger levels as measured by The Anger Situation Inventory (Hoshmand & Austin, 1987), though no differences were found in response to severe triggers (Goldstein et al., 1989). Further, because the majority of youth who commit another delinquent act do so within six months following their release, researchers collected data on recidivism rates of participants. Participants who received ART only had a 30% re-arrest rate, which was significantly lower than those who did not receive ART (43%), especially if families participated in ART as well (15%) (Goldstein et al., 1989).

As the aforementioned community-based evaluation demonstrated that ART can improve prosocial behavior when implemented in collaboration with a youth's family environment, Goldstein and colleagues' next step was to determine if the same improvements could be effected among peer groups of gang-involved delinquents (Goldstein, Glick, Carthan, & Blancero, 1994). The study was conducted in two community-based agencies in Brooklyn, New York– the Brownsville Community Neighborhood Action Center and Youth DARES of Coney Island. While researchers described the racial composition of the resident gangs where this study was conducted, specific demographic data of study participants was not reported. Staff members each administered three 4-month sequences of ART to members of the same gang– six gangs participated in ART (N = 38) and six participated in a non-ART control group (N = 27). Members of the same gang comprised each treatment group so that trainers could attempt to cultivate a more prosocial environment to ideally be transferred to their real-world environments. ART participants performed superiorly in the total Skill streaming skills score, as well as six target skills as measured by The Skills Checklist (Goldstein, Glick, Carthan, & Blancero, 1994). However, no significant differences were found between treatment groups in the domains of anger control and community functioning, with the exception of work adjustment, which improved in ART participants. Regarding recidivism, 13% of ART participants were rearrested, compared to 52% of the control group. These positive findings suggest that a more prosaically reinforcing peer environment was created among gang members.

In addition to Goldstein and colleagues' analyses, dissertation studies investigating the effectiveness of ART have been conducted. Kennedy (1989) investigated the effectiveness of implementing two ART components– Skill streaming and Anger Control Training– individually and conjointly with 37 incarcerated adult offenders with anger management difficulties. Analyses revealed that both interventions implemented singly and in combination, resulted in more appropriate anger expression, increased prosaically attitudes, decreased self-reported anger, and fewer adverse consequences following aggressive incidents as measured by self-report and objective behavioral ratings. In 1991, Jones compared three groups of aggressive male high school students (n=18) in Brisbane, Australia: ART, moral education, and a non-treatment control. Jones found that participants in the ART group exhibited a reduction in aggressive behaviors and impulsivity, but that these teens were also more aggressive at pre-test than those in the other groups.

Curulla (1991) conducted an effectiveness evaluation comparing three groups: (a) a 14-week ART program, (b) ART without Moral Reasoning Training, and (c) a non-ART control condition. Participants included 67 young adult offenders of a community agency in Seattle. Curulla reported that non-dilemma and control groups evidenced no reduction in recidivism, as measured by the Weekly Activity Record. While the dilemma group demonstrated the lowest actual recidivism rate, no significant difference in recidivism rates was found between groups. In 1998, Nodarse conducted an evaluation of ART implemented with 50 male and female adolescents, ages 12-14, with emotional handicaps: 25 were assigned to the experimental group, and 25 served as the control group. All participants received individual therapy, crisis management, adventure group therapy, and educational services, while the experimental group also received ART three times a week for ten weeks. Overt aggression and social skills were assessed by the Behavior Scale for Children prior to and following the intervention. Nodarse (1998) found that participation in ART was associated with statistically significant reductions in aggression and improvements in social skills and higher stage thinking however, the specific impact of ART alone could not be determined. Most recently, Grimes (2015) conducted a quasi-experimental evaluation of ART in mainstream school settings for 41 adolescents (treatment group=23, control group=18). Data on problem behavior, social-moral reasoning and social skills were gathered from teacher, parents and the students themselves. The only significant improvement occurred for the treatment group in moral reasoning, with a large effect size of $r=0.64$).

Leeman, Gibbs and Fuller (1993) implemented a constructive treatment strategy by adding ART to an existing motivation-oriented treatment for male youth in a medium security institution. In this quasi-experimental study, boys were assigned to one of three conditions: the motivation treatment, the motivation treatment plus ART, or a non-treatment control. Those who received the ART-enhanced intervention were found to have fewer conduct problems as measured by staff assessment of behavioral incident reports and unexcused school absences and by a self-report questionnaire designed by study authors. Upon release from the facility, recidivism rates for ART youth were 15% at six and twelve-month follow-up assessments, whereas the motivation treatment only and the non-treatment youth had recidivism rates from 25-30% at 6 months to 35-40% at twelve months respectively. These findings suggest that ART outcomes are stable to some degree. In a large study, Nugent, Bruley and Allen (1999) studied the effect of ART on 522 male and female adolescents in a runaway shelter. In this study, ART was condensed into 15 sessions and only included Skill streaming and Anger Control Training components.

Researchers conducted an interrupted time series design in which rates of participants' antisocial behavior were assessed prior to and following administration of ART. The results indicated that for all youth, there was a 20% decrease in antisocial behaviors after participation in ART. For girls, there was a uniform reduction in antisocial behaviors while for boys, antisocial behaviors were only reduced when there were smaller numbers of boys in the facility. The researchers suspect that may have been due to an increased number of conflicts when there are more teens or that staff may then change their criteria for determining antisocial behaviors when there is greater activity in the program (Nugent, et al,1999). The findings of this study are limited by the lack of experimental design, types of data gathered and the vague measure of "antisocial behavior," defined as "(a) a violation of rules and/or behavioral guidelines at the shelter, (b) a violation of legal or social norms, (c) a violation of another person's property, or (d) would be considered as aggression toward another person's physical or emotional well-being" (p. 470). Many of these evaluation studies show limited improvements for ART and it is difficult to separate out the effects of ART from other treatments provided for the adolescents.

