

Aggression Replacement Training in Norway: Outcome evaluation of 11 Norwegian student projects

Knut Gundersen^{a*} and Frode Svartdal^{a,b}

^aRogaland College, Norway; ^bUniversity of Tromsø, Norway

Eleven groups of students performed a 24-session intervention based on Aggression Replacement Training (ART) as part of their further education programme. Subjects were 65 children and young people with varying degrees of behavioural problems. Forty-seven subjects received the ART programme. Eighteen received standard social and educational services and served as comparison subjects. Social problems and skills were assessed before and after the ART intervention using multi-informant instruments (SSRS, CADBI, HIT, CBCL). Informants in the ART group indicated significant improvement following the intervention, both in terms of increased social skills and reduced behavioural problems; in contrast, informants in the comparison group did not generally indicate improvement.

Keywords: Aggression replacement training; Anger control; Social competence; Social skills

Introduction

Children and adolescents with behaviour problems can create substantial difficulties for their families, the youths themselves, and society in general. Behaviour problems include less serious forms like quarrelling and not following orders or rules, as well as more serious forms such as violent behaviour against others, destruction of property, and truancy. Kazdin (1997) estimated that 10% of children and adolescents exhibit behaviour problems, of which 5% are the more serious forms. Norwegian studies have indicated a similar overall prevalence of behaviour problems, but the prevalence of serious behavioural problems is believed to be 1–2% (Ogden, 2002; Wichstrøm, Skogen, & Øya, 1996).

Behavioural problems are related to a variety of risk and causal factors, and successful prevention and treatment programmes should be expected reflect this complexity (Andrews, Zinger, & Hodge, 1990). Several studies have evaluated

*Corresponding author. Rogaland College, N-4365 Nærbø, Norway.
Email: Knut.Gundersen@rlvphs.no

single- and multi-treatment intervention designs, and there is now an increasing agreement on the factors and combinations of factors that are important in effective prevention and treatment. Interventions that target risk factors associated with behaviour problems (e.g., changing antisocial values and beliefs, reducing antisocial peer association, increasing self-control, promoting familial monitoring, and replacing skills of lying, stealing, and aggression with more pro-social alternatives) seem to be particularly promising (Andrews, 1995).

Among the intervention methods that address these factors, those with an origin in cognitive-behavioural approaches seem to be especially well documented (Dowden & Andrews, 2000; Garrett, 1985; Lipsey, 1992). These approaches include moral reasoning training (Arbuthnot & Gordon, 1986; Gibbs, Arnold, Ahlborn, & Cheesman, 1984), anger control training for adults (Novaco, 1975) and adolescent (Feindler, 1995), social skills training for adolescents and pre-adolescents (Goldstein, 1980), and aggressive children (Kettlewell & Kausch, 1983), and parent management training (Patterson, Reid, & Dishion, 1992). Thus, many of the programmes proven effective are, directly or indirectly, linked to social competence. Consistent with this, Gottfredson (1997) concluded that the most effective programmes are those that include “a range of social competency skills (e.g. developing self control, stress management, responsible decision making, social problem solving, and communication skills” (p.55). Further, Sørli (2000), concluded that the most promising method for treatment of behavioural problems is the training of a broad spectrum of social skills in relation to peers as well as adults, combined with the correction of problems in conduct.

At present there are several programmes in use in Norway aimed at the prevention and moderation of behavioural problems. The most ambitious enterprise in this respect is the nation-wide establishment of parent management training (PMT) (Patterson et al., 1992). Other programmes are Webster-Stratton’s programme for children (Webster-Stratton & Herbert, 1996), and multi-systemic therapy (MST) (Henggeler, Schoenwald, Bordun, Rowland, & Cunningham, 1998). Aggression replacement training (ART) (Goldstein, Glick, & Gibbs, 1998), is also gaining momentum in Norway as well as in other Scandinavian countries. In Norway, ART is approved for the prevention and treatment of behaviour problems in schools (Nordahl, Sørli, Tveit, & Manger, 2003) and homes for young people with behaviour problems (Andreassen, 2003), and is implemented in a number of schools and institutions (e.g., Lindoy Child and Youth Care Centre, Rogaland, several schools in Larvik, Vestfold, and Haa, Rogaland, and Glenne Autism Centre, Vestfold). A 2-year further education ART programme is currently offered at Rogaland College.

Although MST, PMT, and ART share many common features, one difference is that whereas MST and PMT are systems-oriented in their approach to the understanding and treatment of behavioural problems, ART is focused more on the development of individual social competencies. In the light of the studies reviewed, this may indicate that ART should be very promising in prevention and treatment of problem behaviour. A number of outcome studies especially in the US (Barnoski & Aos, 2004; Goldstein

& Glick, 1987; Goldstein, Glick, Carthan, & Blancero, 1994; Nugent, Bruley, & Winimaki, 1999) do indeed indicate good effect of ART interventions by increasing social skills and reducing behavioural problems (see also Andreassen, 2003).

Evaluation studies of the Norwegian PMT and MST programmes are currently being conducted (e.g., Ogden & Halliday-Boykins, 2004), and the outcomes seem to be positive and agree well with conclusions from previous American studies. As for ART, except for promising results obtained with enhanced ART with autistic children (Moynahan, 2003), no study has so far documented its efficacy when applied to Norwegian or Scandinavian children or young people. The primary goal of the present paper was, then, to determine whether positive effects observed in American studies could be obtained in a Norwegian sample. This goal is of interest for at least two reasons (Ogden & Halliday-Boykins, 2004). First, differences in the social and political climate between the US and Norway regarding the care of young people exhibiting behavioural problems may affect the efficacy of interventions. A young person who commits a criminal act in the US is typically subject to intervention by the juvenile justice authorities, and a likely reaction is punishment and dispositions designed to keep the community safe. In contrast, in Norway, measures for offenders under 15 years of age cannot be punitive or designed to protect the community (no one under 15 years of age can be imprisoned); instead, they are made in the best interests of the child. Second, since social services, healthcare, and education are more comprehensive and more universally accessible in Norway than in the US, the baselines for interventions are different between these countries. Thus, the expected gain of an intervention (be it ART, MST, or PMT) over the services already provided to youths in Norway is likely to be smaller compared to the corresponding gain seen for interventions in the US (Ogden & Halliday-Boykins, 2004).

