Are Group-based Parenting Programs to Prevent Child Problem Behaviors Equally Effective for Children with Clinical and Non-clinical Problem Levels?

Frida Åström

Master’s Program in Psychology

Örebro University

Submitted in Partial Fulfillment of the Master’s Program in Psychology with Prevention Focus

Independent Research Report (15 credits)

Spring 2014

Supervisor: Metin Özdemir, Ph.D.
Abstract

The effectiveness of parenting programs in preventing child problem behaviors have been increasingly supported in the literature. However, studies revealed contradictory findings of how children with severe problem levels benefit from the programs. In this study, I performed secondary analysis on parent-reported externalizing behaviors, ADHD and ODD symptoms of 749 children, aged 3-12 years, to evaluate the effectiveness of four different parenting programs on children with clinical and non-clinical problems. Mixed design ANOVA models revealed that the programs significantly reduced children’s externalizing behaviors, ADHD and ODD symptoms at post-test. Furthermore, the same mixed design ANOVA models and follow-up tests revealed consistent significant differences in program response for children with and without initial clinical problems on all outcomes. Cohen’s $d$ within-group change effect sizes showed consistently large program effects for children with clinical problem levels on all outcomes. Findings indicate that group-based parenting programs are effective for children with externalizing behavior, ADHD and ODD symptoms, at least in the short term, and that children with clinical problem levels may benefit substantially.

*Keywords:* Parenting program, externalizing, problem behavior, ADHD, ODD, effectiveness study
Are Group-based Parenting Programs to Prevent Child Problem Behaviors Equally Effective for Children with Clinical and Non-clinical Problem Levels?

Externalizing problems among children and adolescents, such as, antisocial, aggressive, destructive, non-compliant, and oppositional behaviors, are among the most common problems that health services need to deal with (Dretzke et al., 2009). These problems are more common among boys, and have a high comorbidity with Attention Deficit Hyperactivity Disorder (ADHD; Furlong et al., 2012). Externalizing problems, especially if they start in early ages, are relatively stable across childhood (Nixon, 2002), and they are likely to develop into criminality and substance misuse (Furlong et al., 2012). Thus, leaving externalizing problems untreated is very costly for society. Parents are often seen as agents of child behavior change, and thus, the use of parenting programs to prevent externalizing problems at an early stage has increased in recent decades (Eisner, Nagin, Ribeaud, & Malti, 2012).

Parenting Programs

Parenting programs to reduce children’s problem behaviors have been studied since the 1960s, especially after it was found that parents could use reinforcement and extinction techniques to influence children’s behaviors (Barlow, Smailagic, Ferriter, Bennett, & Jones, 2010). Today, many parenting programs rely on behavioristic and social learning theories (Nixon, 2002), and these, so-called, behavioral programs are the most empirically supported parenting programs (Taylor & Biglan, 1998). Many programs are also increasingly delivered in group format (Dretzke et al., 2009), because of the economic advantages, and the possibility to help more families in a limited time frame (Taylor & Biglan, 1998). Other types of parenting programs exist and are often referred to as non-behavioral programs (Lundahl, Risser, & Lovejoy, 2006). These programs are based on different theories, such as, effective communication, attachment theory, and systems theory, and often emphasize relational
factors. However, non-behavioral programs have not been evaluated with the same rigor as behavioral programs, and their effectiveness on child problem behaviors remain largely unknown. Many reviews and meta-analyses have supported the effectiveness of parenting programs (Taylor & Biglan, 1998; Piquero, Farrington, Welsh, Tremblay, & Jennings, 2009; Dretzke et al., 2009), and recently, a rigorous meta-analysis found support for the effectiveness of behavioral parenting programs, at least in the short term (Furlong et al., 2012). Thus, the general effectiveness of particularly the behavioral parenting programs to reduce child problem behaviors is well supported.

**Differential Program Effect**

Despite the increased evidence for the short-term effects of parenting programs, there are a number of unanswered questions regarding the effectiveness of parent interventions. One of the unanswered questions is whether parenting programs are equally effective for all children with varying levels of problem severity. Specifically, meta-analyses have provided conflicting results on whether parenting programs are equally effective for children with different problem levels. Some meta-analyses reveal that program effectiveness is greater for children with severe problem levels (Lundahl et al., 2006; Reyno & McGrath, 2006; Menting, Orobio de Castro, & Matthys, 2013), while, another high quality meta-analysis found no greater effect for children with severe problem levels (Furlong et al., 2012). The opposite has also been found in narrative reviews, with less program effect for children with severe problem levels (Kazdin, 1997; Nixon, 2002). In conclusion, it is still unclear whether parenting programs are equally effective for children with different problem severity. Furthermore, it is also unclear whether different types of parenting programs are equally effective for children with the most severe problem levels. For example, it might be that one parenting program has a greater effect on children with severe problem levels, but this effect might be absent for other types parenting programs. These potential interaction effects are
difficult to examine with meta-analyses, since program effects are averaged across many different parenting programs (Gardner, Hutchings, Bywater, & Whitaker, 2010). Thus, we need more empirical studies that test the effectiveness of different parenting programs for children with severe problem levels.

Differential effects of parenting programs may be conceptualized in two ways. First, there might be predictors of program effect on the target group. Predictors influence program effect in the same way across all treatment conditions (Beauchaine, Webster-Stratton, & Reid, 2005). For example, if children with clinical problem levels respond differently than children without clinical problem levels, but consistently across different parenting programs, then problem level is a predictor of program outcome. The second conceptualization is the moderation of program effect. Moderators influence program effect differently across the treatment conditions (Beauchaine et al., 2005). For example, if children with clinical problem levels respond better to one parenting program than children without clinical problem levels, but not on another, then problem level is a moderator of program effect (Lavigne et al., 2008). Performing predictor and moderator analyses help to clarify for whom parenting programs work.

