Child prevention and group based parenting programs
Till alla barn
stora som små
Child prevention and group based parenting programs

Effectiveness and implementation
Abstract


Approximately 10–25% of children and youth suffer from mental health problems, such as depression, emotional difficulties, and disruptive behaviors. The evidence base of the effectiveness of preventive interventions targeting youth mental health currently delivered in regular care is weak. Also, little is known about what is needed for continued delivery of preventive programs in regular care. Hence, there is an evident need of effectiveness evaluations of preventive interventions and their implementation in regular care. In childhood, parenting is an important risk or protective factor for child development, and many programs to improve parents’ parenting has been developed used as preventive interventions. Using an ecological approach to prevention and the prevention research cycle as the theoretical framework this dissertation aim to investigate: 1) the long-term effectiveness of four parenting programs (Cope, Comet, Connect, and the Incredible Years); 2) whether the programs work better for some compared to others; 3) if it matters where parents attend the programs; 4) the field of implementation research regarding group based parenting programs; and 5) implementation challenges specifically related to such parenting programs. Overall, the long-term results reveal that there are no significant difference in effectiveness across the programs over time. Also, the programs does not seem work better for some compared to others, and neither does it seem as if program effectiveness is much influenced by the sectors delivering the programs (child and adolescent psychiatry, social care, or school). Concerning the implementation of preventive interventions the research base is small, and conclusive evidence concerning implementation aspects of group based parenting programs are non-existing. Thus, it was not possible to draw firm conclusions about their implementation. Nonetheless, existing research clearly suggest that program specific challenges can influence the implementation of group based parenting programs. Implications for practice and research are discussed, for instance, adaptations to the prevention research cycle.

Keywords: Prevention, child, mental health, parenting programs, effectiveness, implementation, follow-up, sectors.

Viveca Olofsson, School of Law, Psychology, and Social Work
Örebro University, SE-701 82 Örebro, Sweden, viveca.olofsson@oru.se
Acknowledgement

I would like to take this opportunity to express my gratitude to each and every one with whom I have lived, worked, studied, or just enjoyed collegially.

The road to the completion of this thesis has been winding, challenging, and interesting. I have learned much and experienced many aspects of research and academia which has enriched and developed me as a person.

In the preparation of this thesis I have worked in several projects: 7-schools, FCU, Solna, Connect (U), and NJF. As a doctoral student in these project and at the University I have had the pleasure of working with research leaders, research colleagues, research assistants, group leaders, test leaders, co-authors, administrators, opponents, students, and parents.

I want all of you to know that you all have been important to me and to the completion of this thesis.

Over the years I have been fortunate enough to study together with several other doctoral students (current and former) within the department of Psychology at Örebro University. Your presence, support, and “pep-talk” has been particularly vital for the completion of this thesis, in particular the support received from the “C-cohort”.

Even though my aim is to avoid mentioning any one in particular, thus not forgetting someone or adding an additional 10 pages (I have been around for a while), I still would like to take this opportunity to thank Maria Tillfors, Lauree Tilton-Weaver, and Håkan Stattin. You have all given me experiences, insights, and a development as a researcher and individual that I did not anticipate when I started out on this journey.

Then - My family – my great extended one! I know that you not always have understood what I have been up to, or understood its importance (?). Still you have accepted this roller-coaster ride I have taken you on, my ups and downs, and I have felt your support through thick and thin. Älskar!!!! 😊

To all of you mentioned above:

Hereby I recognize and thank each and every one of you
– from the bottom of my heart –
For your understanding, support, knowledge, company,
practical help, and guidance!