In the present study, intervention effects were assessed using a between-groups design with repeated measures. ART subjects received 24 sessions of ART training; comparable comparison subjects received standard social and educational services. Before the ART intervention, a comprehensive battery of instruments was administered to assess behavioural problems and social skills. Following the intervention, the same test battery was repeated. The battery included the social skills rating system (SSRS) (Gresham & Elliot, 1990), Child and Adolescent Disruptive Behaviour Inventory 2.3 (CADBI) (Burns, Taylor, & Rusby, 2001a, b), How I Think (HIT) (Gibbs, Barriga, & Potter, 2001), and Child Behaviour Check List (CBCL) (Achenbach & Rescola, 2001). In addition, a custom-made self-report problem behaviour questionnaire was developed (Gundersen & Svartdal, 2003b). Since these instruments varied according to focus (behavioural problems, social skills, moral thinking) and informant source (parent, teacher, young person), they allowed for specific analyses across informants and foci.

Overall, the ART intervention as implemented in this study should generate improved social skills and moral reasoning, and reduced behaviour problems; no significant change should be observed in the comparison group. Since the ART

intervention here is implemented primarily in the school context, teachers' judgements might be expected to be the most reliable source for intervention outcome evaluation. But because ART interventions are designed to secure generalisation of skills by assigning homework, use of realistic life dilemmas, and so on (Goldstein & Martens, 2000), we expected that the ART intervention should generalise well, and hence that parents should indicate beneficial effects. However, it is known that teachers and parents do not always agree in their estimates of problem behaviour (Gomez, Burns, Walsh, & Moura, 2003; Achenbach, McConaughy, & Howell, 1987). This difference is most probably not simply an effect of differences in time spent with the young person, but more likely reflects the fact that parents and teachers observe children in different contexts (e.g., Messick, 1995). These considerations complicate the outcome predictions for this study, but the most optimistic is one implying positive effects by all informants in the ART group, and no change for informants in the comparison group.

Method

Subjects

Student participants. The ART trainers were 23 students (14 females, 9 males, mean age=38.1 years) participating in a 60-credit further education course in training of social competence at Rogaland College, Norway. All students had 3 years of college education as teachers or authorised social educators, and all were employed full-time at local schools and institutions. Prior to the ART intervention, all students had theoretical and practical training in ART and related topics corresponding to 30 credits of education. This included responsibility for 24 sessions of ART intervention with colleagues or students.

Children and youths. Subjects were 65 children and young people from schools and institutions in western and eastern Norway. Most of them were recruited from the ART students' workplace. There were 16 girls (mean age=14.1 years) and 49 boys (mean age=12.6 years). Potential participants were screened with a simple checklist (Gundersen & Svartdal, 2003a) to differentiate three general levels of behaviour problems:

- None: the young person may need training in pro-social skills, but has not developed significant behaviour problems;
- Mild: the young person displays some degree of behaviour problems;
- Some: the young person displays some degree of serious behaviour problems.

Consent and approval by ethics committee. Parents, teachers, and young people involved in the study received extended written and oral information about the investigation. The information was presented by the ART students in meetings with

groups and/or individually. Participants signed to show their consent. The regional ethics committee approved the protocol for the project.

ART Intervention

ART (Goldstein et al., 1998) focuses on the development of social skills by the individual. More specifically, ART consists of three distinct components addressing training of social skills, anger control, and moral reasoning. Anger control training (the affective or emotional component) entails youths being trained to recognise their external and internal triggers for aggression, aggression signals, and to control anger using various techniques. Social skills training (the action component) focuses on training a selection of 50 structured skills. Some are quite simple (e.g., listening to someone else, starting a conversation), others more complex (e.g., avoiding disruption, handling group pressure). In moral reasoning training (the thought and values component), participants learn to reason in a more advanced manner in regard to challenging ethical and moral dilemmas, and to handle situations in their own lives in line with their own moral and ethical values. Even if the three components are trained separately, elements of each are partly integrated in the others (e.g., social skills as part of anger control training).

Training takes place in groups with 5 to 8 participants. Groups are composed with regard to age, similarity of behavioural challenges, and friendship between participants. Elliot and Gresham (1991) also recommend the inclusion of group members with a higher level of social competence as positive role models. Two trainers conduct ART sessions. Rules and consequences for infractions are clearly defined. Participation is voluntary and the use of positive reinforcement and small games are highly recommended to secure the motivation of trainees. There is a firm structure in the programme including defining the theme of the session, demonstration, role-playing, questioning where and when to use the skill, feedback/evaluation, and homework. The purpose of feedback to the youths is primarily to shape proper behaviour. All participants have different observation tasks. Goldstein et al. (1998) recommend that all three components (skill streaming, anger control training, and moral reasoning training) are scheduled for training at least once per week over a period of 10 weeks. In order to transfer and maintain skills, it is important to establish contact with important individuals (family members, teachers, club leaders, etc.) in the participants' environments.

The ART group received a standard 24 hours of ART intervention with at least 4 sessions from each component of skills training, anger training, and moral reasoning training. The duration of the intervention period was 10 weeks. For practical reasons, the number of sessions was reduced from the recommended number of 30 (Goldstein et al., 1998). On average, the interventions had 10.8 sessions of social skills training, 8.4 sessions of anger control training, and 4.8 sessions of moral reasoning.

Treatment fidelity. Substantial information and supervision was made available to the ART students before and during the implementation of the ART interventions.