In 2006, Gundersen and Svartdal published a mixed-design study evaluating the effectiveness of ART in Norway for 65 youth with behavioral difficulties (mean age = 14.1 years). Forty-seven participants from schools across Norway were divided into 11 groups of children who received 24-sessions of ART. The remaining 18 participants comprised the control group, receiving standard social and educational services. Researchers assessed participants' social problems and social skills prior to and following ART as measured by the Child Behavior Checklist (CBCL, Achenbach & Rescola, 2001), Child and Adolescent Disruptive Behavior Inventory 2.3 (CADBI, Burns, Taylor, & Rusby, 2001), Social Skills Rating System (SSRS, Gresham & Elliot, 1990), and How I Think (HIT, Gibbs, Barriga, & Potter, 2001 as cited in Gundersen & Svartdal, 2006). This study was one of few to utilize a battery of reliable assessment tools. Results indicated that ART participants demonstrated superior improvement in social skills and reduction in behavior problems (Gundersen & Svartdal, 2006). Contrast analyses were conducted between ART and control group participant performance on the SSRS, as reported by parents, teachers and youth. Results indicated that parents and teachers reported significant improvements in social skills among ART participants, as compared to the control group.

No significant improvements were reported by youth, with the exception of responses on a self-report measure created by the researchers. Finally, comparison analyses were conducted on parent, teacher and youth scores on the CADBI and CBCL. Again, parent and teacher reports produced significant reductions in problem behavior among ART participants as compared to control group participants. Youth responses on the aforementioned self-report measure supported this finding, but youth scores on the CBCL indicated that both groups reported decreases in problem behavior. Again, researchers suggested that this could be due to dilution, or a possible response bias.

Additional studies after 2000 have shown some gains in social skills, decreased aggression, improved cognitions and reduced recidivism. The first is a study that was conducted by the Washington State Institute for public Policy (WSIPP) (Barnoski, 2004). In 1997, The Washington the Community Juvenile Accountability Act established “research-based” programs in the state’s Juvenile Court. Research found that when competently delivered, ART has positive outcomes with estimated reductions in 18-month recidivism of 24% and a benefit to cost ration of \$11.66. The study was initiated in July 1999 with the final report submitted December 2003. Two unique qualities were that they did have a control group $n=525$ and the program group $n=704$ for a total sample population of 1,229 and they developed a specific Quality Assurance program to assure model adherence. The finding for felony recidivism is statistically significant at the $p=.125$ probability level. They also were so intent on program integrity and a supervisor provided oversight for training, fidelity compliance and quality assurance. The second large scale study was conducted by the California Institute for Mental Health (CMHI, 2009). Their biannual report reflects a commitment to support real world implementation of ART across the state and reflected 3,482 participants that were provided ART in 4 different settings: Juvenile Hall, Group Home/ Residential Treatment Settings, Secure Juvenile Justice Settings (Camps and Ranches) and Outpatient Settings.

All sites trained to implement ART, were encouraged to collect and submit program evaluation data. The CMHI supports quality improvement evaluation on an ongoing basis, thereby providing motivated sites with the ability to monitor the performance of their ART programs. The aggregate data suggested an extremely diverse population: 21.6% female and 76.4% male; African American 13.5%, Asian 2.8%, Caucasian 30%, Hispanic/Latino 43.8%, Native American .6%, and Other 1.4%. The delivery of the model varied from site to site; however, the outcomes across sites are measured by a core set of identified tools including: The Skill Streaming Checklist (Goldstein, 1998), The Aggression Questionnaire: AQ (Buss & Perry, 1992) and the HIT (Gibbs, Barriga & Potter, 2001). Staff reported an increase of 24% in youth’s use of constructive, prosocial skills; and youth self-reports identified a 12% improvement in the use of prosocial skills, an 11% improvement in the ability to manage aggression, and a 9% increase in the use of more mature, constructive thought processes. All improvements from pre- to post-ART were statistically significant differences ($p<.01$).

Another evaluation with juvenile offenders ($n=57$) in residential treatment in Sweden was reported by Holmqvist, Hill and Lang (2009). ART training was combined with a traditional token economy for this population between the ages of 16 and 19. Compared to other treatment at the facilities, ART did not result in any identifiable changes in usual offender outcome markers. Case reports indicated that individual motivation for treatment is an important consideration. More recently, Currie, Wood, Williams & Bates (2012), in an evaluation of ART in Australia with a small sample ($N=20$) young adult offenders in a custodial setting, specifically examined changes in self-reported impulsivity, aggression, cognitive distortions and social problem solving. Significant decreases in these measures (HIT, AQ, I-Questionnaire and SPSI) were reported at multiple assessment times including follow-up, however, no changes were reported on the Adult Behavior Checklist. Kopusov, Gundersen and Svartdal (2014) reported an evaluation of an extended version of ART which included character education and was implemented by teachers in schools in Russia. The 30 sessions, delivered over 10 weeks occurred in small groups of children. This study included a large sample of children from 6-16 years of age ($n=232$), parents and teachers and compared program outcomes to children in a control sample.

Comparative results from both parent and teacher versions of the SSPS (Social Skills Rating System) and children’s self-report indicated that the greatest changes in social competence ratings and problem behavior scores occurred for the younger children. There were however, changes also noted for children in the control group across time. The authors note that the “diffusion of treatment” played a role, as the children from the intervention and control groups interacted during the school day and might be considered a positive side-effect (Kopusov et al, 2014).

ART research has progressed substantially since its creation by Goldstein and colleagues over two decades ago. It has been implemented in at least 45 states, six Canadian provinces and numerous foreign countries, and is used with a variety of juvenile populations (Glick, 2006). ART has gradually gained support as an evidenced based treatment effective in increasing prosocial behavior and moral reasoning abilities, and reducing impulsivity and antisocial behavior. Although the available literature strongly suggests that ART is effective in teaching social skills, improving anger management skills, and reducing self-reported anger and recidivism rates among juvenile delinquents (Goldstein et al., 2004), a recent meta-analysis of ART interventions for adults and adolescents concluded something different. Brannstrom, Kaunitz, Andershed, South & Smedslund (2016) conducted a systematic review of ART programs for adults and adolescents and initially identified 749 research reports of ART evaluations with adolescents and adults. However, only 16 studies were included in their review and they had considerable clinical and methodological diversity. Further, as there were no consistencies in evaluation measures used (recidivism was primary outcome measure and change in any of the three domains-anger control, social skill and moral reasoning- were secondary outcomes) or timing of assessments, a statistical meta-analysis was not possible. Brannstrom et al (2016) concluded that ART only seems successful in that it is wide-spread and frequently used for offender populations and recommended that more reliable and valid research methodology be used to explore the specific components of ART with clearly defined target groups and well established protocols.