Several factors have been proposed as predictors or moderators of program effect, but few have received consistent empirical support. Child age has been proposed as a predictor based on theoretical reasoning, such that, younger children are more malleable than older children (Hautmann et al., 2010). However, one meta-analysis (Lundahl et al., 2006), and a pooled study, which combined data from six different randomized trials (Beauchaine et al., 2005), found no relation between child age and program outcome. Child gender has also been proposed as a predictor of program effect, but studies have produced ambiguous results (Hartmann et al., 2010). The effect of child gender has not been subject to meta-analytic procedures, possibly because treatment studies naturally include more boys. However,
Beauchaine and colleagues (2005) combined unweight data from six trials to achieve a larger sample of girls, but found no effect of child gender. Another commonly proposed predictor, socioeconomic status (SES), has been included in several meta-analyses. Some meta-analyses found that program effectiveness was greater for families with low SES (Lundahl et al., 2006; Reyno & McGrath, 2006), whereas another meta-analysis found that SES only had an effect when children’s problem levels were low, but not when children’s problem levels were high (Leijten, Raaijmakers, Orobio de Castro, & Matthys, 2013). In contrast, a high quality meta-analysis found no moderator role for economic disadvantage on program effectiveness (Furlong et al., 2012). Consistent with Furlong and colleagues, a recent study comparing the short-term effectiveness of four parenting programs found no consistent role of child age, child gender, parent age, parent education, SES or immigrant status, on program effectiveness for child externalizing problems, or parents’ behaviors, competence or well-being (Stattin et al., under review). In sum, more research are needed to understand whether the effectiveness of parenting programs is about equal for children of different ages and sexes, as well as across families of different socioeconomic status. However, the evidence suggesting no substantial roles is stronger than the evidence suggesting significant roles of child age, child gender, and socioeconomic status.

As for child problem level, both empirical findings and theoretical reasoning lead to contradictory conclusions. Hautmann and colleagues (2010) explained how two different hypotheses about child problem severity can be equally true from a theoretical perspective. First, it can be argued that children with severe problems have greater room for change, compared to children with less problems. This hypothesis suggests that program effectiveness would be greater for children with severe problems, compared to children with less problems. Second, it can also be argued that children with severe problems are more resistant to change, compared to children with less problems. This hypothesis instead suggests that program
effectiveness would be greater for children with less problem, compared to children with severe problems. Thus, we need more empirical studies that test these perspectives in different contexts to understand the role of problem level in the effectiveness of parenting programs.

The Present Study

The goal of the present study is to examine the overall effectiveness of four group-based parenting programs for children with clinical and non-clinical problem levels. I will also examine potential differences in effectiveness across programs, for children with clinical and non-clinical problem levels. To answer my research questions, I will use data from a randomized effectiveness trial, which evaluated how four different programs work in real life settings (Stattin et al., under review). The current study focuses on changes in externalizing behaviors, ADHD symptoms, and Oppositional Defiant Disorder (ODD) symptoms for children aged 3-12 years.

The current study compares effectiveness of four parenting programs that are commonly used in various service sectors in Sweden, such as, child and adolescent psychiatry clinics, social welfare, and schools. Two of these programs (Incredible Years and Komet) are behavioral programs. These programs focus on skills, such as, praise and reward, limit setting, dealing with misconduct, and play skills, but Komet has a particular emphasis on praise, and consists of smaller groups (parents of 6 children). In both programs, behavioral principles are taught, and parents watch videotapes of interactions between parents and children in different situations, and discuss the content with the support of two group leaders. The parents then role play the techniques, and homework are used between the 11 sessions to further improve the learning (Menting et al., 2013; Kling & Sundell, 2006). The third program, Cope, is based on several theories, such as social learning, family systems, and social-cognitive psychology. It also uses role plays, homework, and goal setting, but is mainly based on non-didactic learning. Cope focuses on strategies, such as, giving attention to positive behaviors, planning,
and balancing time among siblings. The program consists of 10 weekly sessions and has a special focus on large groups (25-30 parents) delivered in community centers or schools (Thorell, 2009). Finally, the Connect program is an attachment-based program, and focuses on building secure attachments between parents and pre-adolescents (9-12 years). The rationale is that the parent-teen relation need to be secure before any parenting techniques can be used effectively. Connect focuses on skills, such as, parental sensitivity, reflection, cooperation, and affect regulation. The program consists of 10 weekly sessions and includes didactic learning, role-plays, and reflective exercises. Much emphasis is put on observation and reflection of behaviors, rather than prescriptive advice (Moretti & Obsuth, 2009).

In sum, the aim of the current study is to test the role of baseline child problem severity for the effectiveness of group-based parenting programs. This research question is tested for all programs together, and then separately, for each of the programs to understand potential differences between the programs.

Method

Design and Study Setting

The current study is based on secondary analysis of data from a randomized controlled trial, which was designed to evaluate program effectiveness of four group-based parenting programs under real-world conditions (Gartlehner, Hansen, Nissman, Lohr, & Carey, 2006). Four research groups took part in the study: Örebro University, Karolinska Institute (Stockholm), Gothenburg University, and University of Lund. These regions represent the most populated areas of Sweden. In these regions, 30 agencies, such as, social welfare agencies, child and adolescent psychiatry clinics, and schools, with the capacity of delivering two of the four programs, were recruited for the trial. The participation in the research project did not change the agencies’ routine practices, and no further training or advertisements for the parenting programs were given.
The parents in the study had mainly sought help themselves, and a few were recruited from primary care, following regular practice. A first parent meeting was held at each site once the group was large enough. At this time, consents were signed, and the first questionnaire assessments were made. These meetings were the only deviation from the units’ routine practice in the study. Random assignment was made after the first assessment, and most parents received information about group allocation by mail, while some received this information at the first meeting. The parenting programs started 1-2 weeks after the first assessment and consisted of 72 parenting groups. As part of ordinary practice, a few participants received interpreter services. Post-test assessment took place immediately after the last program session (about 3-4 months after pre-test assessment), while questionnaires were sent by mail to parents who did not attend the last session.

**Randomization**

Because the ethics committee disagreed to include a control group in the project’s first year, the comparison condition consisted of a self-help condition. Participants in the self-help group were given a book on parent management with accompanied reading instructions. Thus, in the first year, participants were randomly assigned to either one of two parenting programs provided by each unit, or to the self-help condition. In the second year, a waitlist condition was permitted, and participants were randomly assigned to either one of the two parenting programs provided by each unit, or to a waitlist condition. The self-help condition and the waitlist conditions are not included in the current study. Findings regarding the program effect in comparison to the waitlist condition are reported elsewhere (Stattin et al., under review).

The randomization had to be stratified by age groups because different parenting programs aimed at different age groups. Parents of children aged 3-8 years were randomized to age appropriate versions of Incredible Years, Cope and Komet, while parents of children aged 9-12 years were randomized to Connect, Cope or Komet.
Participants

At the first assessment, parents of 1113 children attended the information meetings. Because of the inclusion criteria, three children younger than 3 years, one child older than 13, and one child diagnosed with autism spectrum disorder were excluded. In addition, six children were excluded because the service units considered manual-based parenting programs to be non-beneficial for these children. A total of 1104 parents were randomized (of which 196 were assigned to the self-help condition, and 159 to the waitlist condition). For the present study, the final sample consisted of 749 parents assigned to the four parenting groups. The flow of participants in the study is presented in Figure 1.