A web page with detailed instructions was available at all times. To ensure sufficient trainer quality, each student was required to complete at least 24 sessions as ART trainer before the ART intervention. During the intervention, supervision (either from videotapes or in vivo) was provided. The students also evaluated each session during the intervention, and the evaluation was subject to approval from the college. Following the intervention, students filled out a questionnaire to identify to what degree they had implemented ART according to treatment fidelity factors described by Hollin (1995). Finally, overall ART competence was evaluated by colleagues and superiors in a questionnaire prior to and after training. The outcome of this evaluation indicated significant improvement. Overall, we believe that extensive prior training and continued supervision during the intervention secured good treatment fidelity (Hollin, 1995).

Comparison Group

The comparison group received the standard child social and educational services as defined and delivered by the actual school/institution. In some cases this meant no special arrangements. For five of the sub-studies, the comparison group was recruited from the same class or school as the ART group. This represents a potential threat to the design of this study, since the intervention (which was distributed over 13 weeks) could be diluted in that the comparison groups were affected by the ART programme and/or changed behaviour in the ART groups. However, this limitation was occasioned by practical circumstances and could not be remedied.

Instruments

Four instruments were selected to assess behavioural problems and social skills, using multiple informants (young person, parent, teacher). The instruments were mainly chosen among internationally validated instruments. Two were available in Norwegian versions (CBCL, SSRS), and two were translated as part of this study (CADBI, HIT). In addition, a custom-made self-report problem behaviour checklist was included.

CBCL. Aseba School-Age Forms & Profiles (CBCL) (Achenbach & Rescola, 2001) comprises an integrated system of multi-informant assessment consisting of the Child Behaviour Checklist for Ages 6–18 (parents), Teacher's Report form, and Youth Self-Report. CBCL is based on practical experiences and research over 40 years, and is now regarded as one of the best validated tools for assessing behavioural functioning in children and youth. CBCL is translated into Norwegian, and studies have demonstrated acceptable validity (Noevik, 1999).

In the present study, raw scores (rather than T-scores) were used. The internal consistency of the subscales included in the analyses was very high; for each informant, α was =.90 (parents), .86 (teachers), and .93 (young people).

CADBI. The Child and Adolescent Disruptive Behaviour Inventory 2.3 (*CADBI*) has a parent and teacher version (Burns et al., 2001a, b). Both versions contain assertions relating to various forms of behaviours and behaviour problems towards adults (parents, teachers) and peers, activity level, and attention skills (e.g., “Starts physical fights with peers”; “Does not seem to show concern for the feelings of others”). Each of the assertions is scored on an 8-point frequency of occurrence scale (never in past month, 1–2 times in past month, 3–4 times in past month, 2–6 times per week, 1 time per day, 2–5 times per day, 6–9 times per day, and 10 or more times per day). In part 6, parents and teachers rate the child’s academic competence on 11 items. These items are rated on a 7-point scale (from severe difficulty to excellent performance).

CADBI has achieved good internal consistency 3-month test–retest reliability (Fitzgerald, 2002; Iredale, 2000; Skansgaard & Burns, 1998), but the validity of version 2.3 is not well documented. For the present investigation, *CADBI* was translated into Norwegian (Moynahan & Stroemgren, 2003). The internal consistency of the subscales included in the analyses was very high ($\alpha=.94$ for the parent scores, and .90 for the teacher scores).

SSRS. Social skills rating system (*SSRS*) (Gershaw & Elliot, 1990), Norwegian version (Ogden, 2003) is a standardised norm-referenced scale measuring children and adolescents’ social skills, problem behaviour, and academic competence. The validity of *SSRS* has been demonstrated in several studies (Demary et al., 1995; Gresham & Elliot, 1990; Ogden, 2003). It was chosen in this study because of its obvious relevance to measurement of social skills (since we included two other instruments to measure problem behaviour, the problem scale of *SSRS* was not used). The *SSRS* items group into four subscales, co-operation, assertion, self-control, and responsibility. Informants (teacher, parent, and young person) rate 30 statements about social behaviour in terms of frequency. For example, co-operation includes behaviours such as helping others, sharing materials, and responding to the actions of others (e.g. “helps with tasks”, “completes task on time”). The rating scale for each item was originally a 3-point scale, but Ogden (2003) increased this to four (1=never, 2=sometimes, 3=often, 4=very often), which was used in the present study.

In the present sample, the internal consistency for the *SSRS* subscales was high ($\alpha=.92$ for the parent scores, and .79 for the teacher scores).

HIT. How I Think (Gibbs et al., 2001) is a questionnaire designed to measure self-serving cognitive distortions in antisocial youths. It contains 54 items, and 39 cognitive distortion items (in addition, there are 8 anomalous responding (AR) items and 7 positive filler items). These are divided into four cognitive distortion subscales: self-centred, blaming others, minimising/mislabelling, and assuming the worst. The subject’s response to each item statement is marked on a 6-point Likert-type scale

from “disagree strongly” (1) to “agree strongly” (6). The 8 AR items are designed to screen for inaccurate or suspect responding. The current version has been evaluated on four validation samples and exhibited generally good validity on every validity measure (Barriga, Gibbs, Potter, & Liau, 2001). HIT also exhibits high test–retest reliability, good internal consistency, and generally good construct validity (Barriga & Gibbs, 1996). High correlations have been reported between HIT and the self-reported externalising scale on the CBCL (Achenbach, 1991) after controlling for internalising disorders.

For the present study, overall HIT was used for all statistical analyses. This index includes all subscales of HIT. In the present sample, the internal consistency for these subscales was high ($\alpha=.90$).