Limitations in Existent Research

The ability for previous research findings to be replicated and generalized to other populations of juveniles who engage in delinquent behavior is limited due to some methodological weaknesses in the available ART research. Homogeneity of samples and a lack of reporting on ethnicity in early research threaten the external validity of past findings. For example, three samples consisted of predominantly white males (Kennedy, 1989; Curulla, 1991; Leeman et al., 1993), racial demographic data were not available in other studies (Goldstein et al., 1989; Jones, 1991; Coleman et al., 1992; Gundersen & Svartdal, 2006), and only four experimental ART studies included multiethnic samples (Goldstein et al., 1987, 1994; Nodarse 1998; Nugent et al., 1999). Given that minorities are overrepresented compared to whites in the delinquent population (Glick, 2003) this shortcoming limits consumers' ability to comprehend findings' implications or replicate findings (Dowden, Antonowicz & Andrews, 2003).

One limitation of many of the studies evaluating ART is the type of measures and data used to draw conclusions. Most ART studies have primarily relied upon recidivism data and subjective ratings by institution staff to determine the treatment's effectiveness (e.g., Barnoski, 2004; Goldstein, Glick, Carthan & Blancero, 1994; Leeman, Gibbs, & Fuller, 1993; Nugent, et al, 1999). While overt behavioral changes observable by staff and incidents of arrest are important markers of change for a juvenile delinquent population, these do not capture all elements of aggressive and antisocial behavior problems that are included in the cognitive-behavioral conceptualization upon which ART is based. In addition, these types of observations should be made using valid and reliable standardized measures. Furthermore, behavioral incident write-ups by staff indicates that only acting-out behaviors occurring within view of staff are recorded, and can have compromised reliability as the raters vary in terms of training and judgment. Recidivism data only accounts for those youth whose behavior problems warrant arrest, and behaviors that occurred in such a manner that the youth "got caught." Although it is hardly possible to capture all of the data one may wish to have on youth who undergo ART, Feindler and Baker (2004) recommend using measures that assess direct treatment targets, and the use of multiple sources of data collection (i.e., avoid relying entirely on ratings by staff or self-report, etc.). This was best represented in the Currie et. al. (2012) ART evaluation with a young adult offenders using several self-report measures tapping into specific domains of content reflected in the ART program. Another limitation is that Goldstein and colleagues (1987, 1994, 1998) were involved in approximately half of the foundational ART studies. While continued evaluation is necessary to the development of a novel treatment package, its creators' close involvement may carry two problems. First, they were not independent observers, and therefore their own biases may have affected their conclusions. Second, their close oversight may have affected the way the treatment was implemented. Treatment integrity may have been upheld more so than it would be in a natural setting, where practitioners may be less experienced with and motivated to conduct ideal ART. Therefore, Goldstein et al.'s analyses and subsequent conclusions may not be representative of ART effectiveness in real-world settings. In a meta-analysis of all ART programs reported in the literature between 1987 and 2014, Brannstrom et al (2016) could only find 16 studies of ART as a stand-alone program and the methodological quality was limited. Half of these studies were reported by researchers with vested interest in the outcomes.

Research in the Context of Usual Care:

Agency and school district budgets, available resources and time constraints inevitably affect treatment delivery in real-world settings. Treatment integrity may suffer, trainer motivation may waiver, and client attrition may prohibit them from obtaining the optimal benefit from ART. As such, Goldstein et al., (2004) asserted that a vital need exists for applied research in real-world settings to bridge the gap between research and practice. Like other ART studies reviewed, program evaluation research utilizes data collected in a ‘real world’ clinical as opposed to laboratory setting. However, there are significant drawbacks to attempting to draw valid conclusions about outcomes of a specific treatment in the context of usual care. Staff turnover, population heterogeneity, and site differences, in addition to the lack of standard controls available in a laboratory setting, such as random assignment, uniform delivery of treatment, and data collection by those trained in research methodology, all threaten the internal validity of findings in this type of research. However, assessing a treatment as it is actually provided within the constraints of a real world setting provides the greatest opportunity for external validity - laboratory studies are often criticized for lack thereof.

The conflict over treatment outcome research findings derived from laboratory studies (known as efficacy research) and their applicability to diverse populations in real-world settings has been widely observed (see Glasgow, Klesges, Dziewaltowski, Bull, & Estabrooks, 2004; Weisz, Jensen, & McLeod, 2005). The traditional model of intervention research begins with efficacy studies to establish the ability of a given intervention to create change in tightly controlled conditions. Promising interventions are then implemented under real world conditions, typically to a more diverse population, and subsequent evaluation of treatment effect is known as effectiveness research. Glasgow, Lichtenstein, and Marcus (2003) argue that this manner of progression in intervention research is flawed because the very conditions that contribute to an intervention’s success in a laboratory setting are at odds with what exists in the real world clinical setting.

Some have proposed alternative models of intervention research that challenge rigid distinction of laboratory investigation from the context of usual care in real world settings. Hoagwood, Hibbs, Brent, and Jensen (1995) propose moving away from the linear model and toward a dimensional model that allows components of each approach to influence research in a bi-directional manner. Garland, Hurlburt, and Hawley (2006) describe their attempts at hybrid research as an effort to utilize tools and techniques from the laboratory to study treatment in the context of usual care, that is, to have more precise measurement of what is actually happening in the real world. Ultimately, Kolko (2006) argues that the most useful research on treatment interventions has to strike a balance between “rigor and relevance” (p. 49). The remainder of this article aims to present a model for program evaluation by making use of measurements that have been established to be valid and reliable indicators of the constructs of concern, while also accounting for issues of treatment fidelity and staff training. In applied program evaluations, the setting of the treatment delivery and the population of youth participants are variable and outside the domain of laboratory controls, which very much resembles the way ART has already been implemented in numerous sites across the country and beyond. We hope that our model provides a reasonable balance of rigor and relevance for those seeking to evaluate the efficacy and effectiveness of intervention programs.

Program Evaluation:

In order to evaluate the impact of ART on distorted cognitions, aggressive behaviors, emotion management and non-aggressive antisocial behaviors in boys and girls, reliable data needs to be collected on youth receiving ART or prosaically skills training.

Instruments designed to measure direct targets of ART and to assess related construct of importance given what is known about aggressive and antisocial behaviors will be most relevant. The following is a detailed description of a program evaluation project and is suggested as one type of comprehensive approach tailored for the “real” clinical setting but also meeting the rigorous standards for providing empirical data.

Tidewater Program Evaluation

Method and Procedures of Evaluation

The data used for this program evaluation were collected from November 2006 to August 2008 at two sites where Tidewater Regional Group Home Commission implemented its ART program: Chesapeake Boys Home and Centerville Girls Home. Independent groups of staff led ART groups and collected data in each group home.