Of these 749 randomized parents, 635 (84.8%) started a parenting program. Of these 635 program starters, there were complete pre- and post-data for 598 (79.8% of those randomized to a parenting program, and 94.2% of the program starters). Fewer parents started the Incredible Years program (75.4%) compared to the other parenting groups. This loss was because of organizational problems in two of the communities offering the program. The communities’ geographical locations made it difficult to recruit a sufficient number of participants, so they recruited participants from neighboring communities. However, of those who started the Incredible Years program, many stayed and completed pre- and post-assessment (93.6%). As for statistical power, the sample size was sufficiently large to identify small effects. Each of the four parenting programs consisted of at least 75 participants, as recommended by Gartlehner and colleagues (2006), and all but the Incredible Years program had twice as many participants.

At pre-test, one fourth of the parents (24.6%) reported having additional help from a special education teacher, school psychologist, speech therapist, counselor, social worker, assistant, or contact person. At post-test, more than one fifth (21.4%) of the parents reported
having additional help. There were no significant differences between the four parenting groups on received additional help at baseline, or at post-test.

Mainly, there was only one parent who participated (78.1%) in the program, typically the mother, but in some cases the biological father. In this study, only one parent was chosen as reporter of child behavior, based on program attendance rates. If attendance was equal among the parents, the mother was chosen. Across the programs, 85% of reporters were mothers. Socio-demographic characteristics of the families are presented in Table 1.

Two thirds of the included children were boys. The mean age of the child was 7.70 years (SD = 2.60), with a range from 3 to 12 years. The mean age of the parent (primary reporter) was 37.70 years (SD = 7.51), with a range from 20 to 60 years. In most of the families (84%) both parents were born in Scandinavia. A large majority of the parents were married or cohabiting (74%), while the rest were singles. The average monthly household income after tax was 30,000 to 40,000 SEK (around 3000 - 4000 €). For some of the families (6.1%) the household income was between 0 - 10,000 SEK, and for one fourth (24.9%) the income was higher than 50,000 SEK. The majority of the parents (62%) reported that the income was sufficient for household expenses, while some (6.3%) reported that the income was insufficient. About half of the parents (45%) had completed some university education, while some (9%) only had compulsory school education.

**Attrition Analyses**

There were complete pre-test data on all but three of the randomized families, allowing predictor analyses to be made for later drop-outs. First, it was examined whether baseline characteristics of the parents and children were related to starting a parenting program, using logistic regression analyses. In the model, all demographic variables, as well as the parent and child outcomes were included as predictors. The variable “starting a parenting program” was the dependent variable (0 = started, and 1 = did not start). None of
the baseline variables significantly predicted parents’ act of starting a program. Second, it was examined if baseline characteristics predicted drop-out at post-test (0 = participated, 1= lost to follow-up), using another logistic regression model. In this analysis, one of the demographic variables, and two of the parent outcome variables significantly predicted drop-out.

Figure 1. Recruitment, randomization, and follow-up.
Table 1.

**Socio-Demographic Characteristics of Participants.**

<table>
<thead>
<tr>
<th></th>
<th>Komet</th>
<th>Cope</th>
<th>Incredible Years</th>
<th>Connect</th>
<th>F/Chi-sq. (df)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Child</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child Gender</td>
<td>64.90%</td>
<td>61.10%</td>
<td>65.60%</td>
<td>65.80%</td>
<td>1.04 (3)</td>
<td>.792</td>
</tr>
<tr>
<td>Child age</td>
<td>7.32a (2.41)</td>
<td>7.07a (2.54)</td>
<td>6.78a (2.17)</td>
<td>9.8b (1.39)</td>
<td>69.78 (3, 625)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Parent</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent gender</td>
<td>88.80%</td>
<td>86.90%</td>
<td>81.30%</td>
<td>81.60%</td>
<td>5.12 (3)</td>
<td>.163</td>
</tr>
<tr>
<td>Parent age</td>
<td>37.96a (6.63)</td>
<td>37.28a (6.07)</td>
<td>37.08a (6.04)</td>
<td>39.98b (5.23)</td>
<td>8.10 (3, 618)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Marital Status</td>
<td>77.10%</td>
<td>71.80%</td>
<td>70.70%</td>
<td>71.10%</td>
<td>2.21 (3)</td>
<td>.657</td>
</tr>
<tr>
<td>Immigrant Status</td>
<td>13.10%</td>
<td>14.00%</td>
<td>18.00%</td>
<td>12.60%</td>
<td>1.62 (3)</td>
<td>.265</td>
</tr>
<tr>
<td>Monthly Income</td>
<td>4.27 (1.54)</td>
<td>4.01 (1.65)</td>
<td>3.75 (1.66)</td>
<td>4.05 (1.62)</td>
<td>1.58 (3, 619)</td>
<td>.194</td>
</tr>
<tr>
<td>Economic Strain</td>
<td>2.79 (.62)</td>
<td>2.74 (.71)</td>
<td>2.64 (.78)</td>
<td>2.68 (.81)</td>
<td>.44 (3, 621)</td>
<td>.723</td>
</tr>
<tr>
<td>Education Level</td>
<td>3.14 (.93)</td>
<td>3.11 (1.07)</td>
<td>2.87 (.95)</td>
<td>3.09 (1.02)</td>
<td>.68 (3, 624)</td>
<td>.580</td>
</tr>
</tbody>
</table>

*Note. Standard deviations of mean values are presented in parentheses. Same superscripts indicate non-significant between-group differences and different superscripts indicate significant group differences. ¹ Child gender was coded as 1 = boy, 0 = girl. Parent gender was coded as 1 = mother, 0 = father. Marital status was coded as 1 = married, 0 = other. Immigrant status was coded as 1 = immigrant, 0 = Swedish.*
Specifically, parents with an immigrant background (OR = 2.26, \(p = .026\)), or higher scores on depressive symptoms (OR = 1.92, \(p = .002\)), or higher sense of parenting competence (OR = 1.49, \(p = .040\)), were more likely to drop out at post-test assessment. However, because these predictors explained only a small amount of the variation in drop-outs (Nagelkerke \(R^2 = .07\)), the conclusion was made that attrition would not seriously bias the study.