Self-Report Questionnaire. Since CADBI has teacher and parent versions, but no youth version, we developed a questionnaire with 56 statements to be filled out by young people. Several of the items were adapted from a Norwegian MST study (Ogden & Halliday-Boykins, 2004). The items were divided into 3 subscales: pro-social skills (19) and problem behaviour (36). The response to the pro-social statements is marked along a 5-point scale from “does not suit” to “suit very well”. Responses to the problem behaviour statements are marked on a 9-point scale according to how many times the actual behaviour has occurred in the last month.

Procedure

The student ART trainers were divided into 11 sub-groups, normally 2 students per sub-study, constituting an ART main trainer and a co-trainer. The formation of these groups was determined primarily by practical considerations (e.g., students’ workplace). Four of the sub-groups performed their ART intervention at the workplace of both participating students. Five sub-groups performed the intervention at the workplace of one of the participating students, and the remaining two sub-groups performed their ART intervention at a location different from their own workplace.

Following screening of all potential participants, each sub-study established two client groups representing comparable behaviour problems. The group composition was performed according to the recommendations by Goldstein (1998). Each group ($n=6$) was required to include at least one young person classified at level 3 (most behavioural problems). Since sex of the young person was not of importance here, the sex factor was not manipulated. One of the groups was then, based on a randomised procedure, allocated to the ART condition, and the other to the comparison group. The allocation procedure—which was performed on groups rather than on individual participants—was carried out by a superior of the student trainers according to directions provided by the project leaders. For some of the sub-studies, however, randomisation was not possible. In one case, miscommunication to the teacher necessitated that both groups at a school were included in the ART group. In three sub-studies, training took part in special schools with small units

(youths with Asperger Syndrome; psychiatric problems; comprehensive behaviour problems), and there were not sufficient young people available to establish two comparable groups. In one sub-study, three of the comparison group members withdrew from the project. We performed complete analyses on the total sample as well as the sub-sample that satisfied the criteria for the randomisation procedure. Since these analyses rendered almost identical outcomes, we decided to include all subjects in the analyses reported here.

Thus, a total of 47 youths received the ART intervention and 18 youths were included in the comparison group. The unequal group size was unfortunate, but we reasoned that would not be critical to the study. An important consideration here is that allocation to groups was not influenced by the youths, teachers, or parents, implying that selection or self-selection was not an issue. Another consideration is that the overall study would compare relatively few participants in heterogeneous sub-groups, implying that the possibility of statistically sound group comparisons (ART vs. comparison group) would be limited. Thus, within-group comparisons (pre vs. post) would probably be the preferred criterion for measurement of change. The latter is supported by the fact that the study involved a large number of repeated measures, thus increasing statistical power.

Six of the sub-studies took part in normal schools, two in special schools for adolescents with behaviour problems, one in an institution for adolescents with behaviour problems, one in a psychiatric child clinic, and one in a specialised group for pupils with Asperger Syndrome (7 subjects). Table 1 shows the actual distribution of young person in the three behaviour problem classifications and over types of institutions.

Design and Statistical Analyses

The overall design was a mixed design with treatment (ART vs. no treatment) as the group factor and pretest–test as the repeated measures factor. To limit the workload related to the assessment procedures, the three participants with most behaviour problems in each group were assessed. General Linear Model (GML) within Statistica (StatSoft) was used to compare differences in scores on the individual instruments between pre-test and test, and between the ART and comparison

Table 1. Number of young people in the three problem behaviour classifications and type of institution

	N	Problem level	Type of institution (distribution of problem levels 3, 2, and 1 in parentheses)
ART	47	3: N=18 (38.3%) 2: N=21 (44.7%) 1: N=8 (17.0%)	School: N=26 (8, 12, 6) Special school: N=10 (6, 2, 2) Institution: N=11 (4, 7, 0)
COMP	18	3: N=5 (27.8%) 2: N=10 (56.0%) 1: N=3 (16.7%)	School: N=15 (4, 8, 3) Special School: N=0 (–) Institution: N=3 (1, 2, 0)

groups. As noted, within-group changes from pre-test to test were the primary treatment measure. Specific predictions, that is, tests of change between pre-test and test levels, were performed by contrast analysis (Rosenthal & Rosnow, 1985). In addition, we ran analyses to check for moderator variables (e.g., sex, age, and institution). These analyses are reported separately.

Results and Discussion

The overall expectation for the present data is that participants in the ART condition should demonstrate improved social skills and reduced occurrence of problem behaviour, whereas little or no change should be seen in the comparison group. In the following presentation, this expectation is examined for each of the three areas of focus of the present study, social skills (SSRS), moral thinking (HIT), and problem behaviour (CADBI; CBCL). Since intervention effects are the main focus of this article, results pertaining to this issue are presented in the text; a summary of results is presented in Tables 2 and 3.

Social Skills: SSRS

Parents. For the parent data, scores on all the four subscales of SSRS, cooperation, assertion, responsibility, and self-control were averaged and subjected to analyses. An overall ANOVA of the Group and pre-test test data showed no overall significant effects of the Group or pre-test factors, but a significant Group x pre-test-test interaction, $F(1, 44)=5.33$, $p=.026$. Contrast analyses demonstrated that the increase from pre-test to test was significant within the ART group, $F(1, 44)=10.97$, $p=.002$, whereas the change in the comparison group was nonsignificant, $F(1, 44)=0.42$. These results indicate that the ART and comparison groups changed differently from pre-test to test, with a significant increase in social skills in the ART group but no change in the comparison group.

Teachers. The teacher checklist includes three subscales, cooperation, assertion, and self-control, which were averaged for the present analysis. The ANOVA showed a significant effect of the pre-test test factor, $F(1, 51)=8.26$, $p<.001$, but nonsignificant effects of group and of the pre-test test x group interaction. Contrast analyses showed that the change from pre-test to test within the ART group was significant, $F(1, 51)=11.07$, $p=.001$, whereas the corresponding change within the comparison group was not significant, $F(1, 51)=1.70$, $p=.20$. Thus, the ART group demonstrated a significant increase in SSRS scores, but the comparison group demonstrated a nonsignificant change.