The amount of experience and ART training varied among staff, and some staff members changed over the course of data collection: The girls' group home had more turnover of ART group leaders and obtained a new site director approximately one year into data collection. Participants (N = 58) consisted of males and females, ages 13 – 17, with a mean age of 15.7 years. They represented diverse ethnic and socio-economic groups. All residents were required to participate in the ART program, and participation commenced upon entry into the group home. Participation in the program evaluation was determined by informed consent of parent or legal guardian and assent by youth resident, both of which were solicited and obtained by agency staff. Residents received ART sessions three times a week for ten weeks, in addition to a number of other services. Upon entry into the ART program and after completion of the program, agency staff administered an assessment packet that included several self-report and some staff-report measures.

Assessment Tools

After a comprehensive review of psychometrically sound assessment devices suited to an adolescent population, the following measures were selected for use in our program evaluation based on the need for measurement of the psychological constructs which the ART programs target for change: social skills, anger management and moral reasoning. Behavior Assessment Scales for Children, Second Edition (BASC-2) (Reynolds & Kamphaus, 2004) The BASC-2 is a comprehensive instrument that measures multiple aspects of a youth's behavior, in both adaptive and maladaptive respects. The measure generates numerous subscales and composite indices. This program evaluation collected data using two versions of the BASC-2: a teacher rating scale (TRS, contains 139 items), and a youth self-report inventory of personality (SRP, contains 176 items). The following portions from this instrument were used: The primary scales Aggression (composed of 10 items) and Conduct Problems (12 items) from the TRS, and Sensation Seeking (9 items) from the SRP and the Anger Control scale from both the TRS (10 items) and the SRP (14 items).

How I Think Questionnaire (HIT-Q; Gibbs, et al, 2001). The HIT-Q is a 54-item self-report scale designed to measure cognitive distortions among youth with antisocial behavior problems, and features several subscales which are hypothesized to correspond with aggressive behaviors. It was initially tested on 147 male subjects, aged 14 – 20 years, in three different settings, including an urban working class high school, a youth correctional facility, and an upper middle class high school (Barriga & Gibbs, 1996). Test-retest reliability was shown to be high, $r(135) = .91$, $P < .0001$. A recent meta-analysis across 29 independent samples indicated that there was a mean alpha coefficient of .93 for internal consistency and the HIT-Q was also shown to have strong convergent validity (Gini & Pozzoli, 2013).

Multidimensional School Anger Inventory (MSAI; Smith, Furlong, Bates & Laughlin, 1998). The MSAI is a self-report instrument, featuring 39 items, designed to measure cognitive, behavioral, and affective dimensions of anger in the context of a school setting. The MSAI includes four subscales; the most appropriate one, Anger Experience, consists of 13 items. The MSAI has been studied on several groups of adolescents, with the largest group (used for establishing psychometric properties) consisting of 1,589 high school students in grades 9 through 12 (Furlong, Smith & Bates, 2002). In the initial instrument evaluation, the MSAI was found to have good internal consistency, with Anger Experience subscale having the highest alpha coefficients across three samples, ranging .84 - .88 (Smith, et al 1998). Convergent validity was established via comparison with the Aggression Questionnaire [AQ; Buss & Perry, 2002] and the Connors Teacher Rating Scale [CTRS, Connors, 1989]. Data were collected for 58 youth upon entry into the group home prior to starting ART. Post-ART assessment packets were completed for 40 of those youth at the end of their group home placement. Eighteen participants (non-completers) were transitioned out of their group home before the second assessment could be administered.

Tidewater Results

Profile of TRGHC Youth

No significant difference was found in pre-test levels of aggression between treatment completers (n = 40) and non-completers (n = 18). No significant differences were found in demographic information between completers and non-completers with the exception of age, $t(69) = 3.052$, $p < .01$. Non-completers were significantly older than completers. Additionally, completers were more likely than non-completers to live with their mother only, while non-completers were more likely than completers to be in the custody of the Department of Social Services.

Several elevated mean scores in the pre-ART assessment indicate that the youth in this sample represent a population with more severe behavioral, cognitive, and emotional problems when compared to youth in the general population. Mean scores for youth at pre-treatment were compared with normative data that accompanied each measure. According to staff ratings, many of the constructs targeted by ART (aggressive behaviors, hostile thinking patterns, emotion management) were found to be in the clinical and subclinical range among both boys and girls in the sample. However, there were no elevated mean scores for either gender on any of the self-report scales on the MSAI or BASC-2.

The BASC-2 TRS (staff report) showed several elevated mean scores for both boys and girls in this study. Only girls were found to have elevated scores in the Clinically Significant range, which were seen on the scales Aggression, Bullying, and Conduct Problems. Scores in the At Risk range for girls were seen on the following scales: Externalizing Problems, Adaptability, Emotional Self Control, Executive Functioning, Negative Emotionality, Resiliency, Depression, Internalizing Problems, Behavioral Symptoms Index and Anger Control. For males, there were no mean scores in the Clinically Significant range on the BASC-2 TRS (staff report) scales. Mean scores in the At Risk range were seen on the following scales: Adaptability, Adaptive Skills composite, Emotional Self Control, Executive Functioning, Negative Emotionality, Resiliency, Aggression, Conduct Problems, Study Skills and Anger Control.

On the Overall HIT scale, both boys ($M = 2.96$, $SD = .84$) and girls ($M = 2.85$, $SD = .70$) had elevated mean scores compared with the normative sample ($M = 2.39$, $SD = .69$) which placed both groups in the "Borderline Clinical" range according to the measure's authors (Barriga, Gibbs, Potter & Liao, 2001). The Overall HIT was the only scale generated by a self-report measure that showed clinically significant elevated mean scores for the youth participants. All other significantly elevated scores were observed only on staff-rating scales. This disparity appears to indicate that staff rated youths' problems more severely than youth rated themselves on the self-report measures.