Measures

Parents responded to questions about the socio-demographic characteristics of themselves and their family at pre-test. Responses to statements of child problem behaviors, ADHD symptoms, and ODD symptoms were given at both pre- and post-test assessments.

Socio-demographic Measures

Parents responded open-ended questions about the gender and age of the child and primary reporter. Immigrant status was assessed by asking parents where they were born (in Sweden, another Scandinavian country, another European country, or a non-European country), and was assumed if one or both of the parents had migrated to Sweden from a non-European country. Marital status was assessed by asking whether the parent was married, cohabitating, single, widowed, or something else. Education level of the primary reporter was measured on a four-point scale, from 1 (compulsory school), to 4 (university degree). Parents rated household monthly income on a six-point scale, from 1 (0-10,000 SEK), to 6 (more than 50,000 SEK) per month. Economic strain was assessed with one item asking about whether family income matched household expenses. The responses were rated on a four-point scale, from 1 (“Our monthly income does not cover our expenses”), to 4 (“Our monthly income is fine and we do not think about what we spend”).

Child Behavior Measures

Externalizing Problems. To assess children’s externalizing problems, the Eyberg Child Behavior Inventory (ECBI; Colvin, Eyberg, & Adams, 1999) was used. The ECBI is a
commonly used 36-item questionnaire with both intensity, and problem perception subscales. On the intensity scale, parents rated each of the child’s problem behaviors on a seven-point scale (1 = never, 7 = always), reflecting the frequency of each problem behavior. Example of items are: “Does not obey house rules, “Whine”, and “Teases siblings”. Cronbach’s alpha for this scale was .92 at pre-test and .94 at post-test. On the problem perception subscale, parents provided yes-no responses to statements of whether they perceived their child’s behavior as problematic. Cronbach’s alpha for this scale was .92 at pre-and post-test.

**Attention Deficit and Hyperactivity Problems.** To assess children’s ADHD symptoms, two subscales from the SNAP-IV questionnaire (Swanson, Nolan, & Pelham, 1992) was used. The two subscales consist of responses on a 4-point scale (0 = not at all, 3 = very much), indicating the frequency of symptoms. Symptoms of inattention were measured with 9 items; e.g. “Often has difficulty sustaining attention in tasks or activities”, and “Often is forgetful in daily activities”). Cronbach’s alpha for this scale was .91 at both pre- and post-test. Symptoms of hyperactivity/impulsivity were measured with 9 items; e.g. “Often talks excessively much”, and “Often has a hard time waiting for his/her turn”. Cronbach’s alpha for this scale was .92 at pre-test, and .91 at post-test.

**Oppositional Defiant Disorder Problems.** To assess children’s ODD symptoms, a third subscale from the SNAP-IV questionnaire was used. This subscale also consist of responses on a 4-point scale (0 = not at all, 3 = very much), and ODD was measured with 8 items; e.g. “Often loses temper”, and “Often argues with adults”. Cronbach’s alpha for this scale was .91 at both pre- and post-test.

**Attendance and Implementation Fidelity**

Program attendance was calculated as percentages, considering the four programs’ different numbers of sessions. Specifically, Komet had 11 sessions á 2 ½ hours, Cope and Incredible Years had 12 sessions á 2 hours, and Connect had 10 sessions á 1 hour. Across the
programs, 70% of the parents attended at least 75% of the sessions, 18.6% attended between 50% and 75% of the sessions, 4.9% attended between 20% and 50% of the sessions and 6.5% attended less than 25% of the sessions. Overall, the attendance rates were high. Yet, there were significantly higher attendance rates for Connect (85.2%) and Komet (78.8%), than for Incredible Years (72.1%) and Cope (69.3%), $F(3, 442) = 13.34, p < .001$.

Implementation fidelity was measured using expert ratings on a random selection of videotaped sessions. Each of 72 parenting program were videotaped on three random occasions, and resulted in 216 videotapes. These videotapes were stratified by the number of sessions for each program, and 25% were randomly selected for ratings ($n = 56$). The raters were two trained specialists from each program who had implemented and trained group leaders. The implementation fidelity, defined as the degree to which the group leader followed the program manual, was rated on a 10-point scale from 1 (not at all) to 10 (totally). Inter-rater agreement was measured as the correlation between the two ratings, and was high ($r = .84$). Across the programs, implementation fidelity proved to be high ($M = 7.93$), ranging from 6.86 to 9.04. However, Levene’s F-test, followed by a post-hoc test, revealed a significant difference between the programs. The ratings of Cope ($M = 9.04$) and Komet ($M = 7.94$) were significantly higher than Connect ($M = 7.52$) and Incredible Years ($M = 6.86$). Overall, the high level of program attendance and implementation fidelity reduced the risk for unbiased results.

**Analyses**

To examine how the children with clinical and non-clinical problem levels benefitted from the parenting programs, I identified the 95th percentile cut-off values for the ECBI, and SNAP-IV measures, based on the baseline assessment. For each child outcome measure, I created a dichotomous variable to identify children with and without clinical levels of problems (coded as 1 and 0, respectively). For ECBI intensity and ECBI problem scores, I
used the Swedish norm values of 95\textsuperscript{th} percentile (Axberg, Johansson Hanse, & Broberg, 2008). The 95\textsuperscript{th} percentile cut-off values for the ADHD symptoms (i.e., inattention and hyperactivity), and ODD symptoms, were obtained from [http://www.adhd.net/SNAP-IV-instructions.pdf](http://www.adhd.net/SNAP-IV-instructions.pdf). I used these dichotomous variables as grouping variables in mixed design ANOVA models to examine the changes in each of the child outcome measures from baseline to post-test, for children with and without clinical problem levels at baseline.

To adjust for attrition, I used intention-to-treat analyses and replaced missing values at post-test with baseline values to obtain less biased estimates. This approach assumes that the problem levels remained unchanged for those who dropped-out, and reduces Type I error rate due to selective attrition in the sample (Steiner & Geddes, 2001). Finally, I estimated effect sizes for within-group over time change separately for the two groups (with and without clinical problem levels), using Cohen’s $d$ formula (Grissom & Kim, 2005). I interpreted the effect sizes using Cohen’s (1988) guidelines of 0.20 small, 0.50 moderate, and 0.80 large.