Children/young people. The children/youth version of SSRS includes 34 (children) and 39 (youth) questions related to subjective evaluations of social skills. The

Table 2. Social skills. Means and within-groups effects (pretest vs. test) for the ART and comparison groups for the SSRS and HIT instruments (SE in parentheses)

	SSRS		HIT	
	Pre	Post	Pre	Post
	F		F	
ART	Par		Par	
	25.0 (0.71)		–	
	26.2 (0.67)		–	
	10.97*		–	
	Tea		Tea	
	22.5 (0.57)		–	
	24.1 (0.69)		–	
	11.07*		–	
	You		You	
	101.3 (3.38)		3.03 (0.13)	
	105.9 (3.20)		2.67 (0.14)	
	1.89		9.58*	
Comparison	Par		Par	
	27.1 (1.13)		–	
	26.7 (1.07)		–	
	.42		–	
	Tea		Tea	
	22.4 (0.92)		–	
	23.4 (1.09)		–	
	1.70		–	
	You		You	
	103.4 (5.22)		2.93 (0.18)	
	109.8 (4.94)		2.49 (0.19)	
	1.55		7.35*	

* $p < .01$. Note that the HIT measure only applies to young people. Also note that subscale scores are provided because different numbers of items are included in each.

ANOVA did not indicate significant effects of group, pre-test test, or the group x pre-test test interaction. Moreover, the contrast analyses indicated a nonsignificant change from pre-test to test within the ART group, $F(1, 42) = 1.89$, $p = .176$, and also a nonsignificant change in the comparison group, $F(1, 42) = 1.55$, $p = .22$. We also obtained a measure of subjective change in social skills by means of our custom-made self report form (Gundersen & Svartdal, 2003). The form includes 19 questions related to social skills. The ANOVA showed a significant effect of the pre-test test factor, $F(1, 52) = 7.37$, $p = .009$, as the skills scores increased from pre-test to test. Contrast analyses demonstrated that this change was significant within the ART group, $F(1, 52) = 8.22$, $p = .006$, but nonsignificant within the comparison group, $F(1, 52) = 1.89$, $p = .17$.

Table 3. Social problems. Means and within-groups effects (pretest vs. test) for the ART and comparison groups for the CADBI and CBCL instruments (SE in parentheses)

	CADBI		CBCL	
	Pre	Post	Pre	Post
	F		F	
ART	Par		Par	
		156.5 (9.80)		18.4 (1.60)
		130.2 (9.76)		13.8 (1.64)
		18.03*		27.59*
	Tea		Tea	
		143.9 (10.38)		17.0 (1.42)
		124.0 (8.92)		12.1 (1.73)
		8.11*		13.49*
	You		You	
	72.0 (8.59)		17.6 (1.66)	
	51.6 (5.77)		14.2 (1.64)	
	5.97*		14.81*	
Comparison	Par		Par	
		98.9 (14.97)		10.4 (2.52)
		93.3 (14.91)		9.6 (2.57)
		0.36		.29
	Tea		Tea	
		134.0 (16.54)		13.2 (2.36)
		117.7 (14.21)		11.8 (2.89)
		2.13		.39
	You		You	
		50.0 (12.36)		17.0 (2.32)
		51.6 (8.30)		14.1 (2.29)
		.02		5.45*

* $p < .01$. Note that the youth measures in the CADBI column are from Gundersen and Svartdal (2003)

Moral Reasoning: HIT

For this analysis, the subscale overall HIT was used. The ANOVA indicated an overall effect of test–retest, $F(1, 54) = 16.06$, $p = .00019$, but no overall group difference or group \times test–retest interaction. Contrast analysis showed that the change within the ART group was significant, $F(1, 54) = 9.58$, $p = .004$, as was the change within the comparison group, $F(1, 54) = 7.35$, $p = .008$. These results indicate that moral reasoning, as measured by HIT, improved significantly in both groups. These results were not affected by removing subjects with true AR score > 4.0 .

Problem Behaviour: CADBI

Parents. The first 5 CADBI subscales were summed for the present analyses. The ANOVA indicated a significant effect of group, $F(1, 38) = 7.78$, $p = .008$, of the

pretest–test factor, $F(1, 38)=8.00, p=.007$, as well as a marginal pre-test test x group interaction, $F(1, 38)=3.27, p=.08$. Contrast analyses within the groups showed that the ART group reduced their scores significantly from pre-test to test, $F(1, 38)=18.03, p<.001$, whereas the corresponding reduction within the comparison group was nonsignificant.

Teachers. The first 5 subscales of CADBI were summed for this analysis. The ANOVA indicated a significant effect of the test–pre-test factor, $F(1, 44)=7.57, p=.008$, reflecting a reduction in scores from pre-test to test in both groups. The overall group difference and the pre-test test x group interaction was not significant. Contrast analyses showed that the reduction from pre-test to test was significant within the ART group, $F(1, 41)=8.11, p=.006$, whereas the change within the comparison group was not significant, $F(1, 44)=2.13, p>.15$.

Children/young people. Since CADBI has no children/youth self-report scale, we obtained a measure of subjective problems by our custom-made self-report form (Gundersen & Svartdal, 2003). The form includes 20 questions related to behaviour problems. The ANOVA did not indicate significant effects of group, pretest–test, or the group x pretest–test interaction. However, contrast analyses indicated a significant reduction of problems from pretest to test within the ART group, $F(1, 41)=5.97, p=.019$, but no significant change in the comparison group.