Gender Differences in Youth at Pre-ART Assessment

Significant differences between genders were only seen on mean scores from the staff report measures; none were found on the self-report scales. According to staff report ratings, female ART participants' scores were found to be significantly more severe compared to their male counterparts on the following scales: Hyperactivity, $t(56) = 4.42$, $p < .001$; Aggression $t(56) = 3.06$, $p < .01$; Conduct Problems, $t(56) = 2.51$, $p < .05$; Externalizing Problems, $t(56) = 3.56$, $p < .01$; Internalizing Problems, $t(56) = 2.17$, $p < .05$; Anxiety, $t(56) = 2.27$, $p < .05$; Attention Problems, $t(56) = 2.81$, $p < .01$; School Problems, $t(56) = 2.45$, $p < .05$; Behavioral Symptoms Index, $t(56) = 2.91$, $p < .01$; Bullying, $t(56) = 4.04$, $p < .001$; Emotional Self Control, $t(56) = 2.71$, $p < .01$; and Executive Functioning $t(56) = 3.07$, $p < .01$. Males were found to have more severe scores than the female participants on the scales Study Skills, $t(56) = 3.88$, $p < .001$ and Leadership, $t(56) = 2.37$, $p < .05$.

The number of scales on which significant differences were found between male and female participants could be indicative of a more clinically severe population of female residents compared to their male counterparts. However, despite these findings on the 14 of the 27 staff report measures, no differences were found on 44 self-report scales among gender groups. This inconsistency suggests that the finding of difference is representative of differing rating practices among the female and male group home staff.

Differences in ratings could be due to any number of context-relevant variables including level of observation, varying rules about in-home behavior, rigidity of beliefs about appropriate and inappropriate behavior for girls or boys, and cohort effects among group home staff. It should also be noted that there were considerable differences in the standard practices of each group home's staff. The research team was informed by the agency that while it was the intention that the same group home staff members fill out both pre- and post-assessment measures for each youth, this was not a consistent practice. The boys' group home staff attempted to have the same individual complete the participant's assessments, while the girls' group home staff chose to fill out measures as a group. Group staff rating versus individual staff rating presents unknown differences in the process of assessing youth behavior.

Differences between the Pre-ART and Post-ART Assessments

Pre-ART and post-ART assessment scores were examined to determine if there were significant changes in youths' scores after completing the ART program. When examining all youth together, there were no significant differences found from pre- to post-treatment scores for any of the scales examined as part of the program evaluation. Given the gender differences in staff rating scales, each gender group was examined independently. One significant difference was found for the male participants only: The post-ART mean score on the Adaptability scale of the BASC-2 TRS was significantly higher than the pre-ART mean score $t(19) = 2.51, p < .05$. A pattern of similar findings on scales that measure related constructs would support a conclusion that behavior demonstrating adaptability in boys improved from the pre to post assessment.

Findings on Aggression

The current sample's mean aggression scores fall in the "at risk" range of aggression (pre-test aggression $M = 64.875, SD = 11.643$; post-test aggression $M = 64.641, SD = 9.557$). Thirty percent of participants scored in the "clinically significant" range. Females had significantly higher scores on aggression scales than males at pre-treatment, $F(57) = 9.075, p < .01$. For the youth in this evaluation, completing the ART program was not associated with a decrease in staff reported aggression as no significant difference was found between pre- and post-test BASC-2 TRS aggression scores, $t(38) = .045, p = .964$ for the 40 completers.

It should be noted that aggression scores for some participants with lower pre-treatment aggression scores actually increased following treatment, suggesting that an iatrogenic effect occurred whereby group processes adversely influenced participants' behavior across treatment. Arnold and Hughes (1999) theorized that risks of grouping deviant youth together for skills training include negative changes in participants' attitudes toward antisocial behavior as well as affiliation and identification with antisocial peers. This is especially likely in a residential treatment center where participants live and socialize with fellow group members.

Obstacles in Assessment of ART as Viable Treatment

The single largest threat to TRGHC's ART program evaluation was the degree of adherence to the ART model. Treatment fidelity was assessed through coding and scoring recorded ART sessions that had been conducted in each group. Agency staff agreed to video-record 1-2 sessions from each ART module for each 30-session period of the ongoing ART groups in order to evaluate treatment fidelity. Nineteen videotapes were made available to the research team for coding in order to determine the degree of adherence to the ART curriculum. The content of treatment fidelity coding forms (See examples in Appendix A) closely mirrors the explicit directives given in the ART treatment manual for each session, with the intent of isolating the specific group leader behaviors (Goldstein, et al, 1998). The adherence coding procedure was based on a binomial rating system that required coders to indicate whether or not the explicit directives in the manual were present in a given taped session. Acceptable inter-observer agreement between the two coders was established using Cohen's kappa ($k = .66$). Adherence scores for each session were calculated as the mean of the two coders' ratings and represent the percentage of the ART curriculum directives that were observed in the recorded group sessions. Of the 19 sessions coded, the mean adherence rating was 47%, with a range of 19% to 75%.

ART groups conducted within the first four months of the staff ART training ($n = 11, M = .57, SD = .14$) had significantly higher adherence ratings than those conducted 15 – 22 months after the training ($n = 8, M = .34, SD = .13$), $t(17) = 3.63, p < .01$. In addition, the ART groups that were conducted by staff in the girls' group home ($n = 8, M = .59, SD = .18$) had significantly higher adherence scores than those conducted by the boys' group home staff ($n = 11, M = .39, SD = .10$), $t(17) = 3.09, p < .01$. The length of ART group sessions ranged from 10.5 minutes to over 57 minutes. ART group session length was found to be positively correlated with treatment adherence scores, Pearson's $r = .71, n = 19, p < .01$. No significant differences were detected among the mean adherence scores for each of the three modules.

The problems associated with adherence to the ART curriculum are likely the most significant obstacle in this evaluation of ART. On average, less than half of the explicit directives in the ART manual were actually part of the recorded ART sessions conducted in the group homes. When taking into account the length of ART sessions, there was a clear relationship between the degree of adherence and the amount of time dedicated to a given session. This finding indicates that shortening sessions (the manual recommends an hour per session) directly impacts treatment fidelity.

Frequently overlooked treatment components in recorded sessions were homework assignment review and asking participants to identify specific situations in their own lives in which the new skills could be useful. Although both components take considerable time they are included in the ART curriculum to maximize the possibility that youth will use the new skills in their lives outside of the ART session. Another significant concern apparent from observations of the taped sessions was that group leaders frequently conducted close approximations of the ART intervention, but did not execute the features of the treatment that are critical for behavior change.