**Results**

**Descriptive Analyses**

The proportion of children at or above clinical cut-off values at baseline, for all outcomes, and across the programs, are described in Table 2. Overall, a substantial number of children had externalizing problems within the clinical range. About 2/5 of the children (41.1\%) had clinical levels on the ECBI intensity measure, and more than half (54.6\%) had clinical levels on the ECBI problem measure. For ADHD symptoms, children had fewer problems within the clinical range, compared to ECBI intensity and ECBI problem measures. Specifically, based on 95\textsuperscript{th} percentile cut-off values for the SNAP-IV measures, 21.1\% of the children had inattention problems, and 35.1\% had hyperactivity problems, at or above clinical levels. For ODD symptoms, children had fewer problems within the clinical range, compared to externalizing problems and ADHD symptoms, with 14.3\% of the children having ODD
Table 2.

*Means, (SDs) at Baseline and Post-test, and Percentages of Children with Clinical Problem Levels at Baseline.*

<table>
<thead>
<tr>
<th></th>
<th>Komet</th>
<th>IY</th>
<th>Cope</th>
<th>Connect</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ECBI Intensity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>3.58 (.85)</td>
<td>2.68 (.86)</td>
<td>3.72 (.76)</td>
<td>2.93 (.73)</td>
<td>3.62 (.82)</td>
</tr>
<tr>
<td>&gt;95th percentile at T1</td>
<td>37.1%</td>
<td>51.7%</td>
<td>42.9%</td>
<td>38.1%</td>
<td>41.1%</td>
</tr>
<tr>
<td><strong>ECBI Problem</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>.39 (.23)</td>
<td>.22 (.21)</td>
<td>.44 (.22)</td>
<td>.29 (.23)</td>
<td>.41 (.22)</td>
</tr>
<tr>
<td>&gt;95th percentile at T1</td>
<td>49.7%</td>
<td>62.9%</td>
<td>53.7%</td>
<td>55.8%</td>
<td>54.6%</td>
</tr>
<tr>
<td><strong>Inattention</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.05 (.73)</td>
<td>.79 (.63)</td>
<td>1.19 (.76)</td>
<td>.85 (.65)</td>
<td>1.19 (.74)</td>
</tr>
<tr>
<td>&gt;95th percentile at T1</td>
<td>16.6%</td>
<td>20.5%</td>
<td>22.4%</td>
<td>24.6%</td>
<td>21.2%</td>
</tr>
<tr>
<td><strong>Hyperactivity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.03 (.78)</td>
<td>.83 (.71)</td>
<td>1.26 (.77)</td>
<td>.96 (.61)</td>
<td>1.18 (.72)</td>
</tr>
<tr>
<td>&gt;95th percentile at T1</td>
<td>28.8%</td>
<td>46.2%</td>
<td>38.9%</td>
<td>32.1%</td>
<td>35.1%</td>
</tr>
<tr>
<td><strong>ODD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.19 (.73)</td>
<td>.81 (.68)</td>
<td>1.19 (.69)</td>
<td>.80 (.56)</td>
<td>1.16 (.64)</td>
</tr>
<tr>
<td>&gt;95th percentile at T1</td>
<td>15.8%</td>
<td>14.3%</td>
<td>8.6%</td>
<td>18.1%</td>
<td>14.3%</td>
</tr>
</tbody>
</table>

*Note.* Problem level was coded as 1 = above 95th percentile, and 0 = below 95th percentile.
problems within clinical range. Across the programs, the Incredible Years had a higher proportion of children with clinical levels on the ECBI intensity, ECBI problem, and SNAP-IV hyperactivity measures, compared to the other programs. In contrast, Komet had the lowest proportion of children with clinical levels on all measures, except on the SNAP-IV ODD measure.

**Did Children with Clinical and Non-clinical Problems Benefit Equally from the Programs?**

To test whether the children with and without clinically high levels of problems benefitted equally from the parenting programs, I conducted a series of mixed design ANOVA models. I used the child behavior measurements as repeated measure, and the dichotomized baseline severity of each behavior measure (i.e., 1 = above 95th percentile, 0 = below 95th percentile cut-off value) as between subject factor (see Figure 2 and Table 3). The results suggested, first, that there was a significant reduction in the ECBI intensity scores across all children, \(F(1, 626) = 952.87, p < .001, \eta^2 = .60\). The results also suggested that the amount of reductions in ECBI intensity scores for the children with and without clinical problem levels at baseline were statistically different, \(F(1, 626) = 44.36, p < .001, \eta^2 = .07\). The follow-up tests showed that the amount of decrease for children with clinically high problems was larger (\(\eta^2 = .64\)) than for the children whose ECBI intensity scores were lower than the 95th percentile (\(\eta^2 = .55\)).

For ECBI problem scores, there was a significant reduction across all children, \(F(1, 619) = 254.78, p < .001, \eta^2 = .29\). Similar to the findings related to ECBI intensity scores, the results suggested that the amount of reductions in ECBI problem scores for the children with and without clinical problem levels at baseline were statistically different, \(F(1, 619) = 88.08, p < .001, \eta^2 = .013\). The follow-up tests showed that the amount of decrease for children with clinically high problems was larger (\(\eta^2 = .43\)) than for the children whose
ECBI problem scores were lower than the 95th percentile ($\eta^2 = .12$). In sum, the children with higher levels of externalizing problems at baseline appeared to benefit more from the parenting programs, than the children with lower levels of externalizing problems at baseline.

For SNAP-IV inattention scores, there was a significant reduction across all children, $F(1, 624) = 238.61, p < .001, \eta^2 = .28$. The results also suggested that the amount of reductions in SNAP-IV inattention scores for the children with and without clinical problem levels baseline were statistically different $F(1, 624) = 36.34, p < .001, \eta^2 = .06$. The follow-up tests showed that the amount of decrease for children with clinically high problems was larger ($\eta^2 = .39$) than for the children whose SNAP-IV inattention scores were lower than the 95th percentile ($\eta^2 = .21$).

For SNAP-IV hyperactivity scores, there was a significant reduction across all children $F(1, 627) = 216.57, p < .001, \eta^2 = .26$. The results also suggested that the amount of reductions in SNAP-IV hyperactivity scores for the children with and without clinical problem levels at baseline were statistically different $F(1, 619) = 86.86, p < .001, \eta^2 = .12$. The follow-up tests showed that the amount of decrease for children with clinically high problems was larger ($\eta^2 = .42$) than for the group whose SNAP-IV hyperactivity scores were lower than the 95th percentile ($\eta^2 = .06$). In sum, the children with higher levels of ADHD symptoms at baseline appeared to benefit more from the parenting programs, than children with lower levels of ADHD symptoms at baseline.