Problem Behaviour: CBCL

Parents. The subscales INT, EXT, and OTHER of CBCL were included in the analysis. The ANOVA showed a significant effect of the group manipulation, $F(1, 43)=4.45, p<.05$, of the pre-test test factor, $F(1, 43)=10.75, p=.002$, as well as for the pre-test test x group interaction, $F(1, 43)=5.61, p=.02$. Contrast analyses demonstrated that the reduction from pre-test to test within the ART group was significant, $F(1, 43)=27.59, p<.001$, whereas there was no significant change in the comparison group, $F(1, 43)=0.29, p>.50$. This indicates that the ART group demonstrated reduced scores on the subscales following the ART intervention; no change was seen in the comparison group.

Teachers. As for the parents' data, the subscales INT, EXT, and OTHER of CBCL were included in the analysis. The ANOVA indicated a significant effect of the pre-test–test manipulation, $F(1, 47)=5.88, p=.019$, reflecting an overall reduction in scores on the three subscales from pre-test to test. The group and the pre-test test effects were not significant. Contrast analyses showed that the reduction within the ART group was significant, $F(1, 47)=13.49, p<.001$, whereas there was no significant change in the comparison group, $F(1, 47)=0.39$.

Children/young people. Again, the subscales INT, EXT, and OTHER of CBCL were included in the analysis. The ANOVA indicated a significant effect of the pre-test test factor, $F(1, 54)=17.11, p=.00012$, reflecting the fact that subscore means were lower on test compared to pre-test. The pre-test test and pre-test test x group interaction were not significant. Contrast analyses showed that the reduction within the ART group was significant, $F(1, 54)=14.81, p<.001$, as was the corresponding reduction in the comparison group, $F(1, 54)=5.43, p=.02$.

Effects of Moderator Variables

We also performed analyses to assess the influence of variables that were not manipulated or included in the design. As noted, due to the requirements of group composition, the sex factor was not manipulated in our design. Seventeen of the subjects were females, nine of which were in the ART condition. Due to the low number of females, we decided not to include sex in post hoc analyses.

Age level. Our subjects spanned from 11 to 17 years of age. Overall, five of seven tests with age level as a covariate indicated that age was not significantly related to outcome. The exceptions were CBCL (Parent scale) and CADBI (Teacher scale). Consistent with existing research (Wilson et al., 2003), the overall conclusions were not affected by these outcomes.

Institution. We ran separate analyses with institution (place of intervention) as a covariate. In none of the tests did this covariate indicate that institution has a significant effect on outcome.

General Discussion

The primary finding of this report was that a 24-hour ART intervention produced a positive and reliable outcome, in that social skills were improved and behavioural problems reduced from pretest to test. Thus, the ART group demonstrated significant improvement on 9 out of 10 tests; the comparison group demonstrated improvement on 2 out of 10 tests. To our knowledge, this is the first demonstration in Scandinavia of the positive effect of ART in treatment of behavioural problems. Our results conform to corresponding US studies (Nugent et al., 1999) and indicate a positive effect of an ART intervention despite cultural differences between the US and Norway and also despite differences in baselines for intervention.

It is noteworthy that this positive outcome was achieved despite the fact that the interventions were implemented by students with relatively limited experience as ART trainers, and interventions also had a limited duration (24 hours).

The explanation of this is most probably that the student ART trainers had extensive theoretical and practical training before the ART intervention commenced, and also that the projects were supervised throughout the intervention phase.

Although the intervention was carried out in school/institution settings, positive effects were obvious also in home settings as evaluated by parents. Thus, as expected, the effects of ART generalised to home settings. The ART programme strongly emphasises methods to enhance transformation and maintenance of skills, and these were also part of the present projects (e.g., homework, hassle log, stimulus variation in role playing, training and reflection of real-life situations). Although even better results might have been expected if other factors specifically aimed at enhancing generalisation, e.g., incorporating peers as functional mediators in the group (Osnes & Lieblein, 2002) and separate training of parents (Calame & Parker, 2003) had been included in the intervention, our results clearly indicate good generalisation of effect. It is also noteworthy that the four of the five self-ratings provided by the youths indicated beneficial effects from the ART intervention. This indicates a cross-situational effect confirming the generalised intervention effect. In this context it should also be noted that the result from the HIT measure was positive, despite the fact that the intervention only contained, on average, 4–5 sessions with moral reasoning training (i.e., half the training spent on skills training and aggression management training). This indicates that training within any one of the three elements of ART (anger control, pro-social skills, and moral reasoning) also generalise to the others (Krevans & Gibbs, 1996).

Our results indicate much the same pattern of positive outcome regardless of informant (parent, teacher, and youth). Previous research has demonstrated that parents and teachers demonstrate differences in their ratings of problem behaviour (Gomez et al., 2003; Achenbach et al., 1987), most likely because parents and teachers observe the young people in different settings. Our data show that parents judge the effects of the ART intervention in much the same manner as the teachers do. It should be noted, however, that our analyses do not specifically address multi-informant agreements in judgements.

The use in this study of a comprehensive test battery to measure changes in problem behaviour and social skills also demonstrated a high degree of consistency in results between tests. The tests used all seemed to reflect the positive outcome of the ART intervention. Again, as for informants, we did not measure agreement among the instruments specifically, but we see no reason to assume that any one of the instruments was more sensitive to treatment effects.

It should also be noted that two out of three subjective measures of this study (CBCL, Y scale, and HIT), indicated positive outcome following the ART intervention, but this outcome also occurred in the control group. This result should be seen in the light of the fact that several of the interventions took place under circumstances that made adequate separation of the ART and comparison subjects quite difficult (i.e., a possible dilution between conditions).

This effect did not, however, show up in teachers' judgements, which most probably would have been the case if dilution between conditions had been a serious problem in this study. The fact that the youths themselves indicated similar improvement irrespective of condition may, however, indicate that some dilution occurred.