Recommendations

Assessment of Readiness

Before an agency or school district evaluates the various strategies for meeting training needs required for a more successful ART program, it may be helpful to assess if an organization has assessed its capacity to undertake the level of training and commitment involved in the implementation of a new treatment program. The growing literature examining obstacles to successful implementation of empirically based treatments has called attention to the critical first step of identifying an organization's readiness to put new programs into effect. Saldana, Chapman, Henggeler & Rowland (2007) point out that "front line" practitioners are most heavily and directly affected by an organization's decision to start a new treatment program, and therefore, attending to the variables that affect an individual's willingness to try something new could be highly worthwhile. In addition, organizational culture, including an agency's priorities, expectations, and values, as well as the climate that encompasses practitioners' perceptions of their work environment may all affect the success of a program implementation strategy (Glisson, Landsverk, Schoenwald, Kelleher, Hoagwood, Mayberg & Green, 2008). Lehman, Greener & Simpson (2002) created an Organizational Readiness for Change Instrument that may assist in the assessment of the system's ability to change. Issues that arise during assessment should be addressed among staff and administration in a collaborative manner.

Training

Training alone is not enough for an agency or school to implement evidence based practice with fidelity. For the average organization, it takes supervision from a Master Trainer for 12 months following the initial clinical training. We recommend a fidelity checklist (see Appendix) be filled out by the facilitators and consultation phone calls with a Master Trainer to assist with program drift. A booster training within 4 months is helpful and finally, we recommend videotapes be submitted to the Master Trainer for observation and feedback with program integrity. Workshops alone, without ongoing supervision and consultation, are insufficient to establish treatment effectiveness (Lochman, Boxmeyer, Powell, Qu, Wells, & Windle, 2009). Availability of supervision and feedback are critical to support a new treatment program. Periodic assessment of how the ART groups are conducted through review of the videotaped sessions is in itself a valuable training component for ART group leaders that ideally would be conducted on an ongoing basis.

Koposov, et al (2014) in their large investigation of an extended version of ART in school settings reported 72 hours of training and 18 training sessions with ART master clinicians before the teachers conducted the 30 sessions. The process of session review offers the opportunity for both quantitative and qualitative evaluations of adherence to the treatment and could be used as a mechanism to provide feedback and reinforcement to new trainees. In fact, Currie et al (2012) assessed treatment fidelity on an ongoing basis by sending session videos to a Washington State ART Quality Assurance Specialist for review.

Time between ART training of staff and administration of ART also appeared to have an important impact on treatment adherence. Groups conducted in the time period immediately following ART training will be more adherent than those conducted more than a year after leaders had received training. This is a strong indicator that frequent trainings are critical for maintaining the integrity of an ART program. Koposov et al (2014) report the use of an implementation scale that as of now has not been translated and disseminated. Restrictive budgets and limited resources are formidable obstacles to obtaining more intense levels of training and support. Some providers have cut costs by organizing interagency trainings in which resources are pooled to engage trainers for staff from multiple sites. Staff members can also take advantage of new technologies, such as supervision blogs to increase their level of engagement in the training process. Innovative ways to maximize staff members' exposure to training opportunities may be vital to the success of a treatment program.

Collaboration with the ART Research Community

When a consultation is initiated between a direct service agency and a research community, it is challenging but important to create an arrangement that balances the needs to protect participants' anonymity, maximize convenience, and allow for clear communication between researchers and agency staff members. Research consultants, administrators, and staff should work collaboratively in the planning process of the organizational change and associated evaluation. Agency staff would then have a sense of ownership in the evaluation, and researchers could share their expertise by way of providing staff with training around measure administration to maximize validity and reliability in the data collection process. Research consultants could advise staff about potential threats to data integrity and assist in the identification of individuals who would have a lower potential for biased ratings, such as teachers or probation officers not associated with the administration of ART groups. Given the issues with staff rating that may occur, future evaluations may need additional sources of data, such as school records or third party observations in the classroom or group home environment.

Conclusion

It is hoped that this review of available research from interventions and program evaluations will help to frame the methodological issues and challenges in providing rigorous data to support ART as an empirically promising intervention. While there are not yet data from randomized clinical trials, enough evidence exists to consider ART as a viable treatment approach for working in residential and school settings with a variety of youth exhibiting aggressive behavior problems. It is recommended that reliable and valid assessment methods be used in a structured fashion so as not to rely solely upon recidivism as a measure of treatment effectiveness. However, staff training and supervision of ART implementation is critical to ensure that the ART curriculum is disseminated as developed.

References

- Arnold, M. E., & Hughes, J. N. (1999). First, do no harm: Adverse effects of grouping deviant youth for skills training. *Journal of School Psychology, 37*, 99 – 115. doi: 10.1016/S0022-4405(98)00027-2
- Barnoski, R. (2004). Outcome evaluation of Washington State's research-based programs for juvenile offenders. Olympia, WA: Washington State Institute for Public Policy
- Barriga, A. Q., & Gibbs, J. C. (1996). Measuring cognitive distortion in antisocial youth: Development and preliminary validation of the "How I Think" questionnaire. *Aggressive Behavior, 22*(5), 333-343. doi:10.1002/(SICI)1098-2337(1996)22:5<333::AID-AB2>3.0.CO;2-K
- Barriga, A. Q., Gibbs, J. C., Potter, G., & Liau, A. K. (2001). *The How I Think Questionnaire manual*. Champaign, IL: Research Press.
- Brannstrom, L., Kaunitz, C., Andershed, A., South, S. & Smedslund, G. (2016) Aggression Replacement Training (ART) for reducing antisocial behavior in adolescents and adults: A systematic review. *Aggression and Violent Behavior, 27*, 30-41. doi:10.1016/j.avb.2016.02.006
- Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology, 63*, 452-459. doi: 10.1037/0022-3514.63.3.452
- California Institute for Mental Health (2009). *Teaching Pro-Social Skills to Antisocial Youth Using the Aggression Replacement Training Curriculum: Aggregate Evaluation Dashboard Report*, Unpublished report, available upon request
- Coleman, M., Pfeiffer, S. and Oakland, T. (1992) Aggression replacement training with behaviorally disordered adolescents. *Behavioral disorders: 54-66*.
- Currie, M, Wood, C., Williams, B. & Bates, G. (2012) Aggression Replacement Training (ART) in Australia: A longitudinal youth justice evaluation. *Psychiatry, Psychology and Law, 19:4, 577-604*, doi:10.1080/13218719.2011.615807
- Curulla, V.L. (1991) *Aggression Replacement Training in the community for adult learning disabled offenders*. Unpublished doctoral dissertation, University of Seattle, WA.
- Dowden, C., Antonowicz, D., & Andrews, D. A. (2003). The effectiveness of relapse prevention with offenders: A meta-analysis. *International Journal of Offender Therapy and Comparative Criminology, 47*(5), 516-528. doi:10.1177/0306624X03253018