Finally, for SNAP-IV ODD scores, there was a significant reduction across all children $F(1, 628) = 234.42, p < .001, \eta^2 = .27$. The results also suggested that the amount of reductions in SNAP-IV ODD scores for the children with and without clinical problem levels at baseline were statistically different $F(1, 628) = 38.63, p < .001, \eta^2 = .06$. The follow-up tests showed that the amount of decrease for children with clinically high problems was larger ($\eta^2 = .47$) than for the children whose SNAP-IV ODD scores were lower than the 95th
percentile ($\eta^2 = .23$). In sum, the children with higher levels of ODD symptoms at baseline appeared to benefit more from the parenting programs than children with lower levels of ODD symptoms at baseline.

Figure 2. Parent-reported change from baseline to post-test for children with and without clinical problem levels at baseline.
Table 3.

Results of the Mixed Design ANOVAs.

<table>
<thead>
<tr>
<th></th>
<th>F (df)</th>
<th>p</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ECBI Intensity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overtime changes</td>
<td>952.87 (1, 626)</td>
<td>&lt;.001</td>
<td>.60</td>
</tr>
<tr>
<td>Time*Problem level interaction</td>
<td>44.36 (1, 626)</td>
<td>&lt;.001</td>
<td>.07</td>
</tr>
<tr>
<td><strong>ECBI Problem</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overtime changes</td>
<td>254.78 (1, 619)</td>
<td>&lt;.001</td>
<td>.29</td>
</tr>
<tr>
<td>Time*Problem level interaction</td>
<td>88.08 (1, 619)</td>
<td>&lt;.001</td>
<td>.13</td>
</tr>
<tr>
<td><strong>Inattention</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overtime changes</td>
<td>238.61 (1, 624)</td>
<td>&lt;.001</td>
<td>.28</td>
</tr>
<tr>
<td>Time*Problem level interaction</td>
<td>36.34 (1, 624)</td>
<td>&lt;.001</td>
<td>.06</td>
</tr>
<tr>
<td><strong>Hyperactivity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overtime changes</td>
<td>216.57 (1, 627)</td>
<td>&lt;.001</td>
<td>.26</td>
</tr>
<tr>
<td>Time*Problem level interaction</td>
<td>86.86 (1, 627)</td>
<td>&lt;.001</td>
<td>.12</td>
</tr>
<tr>
<td><strong>ODD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overtime changes</td>
<td>234.42 (1, 628)</td>
<td>&lt;.001</td>
<td>.27</td>
</tr>
<tr>
<td>Time*Problem level interaction</td>
<td>38.63 (1, 628)</td>
<td>&lt;.001</td>
<td>.06</td>
</tr>
</tbody>
</table>

*Note.* η² = Partial eta-squared. Problem level was coded as 1 = clinical, 0 = non-clinical.

The Magnitude of Reductions for Children with and without Clinical Problem Levels

I computed Cohen’s *d* effect size estimates for within-group over time change separately, for children with and without clinical problem levels, to examine the magnitude of the overall program effect (see Table 4). The results suggested, first, that for the ECBI intensity measure, parenting programs led to larger changes for children whose baseline scores were above 95th percentile (*d* = 1.91), than for the children whose baseline scores were below 95th percentile (*d* = 1.15). However, for both groups, the magnitude of change was large, according to Cohen’s (1988) guidelines. On the ECBI problem measure, parenting programs led to substantially larger changes for children whose baseline scores were above 95th percentile (*d* = 1.22), than for children whose baseline scores were below 95th percentile...
(d = 0.47). Here, the magnitude of change was large for the clinically high group, and small-moderate for the non-clinical group, according to Cohen’s guidelines. In sum, there were consistently larger program effects for children with clinical levels of externalizing behaviors at baseline, compared to children without clinical levels of externalizing behaviors at baseline. However, the difference between the two groups was more evident for the ECBI problem measure, than for the ECBI intensity measure.

Overall, the differences in program effect for children with and without clinical problems were larger for ADHD symptoms, than for externalizing problems. Specifically, on the SNAP-IV inattention measure, the parenting programs led to substantially larger changes for children whose baseline scores were above 95th percentile (d = 1.58) than for children whose baseline scores were below 95th percentile (d = 0.40). The magnitude of change was large for the clinically high group, and small for the non-clinical group, according to Cohen’s guidelines. Consistently, on the SNAP-IV hyperactivity measure, the parenting programs also led to substantially larger changes for children whose baseline scores were above 95th percentile (d = 1.01), than for children whose baseline scores were below 95th percentile (d = 0.24). The magnitude of change was large, for the clinically high group, and small or insignificant, for the non-clinical group. In sum, there were consistently larger program effects for children with clinical levels of ADHD symptoms at baseline, compared to children without clinical levels of ADHD symptoms at baseline.

Finally, on the SNAP-IV ODD measure, the programs led to substantially larger changes for children whose baseline scores were above 95th percentile (d = 2.18), than for children whose baseline scores were below 95th percentile (d = 0.50). The magnitude of change was large for the clinically high group, and moderate for the non-clinical group, according to Cohen’s guidelines. In sum, there was a larger program effect for children with
Table 4.

*Cohen’s d Effect Sizes for Change from Baseline to Post-test for Children with and without Clinical Problem Levels.*

<table>
<thead>
<tr>
<th></th>
<th>Komet</th>
<th>Connect</th>
<th>Incredible Years</th>
<th>Cope</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>&gt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>&gt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>ECBI Intensity</td>
<td>1.54</td>
<td>2.37</td>
<td>.81</td>
<td>1.31</td>
<td>1.32</td>
</tr>
<tr>
<td>ECBI Problems</td>
<td>.81</td>
<td>1.57</td>
<td>.12</td>
<td>.84</td>
<td>.43</td>
</tr>
<tr>
<td>Inattention</td>
<td>.40</td>
<td>2.34</td>
<td>.35</td>
<td>1.08</td>
<td>.53</td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>.27</td>
<td>.86</td>
<td>.15</td>
<td>.73</td>
<td>.27</td>
</tr>
<tr>
<td>ODD</td>
<td>.54</td>
<td>2.42</td>
<td>.37</td>
<td>1.29</td>
<td>.47</td>
</tr>
</tbody>
</table>

*Note.* Problem level was coded as 1 = above 95<sup>th</sup> percentile, 0 = below 95<sup>th</sup> percentile.
clinical levels of ODD symptoms at baseline, compared to children without clinical levels of ODD symptoms at baseline.