Although the overall outcome of the interventions analysed here was positive, some qualifications must be pointed out. First, our total sample was relatively small, and the randomisation procedure did not succeed in establishing proper treatment and control groups. Our conclusion is therefore not based on a proper randomised control group design, but primarily on within-groups comparisons before and after treatment. Although proper control group comparisons definitely would have been preferable, it should be noted that randomisation of participants is complicated by the nature of the ART intervention. Rather than focusing on individuals, the ART intervention focuses on a group, and this group should be composed in a specific manner (Goldstein et al., 1998). As a consequence, randomisation must be performed at group level, and this procedure is, as we have experienced, quite vulnerable. Future research could compensate for this situation by including larger samples selected from larger pools of potential participants.

A second issue is the role of moderator variables on outcome of intervention. Consistent with earlier research (Wilson et al., 2003), age level did not modulate the effect of intervention. Further, we could not detect systematic variations in outcome as a function of institution (place of intervention). From one perspective this is not surprising. The ART trainers were students educated in the same class, they acted as trainers in pairs, and received common instructions and supervision. On the other hand, it is well known that a variety of moderator variables do affect intervention outcome, and trainer quality is one of a number of relevant variables here (e.g., Barnoski & Aos, 2004; Hollin, 1995). Among other relevant factors are the frame conditions for implementation (e.g., support in organisation, group size, peers as role models). Analyses of such factors should attempt to quantify the quality of these variables to assess their effect on outcome.

Finally, the positive outcome of this study must be interpreted in light of the fact that our ART trainers, being highly selected and recruited from all over Norway, most probably were highly motivated and particularly well skilled. Wilson et al. (2003) also found that routine programmes showed smaller effects than those introduced and supervised by researchers. Thus, it would be a sound recommendation that future research evaluated ART interventions that were not performed as "pioneer" projects. Future research on the effect of ART interventions should also plan for systematic follow-up studies. We suggest that such follow-up studies should include subject age levels that secure the participation of all informants in all phases of the investigation. In Norway, a sound follow-up investigation should take into consideration the fact that most young people change schools between the ages of 12 and 14.

References

- Achenbach, T. M. (1991). *Manual for the youth self-report report and 1991 profile*. Burlington, VT: University of Vermont.
- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioural and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213–232.
- Achenbach, T. M., & Rescola, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington, VT: University of Vermont.
- Andreassen, T. (2003). *Behandling av ungdom i institusjoner Hva sier forskningen* [Treatment of adolescents in institutions. What does research tell us?; in Norwegian]. Oslo, Norway: Kommuneforlaget.
- Andrews, D. A. (1995). The psychology of criminal conduct and effective treatment. In J. McGuire (Ed.), *What works: Reducing reoffending*. Chichester, England: Wiley.
- Andrews, D. A., Zinger, I., & Hodge, R. D. (1990). Meta-analysis of treatment programmes for young offenders: The effect of clinically relevant treatment. *Rediscovering Psychology. Criminal Justice and Behaviour*, *17*(1), 19–52.
- Arbuthnot, J., & Gordon, D. A. (1986). Behavioural and cognitive effects of a moral reasoning development intervention for high-risk behaviour-disordered adolescents. *Journal of Consulting and Clinical Psychology*, *54*, 208–216.
- Barnoski, R., & Aos, S. (2004). Outcome evaluations of Washington State's research-based programmes for juvenile offenders. Washington: 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 Behaviour*, *22*, 333–343.
- Barriga, A. Q., Gibbs, J. C., Potter, G. B., & Liau, A. K. (2001). *How I Think (HIT) questionnaire manual*. Champaign, IL: Research Press.
- Burns, G. L., Taylor, T. K., & Rusby, J. C. (2001a). *Child and adolescent disruptive behaviour inventory 2.3: Parent version*. Pullman, WA: Washington State University.
- Burns, G. L., Taylor, T. K., & Rusby, J. C. (2001b). *Child and adolescent disruptive behaviour inventory 2.3: Teachers version*. Pullman, WA: Washington State University.
- Calame, R., & Parker, K. (2003). Reclaiming youth and families with "Family ART". *Reclaiming Children and Youth*, *12*(3), 154–157.
- Demary, E., Maray, M. K., Ruffalo, S. L., Carlson, J., Busse, R. T., Olson, A. E., et al. (1995). Social skills assessment: A comparative evaluation of six published rating scales. *School Psychology Review*, *4*, 648–671.
- Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent reoffending; A metaanalysis. *Canadian Journal of Criminology*, October, *42*(4), 449–467.
- Elliot, S. N., & Gresham, F. M. (1991). *Social skills intervention guide: Practical strategies for social skills training*. Circle Pines, MN: American Guidance.
- Feindler, E. L. (1995). An ideal treatment package for children and adolescents with anger disorders. In H. Kassinove (Ed.), *Anger disorders: Definitions, diagnosis, and treatment*. London: Taylor & Francis.
- Fitzgerald, J. (2002). *Internal validity of constructs within the conduct problem behaviour domain*. Unpublished manuscript.
- Garrett, C. J. (1985). Effects of residential treatment on adjudicated delinquents: A meta-analysis. *Journal of Research in Crime and Delinquency*, *22*(4), 287–309.
- Gershaw, N. J., & Elliot, S. N. (1990). *Social skills rating system*. Circle Pines, MN: American Guidance Service.
- Gibbs, J. C., Arnold, K. D., Ahlborn, H. H., & Cheesman, F. L. (1984). Facilitation of sociomoral reasoning in delinquents. *Journal of Consulting and Clinical Psychology*, *52*, 37–45.