- Feindler, E. L., & Baker, K. (2004). Current issues in anger management interventions with youth. In A. P. Goldstein, R. Nensen, B. Daleflod, & M. Kalt (Eds.), *New perspectives on Aggression Replacement Training: Practice, research and application* (pp. 31-50). Hoboken, NJ: John Wiley & Sons.
- Feindler, E. L., & Weisner, S. (2005). Youth anger management treatments for school violence prevention. In S. R. Jimerson & M. J. Furlong (Eds.), *Handbook of school violence and school safety: From research to practice* (pp. 353-363). Mahwah, NJ: Lawrence Erlbaum.
- Furlong, M., Smith, D. & Bates, M. (2002) Further development of the Multidimensional School Anger Inventory: Construct validation, extension to female adolescents and preliminary norms. *Journal of Psycho educational Assessment*, 20, 46-65.
- Garland, A. F., Hurlburt, M. S., & Hawley, K. M. (2006). Examining psychotherapy processes in a services research context. *Clinical Psychology: Science and Practice* 13(1), 30-46. doi: 10.1111/j.1468-2850.2006.00004.x
- Gibbs, J. C., Basinger, K. S., & Fuller, D. (1992). *Moral maturity: Measuring the development of socio moral reflection*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gibbs, J. C., Barriga, A. Q., & Potter, G. B. (2001). *How I think (HIT) questionnaire*. Champaign, IL: Research Press.
- Gini, G. & Pozzoli, T. (2013) Measuring self-serving cognitive distortions: A meta-analysis of the psychometric properties of the How I Think Questionnaire (HIT). *European Journal of Developmental Psychology*, 10, 510-517. doi:10.1080/17405629.2012.707312
- Glasgow, R. E., Klesges, L. M., Dzewaltowski, D. A., Bull, S. S., & Estabrooks, P. (2004). The future of health behavior change research: What is needed to improve translation of research into health promotion practice? *Annals of Behavioral Medicine*, 27, 3-12. doi: 10.1207/s15324796abm2701_2
- Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health*, 93(8), 1261-1267. doi: 10.2105/AJPH.93.8.1261
- Glick, B. (2003). Cognitive programs: Coming of age in corrections. *Corrections Today*, 65(1), 78-80.
- Glick, B. (2006). *Cognitive behavioral interventions for at-risk youth (Vol. 1)*. Kingston, NJ: Civic Research Institute, Inc.
- Glisson, C., Landsverk, J., Schoenwald, S., Kelleher, K., Hoagwood, K. E., Mayberg, S., & Green, P. (2008). Assessing the organizational social context (OSC) of mental health services: Implications for research and practice. *Administration and Policy in Mental Health and Mental Health Services Research*, 35, 98-113. doi: 10.1007/s10488-007-0148-5
- Goldstein, A. P., & Glick, B. (1987). *Aggression Replacement Training: A comprehensive intervention for aggressive youth*. Champaign, IL: Research Press.
- Goldstein, A. P., Glick, B., Irwin, M. J., Pask-McCartney, C., & Rubama, I. (1989). *Reducing delinquency: Intervention in the community*. Pergamon Press.
- Goldstein, A. P., Glick, B., Carthan, M., & Blancero, D. (1994). *The prosocial gang: Implementing Aggression Replacement Training*. Thousand Oaks, CA: Sage.
- Goldstein, A. P., Glick, B., & Gibbs, J. C. (1998). *Aggression Replacement Training: A comprehensive intervention for aggressive youth (Rev. ed.)*. Champaign, IL: Research Press.
- Goldstein, A., Nensen, R., Daleflod, B. & Kalt, M. (2004) *New perspectives on Aggression Replacement Training: Practice, Research and Application*. Chichester, John Wiley and Sons.
- Gresham, F. & Elliott, S. (1990) *Social skills rating system (SSRS)*. American Guidance Service.
- Grimes, S. (2015) *An evaluation of Aggression Replacement Training: The impact of a multi- component, CBT based intervention on the problem behavior, pro-social skills and moral development of pupils in English secondary schools*. Unpublished dissertation, University of Nottingham, UK.
- Gundersen, K., & Svarddal, F. (2006). Aggression replacement training in Norway: Outcome evaluation of 11 Norwegian student projects. *Scandinavian Journal of Educational Research*, 50(1), 63-81. doi: 10.1080/00313830500372059
- Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Introduction to the special section: Efficacy and effectiveness in studies of child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology*, 63(5), 683-687. doi: 10.1037/0022-006X.63.5.683