**Did the Effects for Children with and without Clinical Problem Levels Differ across the Programs?**

To examine whether the overall finding of differential effects for children with baseline clinical or non-clinical problem levels were consistent across the programs, I fitted another mixed design ANOVA model. I used the child behavior measurements as repeated measure, and the dichotomized baseline severity of each behavior measure (i.e. 1 = above 95th percentile, 0 = below 95th percentile), as well as the type of program, as between-subject factors. The results suggested, first, that across all children, there were significant over time reductions for all outcome measures. The results also suggested, that on four of the outcome measures, the differences in program response for children with and without baseline clinical problems were consistent across the parenting programs. Specifically, for the ECBI intensity measure, the differential effect for children with and without clinical problem levels did not statistically vary between the programs, $F(3, 620) = .275, p = .936, \eta^2 = 0.00$. Similarly, for the ECBI problem measure, the differential effect for children with and without clinical problem levels did not statistically vary between the programs, $F(3, 613) = .140, p = .844, \eta^2 = 0.00$. Similarly, the findings for the SNAP-IV inattention measure suggested that the differential effect for children with and without clinical problem levels did not statistically vary between the programs, $F(3, 618) = 2.129, p = .095, \eta^2 = 0.01$. Consistently, for the SNAP-IV hyperactivity measure, the differential effect for children with and without clinical problem levels did not statistically vary between the programs, $F(3, 620) = .275, p = .844, \eta^2 = 0.00$. However, for the SNAP-IV ODD measure, the results suggested that the effects for children with and without clinical problem levels did vary between the programs, $F(3, 622) = 6.495, p < .001, \eta^2 = 0.03$). The follow-up test revealed that the decrease in ODD symptoms
was not statistically different for children with and without clinical problem levels in the Connect program, $F(1, 191) = 3.546, p = .061$, $η^2 = .02$, and in the Cope program, $F(1, 173) = 3.309, p = .071$, $η^2 = .02$. However, the decrease in ODD symptoms was statistically different for children with and without clinical problem levels in the Komet program, $F(1, 169) = 17.273, p < .001$, and in the Incredible Years program, $F(1, 89) = 40.852, p < .001$. In sum, across the parenting programs, the children with and without baseline clinical problem levels benefitted differently from the programs on four of the outcomes, except on ODD symptoms.

In order to examine the magnitude of program effects across the different programs, I computed Cohen’s $d$ effect size estimates for within-group over time change separately, for children with and without clinical problem levels, and by programs (see Table 4). Overall, the results suggested, that all four parenting programs led to larger changes for children with clinical problem levels at baseline, compared to children without clinical problem levels at baseline. This pattern of results was found for all outcome measures, including the SNAP-IV ODD measure. However, the differences in magnitude of change for children with and without clinical problem levels were less pronounced for the ECBI intensity measure, compared to the other measures. Across the programs, the magnitude of change was consistently large for children with clinical problem levels (except for the hyperactivity measure in the Connect program), according to Cohen’s guidelines. In sum, across the programs, there were consistent differences in magnitude of change for children with and without clinical problem levels, with consistent large effects for children with clinical problem levels.

**Discussion**

Previous research has revealed ambiguous findings related to the effectiveness of parenting programs in helping children with severe problem behaviors. To contribute to the
understanding of the effects of parenting programs on children with and without clinically high levels of externalizing problems, ADHD symptoms, and ODD symptoms, I analyzed data from a national randomized trial of four commonly used parenting programs.

The current study revealed that all four parenting programs were effective in reducing child externalizing problems, across all children. However, the program effects were consistently larger for children with severe externalizing problems, than for children with less severe externalizing problems. This pattern of results was found across all four parenting programs, and for both behavior intensity and problem perception measures, even though the difference was more evident for the problem perception measure. Also, the magnitude of changes for children with severe externalizing problems were consistently large across the programs. The present findings suggest that children’s initial problem severity is a predictor of program effect, rather than a moderator of program effect. Specifically, children with different levels of externalizing problems responded differently to the parenting programs. The children with higher problem severity benefitted more than children with less problem severity, and they did so consistently across the four programs. Thus, despite previous reviews proposing that children with severe externalizing problems benefit less (Kazdin, 1997; Nixon, 2002), or about the same amount (Furlong et al., 2012) as children with less problems, this study suggests that children’s problem levels predict program effect, and that children with severe externalizing problems may benefit substantially from group-based parenting programs.

Another important finding of the current study was concerned with the program effects on ADHD symptoms. The results revealed that all four parenting programs were effective in reducing ADHD symptoms, across all children. Consistent with the findings related to externalizing problems, the program effects were also consistently larger for children with severe ADHD symptoms, than for children with less severe ADHD symptoms. Specifically,
children with severe problems of inattention and hyperactivity/impulsivity benefitted more from the parenting programs overall, and across the four parenting programs, compared to children with less severe ADHD symptoms. The differences were also more pronounced for ADHD symptoms, than for externalizing problems. The magnitude of changes for children with severe ADHD symptoms were also consistently large. These findings were promising considering the findings of previous research.

Prior studies have commonly found that ADHD has a strong biological basis (e.g., Purdie, Hattie, & Carroll, 2002), and that ADHD symptoms have a “relatively high stability” (Larsson, Larsson, & Lichtenstein, 2004, p. 1272) during childhood. Furthermore, the few existing meta-analyses which summarized the effectiveness of parent interventions on ADHD symptoms reported mixed findings. For example, in a recent meta-analysis, Zwi, Jones, Thorgaard, York, & Dennis (2011) examined the effectiveness of both randomized and quasi-randomized trials of parenting programs for children diagnosed with ADHD. Nevertheless, they were able to locate only four studies with ADHD-related measures (hyperactivity-impulsivity and inattention). The studies, which met the inclusion criteria, lacked methodological rigor and could not be included in the meta-analysis. Other meta-analyses with more flexible inclusion criteria have revealed modest results. For example, Purdie and colleagues (2002) included only four studies of parenting programs in their evaluation of ADHD interventions, and found a small effect ($d = 0.31$) across broad ADHD-related outcomes (behavioral, social, cognitive, and emotional). However, in the short-term, no effects were found on core ADHD symptoms, such as, hyperactivity and impulsivity, and only a small effect was found on attention. Also, for the few studies incorporating long-term follow-up assessments, the effects were not maintained for any type of intervention. Similarly, Corcoran & Dattalo (2006) performed a meta-analysis of different parent-focused interventions (including therapies) for children with ADHD diagnosis or symptoms, and
found a small effect \((d = 0.40)\) on ADHD symptoms across different informants (parent, teacher, and child). Finally, Lee, Niew, Yang, Chen, & Lin (2013) performed a meta-analysis on the effect of behavioral parenting programs for children with ADHD, and found a moderate effect size \((r = .39)\) on parent-reported child behavior (disruptive behavior and positive behavior) in the short-term. However, they found a smaller program effect for children comorbid with externalizing problems, and the overall program effect was reduced from moderate to small at long-term follow-up.