- Gibbs, J. C., Barriga, A. Q., & Potter, G. B. (2001). *How I think (HIT) questionnaire*. Champaign IL: Research Press.
- Goldstein, A. P. (1980). *Skill-streaming the adolescent: A structured learning approach to teaching prosocial skills*. Champaign, IL: Research Press.
- Goldstein, A. P., & Glick, B. (1987). *Aggression replacement training: A comprehensive intervention for aggressive youth*. Champaign, IL: Research Press.
- Goldstein, A. P., Glick, B., Carthan, W., & 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*. Champaign, IL: Research Press.
- Goldstein, A. P., & Martens, B. (2000). *Lasting change: Methods for enhancing generalization of gain*. Champaign, IL: Research Press.
- Gomez, R., Burns, G. L., Walsh, J. A., & de Moura, M. A. (2003). A multitrait-multisource confirmatory factor analytic approach to the construct validity of ADHD rating scales. *Psychological Assessment*, 15, 3–16.
- Gottfredson, D. (1997). School-based crime prevention. In L. W. Sherman, et al., *Preventing crime: What works, what doesn't, what's promising: A report to the United States Congress*. Washington DC: Department of Justice.
- Gresham, F. M., & Elliot, S. N. (1990). *Social skills rating system*. Circle Pines, MN: American Guidance Service.
- Gundersen, K., & Svartdal, F. (2003a). *Presjekk, en grovsortering for klassifisering av problematferd*. [Precheck, a preliminary classification of problem behaviour; in Norwegian]. Unpublished manuscript.
- Gundersen, K., & Svartdal, F. (2003b). *Selvrapporteringsskjema* [Self-report problem behaviour questionnaire; in Norwegian]. Nærbø, Norway: Rogaland Høgskole.
- Henggeler, S. W., Schoenwald, S. K., Bordun, C. M., Rowland, M. D., & Cunningham, P. B. (1998). *Multisystemic treatment of antisocial behaviour in children and adolescents*. New York: Guildford.
- Hollin, C. R. (1995). The meaning and implications of “program integrity”. In J. McGuire (Ed.), *What works: Reducing reoffending-guidelines from research and practice* (pp. 195–208). Chichester, England: Wiley.
- Iredale, T. (2000). *Internal validity of the DSM-IV oppositional defiant disorder, attention-deficit/hyperactivity disorder, and conduct behaviour problem symptoms*. Unpublished manuscript.
- Kazdin, A. E. (1997). Practitioner review: psychosocial treatments for conduct disorder in children. *The Journal of Child Psychology and Psychiatry*, 38, 161–178.
- Kettlewell, P. W., & Kausch, D. F. (1983). The generalization of the effects of a cognitive-behavioural treatment program for aggressive children. *Journal of Abnormal Child Psychology*, 11, 101–114.
- Krevans, J., & Gibbs, J. C. (1996). Parents' use of inductive discipline: relations to children's empathy and prosocial behaviour. *Child Development*, 67, 3263–3277.
- Lipsey, M. W. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In F. T. Cullen & B. K. Appelgate (Eds.), *Offender rehabilitation, effective correctional intervention*. Aldershot, England: Dartmouth.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749.
- Moynahan, L. (2003). Enhanced aggression replacement training with children and youth with autism spectrum disorder. *Reclaiming Children and Youth*, 12(3), 174–180.
- Moynahan, L., & Stroemgren, B. (2003). *Cadbi foreldre og Cadbi lærer*. Horten, Norway: Glenne Senter. Unpublished manuscript.

- Noevik, T. S. (1999). *Validity and use of the child behaviour checklist in Norwegian children and adolescents. An epidemiological and clinical study*. Oslo, Norway: University of Oslo, Centre for Child and Adolescent Psychiatry.
- Nordahl, T., Sørli, M. A., Tveit, A., & Manger, T. (2003). *Alvorlige atferdsvansker. Effektiv forebygging og mestring i skolen* [Serious behaviour problems: Effective prevention and mastering in school; in Norwegian]. Oslo, Norway: Læringscenteret.
- Novaco, R. W. (1975). *Anger control: The development and evaluation of an experimental treatment*. Lexington, DC: Heath & Co.
- Nugent, W. R., Bruley, C., & Winimaki, L. (1999). The effects of aggression replacement training on male and female antisocial behaviour. *Research on Social Work Practice*, 9(4), 466–482.
- Ogden, T. (2002). *Children and adolescents' life conditions in the Nordic countries. Marginalisation of children and adolescents in the Nordic countries—an overview of current knowledge*. Copenhagen, Denmark: Nordic Council of Ministers.
- Ogden, T. (2003). The validity of teacher ratings of adolescents' social skills. *Scandinavian Journal of Educational Research*, 47, 63–75.
- Ogden, T., & Halliday-Boykins, C. A. (2004). Multisystemic treatment of antisocial adolescents in Norway: Replication of clinical outcomes of the U.S. *Journal of Child and Adolescent Mental Health*, 9, 77–83.
- Osnes, P. G., & Lieblein, T. (2002). An explicit technology of generalization. *The Behaviour Analyst Today*, 3, 364–374.
- Patterson, G. R., Reid, J., & Dishion, T. J. (1992). *Antisocial boys*. Eugene, OR: Castalia.
- Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis. Focused comparisons in the analysis of variance*. Cambridge, England: Cambridge University Press.
- Skansgaard, E. P., & Burns, G. L. (1998). Comparison of DSM-IV ADHD combined type and predominantly inattentive types: Correspondence between ratings and direct observations of inattention hyperactivity/impulsivity, slow cognitive tempo, oppositional defiant, and overt conduct disorder symptoms. *Child and Family Behaviour Therapy*, 20, 1–14.
- Sørli, M. A. (2000). *Alvorlige atferdsproblemer og lovende tiltaks i skolen. En forskningsbasert kunnskapsanalyse* [Serious behaviour problems and promising treatment in school: A research based analysis of knowledge; in Norwegian]. Oslo, Norway: Praxis Forlag.
- Webster-Stratton, C., & Herbert, M. (1996). *Trouble families: Problem children*. Chichester, England: Wiley.
- Wichstrøm, L., Skogen, K., & Øya, T. (1996). The increased rate of conduct problems in urban areas: what is the mechanism? *Journal of American Academy of Child and Adolescent Psychiatry*, 35, 471–479.
- Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 71, 136–149.