Thank you for being there and being you!
List of studies

This dissertation is based on the following studies, which hereafter will be referred to in the text by their Roman numerals:


IV. Olofsson, V., Skoog, T., & Tillfors, M. Implementing Group Based Parenting Programs: A Narrative Review. Manuscript submitted for publication.
# Table of Contents

INTRODUCTION .......................................................................................... 15
Descriptions of the included parenting programs and sectors .................. 16
   The parenting programs included in the NJF project ......................... 17
   *The behaviorally-oriented programs* .................................................. 17
   *The blended program* ....................................................................... 18
   *The attachment-based program* ....................................................... 18
   The evidence base of the included programs ........................................ 18
   The sectors ......................................................................................... 21
Theoretical Framework ........................................................................... 22
   Prevention ........................................................................................... 22
   An ecological approach to prevention ............................................... 22
   The prevention research cycle ............................................................ 23
   *Problem identification* ..................................................................... 24
      Developmental psychopathology ..................................................... 25
   *Program development* ..................................................................... 26
      Risk and protective factors ............................................................. 26
   The influence of parents and their parenting ........................................ 29
   *Reciprocity in the parent-child relationship* ..................................... 35
      Program impact theory .................................................................... 35
      Parenting programs ......................................................................... 37
   *Efficacy trials* ................................................................................ 39
   *Effectiveness trials* ......................................................................... 41
   *Dissemination and Implementation* ................................................ 43
   Gap between preventive theory and research practice ......................... 45
Limitations of previous research ............................................................. 46
   Program effectiveness ........................................................................ 46
   Implementation research .................................................................... 51
This dissertation ....................................................................................... 52
DESCRIPTION OF THE INCLUDED STUDIES ....................................... 53
Methods ................................................................................................... 53
   Studies I, II, and III .......................................................................... 53
   *Participants* .................................................................................... 53
      Participants in study I .................................................................... 54
      Participants in study II ................................................................... 56
      Participants in study III ................................................................... 56
   *Procedure* ....................................................................................... 57
Design ............................................................................................. 58
Analyses ............................................................................................. 60  
Attrition analyses ............................................................................ 60  
Study IV .............................................................................................. 61  
Procedure ............................................................................................ 61
Measures ............................................................................................. 63
Studies I, II, and III .............................................................................. 63  
Child measures .................................................................................... 63  
  Externalizing behaviors (studies I, II, and III) .................................. 64
Parent measures (study I) ................................................................... 65  
  Parents’ negative reactions to their children .................................... 65  
  Parents’ positive reactions to their children ..................................... 66  
  Parents’ sense of competence and emotions ..................................... 66
Results................................................................................................. 67
Study I ................................................................................................. 67  
  The aim of study I ............................................................................... 67  
  Findings from study I .......................................................................... 68
Study II ................................................................................................ 69  
  The aim of study II ............................................................................. 69  
  Findings from study II ......................................................................... 69
Study III .............................................................................................. 70
  The aim of study III ............................................................................ 70  
  Findings of study III ........................................................................... 70
Study IV .............................................................................................. 71
  The aim of study IV ............................................................................ 71  
  Findings of study IV ........................................................................... 71
DISCUSSION .......................................................................................... 75
Adding new knowledge to prevention research ....................................... 75
  Long-term effectiveness ...................................................................... 75  
  Do socio-demographic characteristics influence improvements? ........ 79
  Does it matter where parents participate? ........................................... 79
Implementation research on group-based parenting programs ................ 80
  Were there format-specific implementation challenges? ................. 81
Limitations and strengths .................................................................... 82
Implications for theory ......................................................................... 86
  The ecological model ........................................................................... 86  
  The prevention research cycle .............................................................. 89
Implications for practice ...................................................................... 92
Introduction

Many children and youth suffer from mental health problems, and so the need for prevention has become increasingly recognized. Today, approximately 10–25% of children and youth growing up in the Western world suffer from mental health problems such as depression, anxiety, social and emotional difficulties, and disruptive or aggressive behaviors (e.g., National Research Council and Institute of Medicine, NRC/IOM, 2009; SBU, 2010; World health Organization, 2013). Resources within welfare systems are commonly scarce, and as a consequence, only a fraction of the children suffering from these problems receive the support they need (e.g., Burns et al. 1995; NRC/IOM, 2009). The overwhelming needs of the children combined with the welfare systems’ limited possibilities to meet these needs prompted the development of preventive efforts in order to “nip the problems in the bud”. In other words, when too many needed support from too few, the answer was to try to prevent the needs from occurring or escalating.