- Hoshmand, L. T., & Austin, G. W. (1987). Validation studies of a multifactor cognitive-behavioral anger control inventory. *Journal of Personality Assessment*, 51(3), 417-432. doi: 10.1207/s15327752jpa5103_9
- Holmqvist, R., Hill, T. & Lang, A. (2009) Effects of Aggression Replacement training in young offender institutions. *International Journal of Offender Therapy and Comparative Criminology*, 53, 74-92. doi:10.1177/0306624X07310452
- Jones, Y. (1991) Aggression Replacement Training in a high school setting. *Australian Journal of Guidance and Counseling*, 1, 1-19..
- Kennedy, S. (1989) Anger management Training with adult prisoners. Unpublished doctoral dissertation, University of Ottawa, Canada.
- Kolko, D. J. (2006). Studying usual care in child and adolescent therapy: It's anything but routine. *Clinical Psychology: Science and Practice*, 13(1), 47-52. doi: 10.1111/j.1468-2850.2006.00005.x
- Koposov, R., Gundersen, K. & Svartdal, F (2014) Efficacy of Aggression Replacement Training among children from North-West Russia. *The International Journal of Emotional Education*, 6,14-24.
- Leeman, L. W., Gibbs, J. C., & Fuller, D. (1993). Evaluation of a multi component group treatment program for juvenile delinquents. *Aggressive Behavior*, 19, 281-292.
- Lehman, W. E. K., Greener, J. M., & Simpson, D. D. (2002). Assessing organizational readiness for change. *Journal of Substance Abuse Treatment*, 22(4), 197-209. doi: 10.1016/S0740-5472(02)00233-7
- Lochman, J. E., Boxmeyer, C., Powell, N., Qu, L., Wells, K., & Windle, M. (2009). Dissemination of the Coping Power program: Importance of intensity of counselor training. *Journal of Consulting and Clinical Psychology*, 77(3), 397 – 409. doi: 10.1037/a0014514
- Nodarse, M. (1998) The effects of Aggression Replacement Training on adolescents with an emotional handicap. *Dissertation Abstracts International*, May, 6424.
- Nugent, W. R., Bruley, C., & Allen, P. (1999). The effects of Aggression Replacement Training on male and female antisocial behavior in a runaway shelter. *Research on Social Work Practice*, 9(4), 466-482. doi: 10.1177/104973159900900404
- Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children—Second Edition manual*. Circle Pines, MN: American Guidance Service Publishing.
- Saldana, L., Chapman, J. E., Henggeler, S. W., & Rowland, M. D. (2007). Organizational readiness for change in adolescent programs: Criterion validity. *Journal of Substance Abuse Treatment*, 33(2), 159-169. doi: 10.1016/j.jsat.2006.12.029
- Smith, D. C., Furlong, M., Bates, M., & Laughlin, J. D. (1998). Development of the Multidimensional School Anger Inventory for males. *Psychology in the Schools*, 35(1), 1-15. doi: 10.1002/(SICI)1520-6807(199801)35:1<1::AID-PITS1>3.0.CO;2-U
- Weisz, J. R., Jensen, A. L., & McLeod, B. D. (2005). Development and dissemination of child and adolescent psychotherapies: Milestones, methods, and a new deployment-focused model. In E. D. Hibbs & P. S. Jensen (Eds.), *Psychosocial treatments for child and adolescent disorders: Empirically based strategies for clinical practice* (2nd ed., pp. 9-39). Washington, DC: American Psychological Association.

Appendix A: Sample Fidelity Coding Forms

Skillstreaming Fidelity Checklist

Facility: _____ Date: _____ Skillstreaming Session Number: _____

Length of Session: _____ Number of Youth Attending: _____ Number of Trainers: _____

Skillstreaming Skill: _____

	Circle Y or N	
<i>Homework Review</i>		
Homework practice from previous week's skill was reviewed	Y N	
Efforts to practice skills were reinforced by trainer(s)	Y N	
<i>Introduction of New Skill</i>		
Name of skill was introduced	Y N	
Explanation/definition of skill was given	Y N	
Youth involved in process of defining the skill	Y N	
<i>Modeling the Skill</i>		
Trainer(s) demonstrated the skill	Y N	
Behavioral steps of skill were made explicit	Y N	
Visual aids were used (skill cards, white board, etc.)	Y N	
<i>Establishing Youth Need for the Skill</i>		
Each youth identified specific situation in which the skill would be helpful to them	Y N	
<i>Role Plays</i>		
Each youth participated in role play as main actor	Y N	
Role plays were related to youth's individually-generated examples	Y N	
Trainers provided encouragement and/or instruction to youth during role play	Y N	
<i>Performance Feedback</i>		
Trainer(s) provided reinforcement (praise, approval, encouragement) for role plays that followed behavioral steps of skill	Y N	
All group members gave performance feedback	Y N	
Order of performance feedback given to role-playing youth was: 1. Co-actor 2. Group Members 3. Co-trainer 4. Trainer 5. Main Actor (preferred order)	Y N	
<i>Homework</i>		
Each participant was assigned to practice skill after role plays	Y N	
Youth asked to identify specifics (when, where, with whom) of plan to practice the skill	Y N	

Coder: _____ Score: _____
(out of 17)

Anger Control Fidelity Checklist

Facility: _____ Date: _____ Anger Control Session Number: _____

Length of Session: _____ Number of Youth Attending: _____ Number of Trainers: _____

Choose the appropriate list based on the session number you are coding

Session 1		Circle Y or N
<i>Introduction</i>		
Trainer introduced concept of controlling anger and aggressive behavior	Y N	
Trainer provided rationale for learning self-control	Y N	
<i>Rules & Procedures</i>		
Trainer outlined procedures and expectations for group including:		
a. Active participation required	Y N	
b. Show respect for other youth in the group	Y N	
c. Completion of homework assignments is required	Y N	
d. Material will be presented by trainer and practiced in role plays	Y N	
e. Conflict situations described in homework will be used for role plays in group	Y N	
<i>A-B-Cs of Anger</i>		
Trainer explained three steps of conflict situations:		
a. A = what triggered the problem	Y N	
b. B = what was the response (what you did)	Y N	
c. C = what were the consequences	Y N	
Trainer provided examples to illustrate three steps	Y N	
Trainer helped youth to identify A-B-C with their examples	Y N	
<i>Review</i>		
Trainer reviewed reasons for learning self control	Y N	
Trainer reviewed expectations/procedures	Y N	
Trainer reviewed A-B-Cs	Y N	

Coder: _____

Score: _____
(out of 15)

Moral Reasoning Fidelity Checklist

Facility: _____ Date: _____ Moral Reasoning Session Number: _____

Length of Session: _____ Number of Youth Attending: _____ Number of Trainers: _____

Problem Situation: _____

<i>Introduce Problem Situation</i>		Circle Y or N
Trainer reminded group of ground rules:		
1. No put downs or threats	Y	N
2. Listen to others	Y	N
3. If you criticize another group member, give that person a chance to answer	Y	N
4. Stay on subject if you disagree	Y	N
5. Never share what is said in group with others outside of the group	Y	N
Trainer checked to ensure group members understood problem situation	Y	N
Trainer involved/engaged youth to relate problem situation to their own lives	Y	N
<i>Cultivate Mature Morality</i>		
Trainer asked youth who gave positive responses to the problem about their decisions (highlighting mature morality)	Y	N
<i>Remediate Moral Developmental Delay</i>		
Trainer elicited explanations/reasons from group members who gave negative responses (e.g., hedonistic or pragmatic) to problem situation	Y	N
Reasons for negative decisions were recorded (white board, pad)	Y	N
Trainer elicited responses to negative (delayed) reasoning from more mature group members	Y	N
<i>Consolidate Mature Morality</i>		
Trainer worked with the group to "convert" those with immature reasoning to agree to the positive decisions (e.g., by asking, "Are there any strong objections if I circle this decision as the group's decision?")	Y	N
Trainer provided reinforcement (praise, encouragement) for group's positive decision-making	Y	N

Coder: _____

Score: _____
(out of 13)