In sum, previous research revealed that ADHD symptoms are difficult to change, and the effects of parenting programs on children with ADHD appear modest and temporary. By contrast, the current findings revealed large effects for children with very high levels of ADHD symptoms. However, because the study used 95th percentile cut-off values to obtain potential clinical and non-clinical samples, the current findings can only be generalized to children with symptoms of ADHD, and not to children with clinical ADHD diagnosis. In addition, the current findings are limited to immediate program effects. Thus, future research needs to examine whether the four programs that were examined here, are equally effective on children with ADHD diagnosis, and if they are effective also in longer terms.

The current study also revealed that all four parenting programs were effective in reducing symptoms of oppositional defiant disorder, across all children. Consistent with the findings for externalizing problems, and ADHD symptoms, the overall program effect was larger for children with severe symptoms of ODD. However, this pattern of results was not found for the Connect and Cope program, but for the Komet and Incredible Years program. Thus, more replications are needed to fully understand this inconsistency. Nonetheless, for ODD symptoms, the effect sizes revealed clear differences in magnitude of change for children with and without clinical problem levels, with large effects for children with severe symptoms of ODD, across the programs.
One important aspect of the current study is that the data came from a randomized effectiveness trial. That is, all four parenting programs were implemented in ordinary service settings through the regular staff members. It is often argued that preventive interventions are more effective when the program developers are involved in the implementation process (Petrosino & Soydan, 2005). By contrast, when the programs are implemented in real life settings, their effectiveness are lower (Eisner et al., 2012). The current findings showed that parenting programs, as implemented in real life settings, could be highly effective in reducing child externalizing problems, ADHD and ODD symptoms, especially for children with high levels of problems. In sum, the observed effects suggest that group-based parenting programs could be promoted as part of regular service schedule of health services. However, before a comprehensive dissemination, the current findings need to be replicated in independent trials.

**Limitations and Strengths of the Study**

Three aspects of the study need to be cautiously considered. First, I used within-group change effect sizes to examine differences in program effect for clinical and non-clinical children. These intra-group comparisons are not corrected for potential improvements in the control group, and in turn, may lead to overestimation of program effects (Leijten et al., 2013). Thus, these intra-group effect sizes cannot be interpreted as measures of overall program effect. They should be interpreted as the magnitude of changes over time, without any reference to the magnitude of effects in comparison to a control condition. Despite its limitations, within-group change effect sizes are useful in comparing how much reductions were observed across different groups, especially when the baseline problems levels are substantially different.

Second, the present study examined whether the baseline problem level was related to how much children were improved following the parenting programs. However, it is important not to only focus on children’s greatest change, but also on children’s actual
problem levels after program completion. A previous study found that children who had the highest problem levels at the start of the program, also had the highest problem levels by the end of the program, despite that they showed the largest decrease (Hautmann et al., 2010).

Thus, it would be preferable to include a measure of clinical significance at post-test to fully evaluate the effects on children with initial clinical problem levels. Specifically, including a blind clinical evaluation could be very informative with respect to the value of the current findings for clinical practice.

Third, this study only included one post-test assessment. The results from this assessment support the view that children with clinical problem levels have more room for change, rather than being more difficult to treat. However, it may be that both theoretical views of severe problem levels are correct. Even though children with severe problem levels may have more room for change in the short-term, they may still remain difficult to treat in the longer-term. For example, children with initial clinical externalizing problems may reveal a large decrease immediately after program completion, but may not maintain the reductions in the days following the end of the program. Thus, additional follow-ups are needed to evaluate whether children with initial clinical problem levels maintain the gains in the long-term.

There are also a number of strengths in the present study. Overall, the study is unique because it examined real-life effectiveness of four parenting programs on a large sample of children with limited attrition. Specifically, performing effectiveness studies in public agencies provide robust tests of program effectiveness, because circumstances of program implementation might not be as ideal as in research settings (Flay et al., 2005). In addition, given the recent finding that larger samples tend to reveal smaller effects (McMahon, Holly, Harrington, Roberts, & Green, 2008), the large sample size in the present study decreased the risk for overestimated results. The study also had a large number of children with clinical
problem levels which allowed for balanced sub-group analyses. Furthermore, because the study included four theoretically different parenting programs which all revealed the same pattern of results, the effect on children with clinical problem levels were supported for both behavioral and non-behavioral programs. Finally, the relatively low amount of drop-out, and use of intention-to-treat analysis in the study allowed for more valid conclusions.

**Conclusion**

A key issue for clinical practitioners is to know which interventions work for children with severe levels of externalizing behaviors, ADHD and ODD symptoms. This study support that group-based parenting programs are effective in reducing externalizing behaviors, as well as ADHD and ODD symptoms, for children with severe problem levels, in real life settings. Notably, the present findings concern short-term effects of four parenting programs (Komet, Connect, Incredible Years, and Cope) offered to parents of children between 3 and 12 years. Thus, the findings cannot be generalized to other parenting programs not included in the study, or to children younger than 3 years, or older than 12 years. In addition, the results cannot be generalized to children with clinical ADHD diagnosis. More long-term follow-ups, independent replications, and blind clinical evaluations are needed to fully evaluate program effects. Nonetheless, group-based parenting programs are promising interventions that can be offered to many families in a limited time-frame, and in turn, remain cost-efficient for society.
References


Hautmann, C., Eichelberger, I., Hanisch, C., Plück, J., Walter, D., & Döpfner, M. (2010). The severely impaired do profit the most: Short-term and long-term predictors of therapeutic change for a parent management training under routine care conditions for


Stattin et al. (under review).

Steiner, D., & Geddes, J. (2001). Intention to treat analysis in clinical trials where there are missing data. *Evidence Based Mental Health, 4*, 70-71. doi:10.1136/ebmh.4.3.70