The stance of this dissertation is based on the recent findings of a Swedish national review of the available preventive interventions targeting children’s emotional health problems conducted by the Swedish Council on Health Technology Assessment (SBU) on behalf of the Swedish Board of Health and Welfare (SBU, 2010). These findings showed that about 200 preventive programs, of which 125 were manualized, were used to prevent children’s mental health problems. The majority of the programs targeted children’s externalizing problems (e.g., defiance and disruptive and/or aggressive behaviors). Only a minority of the preventive interventions had been evaluated in effectiveness trials; that is, in regular care settings. Long-term follow-ups of program effects were rare. At the time of the report, only a few programs had been evaluated in a Swedish context and no program had yet shown a sufficient evidence base regarding efficacy or effectiveness in Sweden (SBU, 2010).

The results from this report and the political stance of the Swedish government in 2009, which increased research funding particularly towards preventive research, resulted in the funding of the National Effectiveness Evaluation of Parenting Programs (NJF project). The purpose of the NJF project was to evaluate three existing and currently used group-based parenting programs (Comet, Cope, and Incredible Years) within regular care (effectiveness evaluation). These programs have similar theoretical underpinnings (behavioral), and therefore a fourth group-based parenting pro-
program, Connect, with a contrasting theoretical base (attachment) was introduced. These programs all targeted parents’ parenting (e.g., use of rewards, rule setting, communication) with the aim of decreasing externalizing problems of children aged 2–12. None of the existing parenting programs targeted internalizing problems (e.g., sadness, low self-esteem, shyness). As a consequence, my work has focused on group-based parenting programs targeting children’s externalizing problems. Three of the included studies evaluate different aspects of effectiveness, long-term and by sector of care.

Another important area of research that influences the effectiveness of preventive interventions is implementation; that is, the way in which a new program is carried out in terms of, for instance, program delivery, recruitment, and other surrounding routines. How these aspects of implementation are carried out does matter, and implementation research shows that the overall effectiveness of a program is affected by how well it is implemented (e.g., Dane & Schneider, 1998; Durlak & DuPre, 2008). During the recruitment of care units, and later when care units recruited parents to the NJF project, I found that the implementation process and recruitment of parents went more smoothly at some care units than at others. For instance, some care units easily recruited enough parents and sometimes even more parents than required by the project, whereas other care units really struggled to reach minimum recruitment. Why? Could it be that it was harder for the care units to implement a group-based parenting program compared to another kind of parenting program (e.g., individual, school-based)? Implementation questions such as this resulted in the fourth article, a review of the current implementation research on group-based parenting programs used as preventive interventions.

The purpose of this dissertation is to fill some of the gaps in present preventive research concerning group-based parenting programs relating to program effectiveness and implementation. As shown in the SBU report (2010), systematic research on long-term program effectiveness is lacking within the Swedish context, which makes the findings in this dissertation particularly important.

**Descriptions of the included parenting programs and sectors**

Below I provide a short description of the group-based parenting programs, their current evidence base, and the sectors which delivered the programs as part of their regular service.
The parenting programs included in the NJF project

The four programs included in the NJF project and in this dissertation are manualized, group-based programs where parent training is the only component. Comet and Incredible Years are based on behavioral and social learning theories, Connect is based on attachment theory, and Cope is blended, incorporating both behavioral and social learning theories with family system theory and group theory. They share common features such as role play and group discussions, and they are all delivered in 10–12 weekly sessions. They differ on some key aspects: group size, dosage (amount and length of the sessions), reflections (e.g., on parenting, the parent-child relationship), inclusion of individual sessions, (tailored) homework, techniques taught to parents, problem solving, and encouraging support between parents within the group (i.e., networking). A brief overview is provided in Table 1.

The behaviorally-oriented programs

Comet and Incredible Years are behaviorally-based, and techniques for positive parenting practices and consistent rule setting are core components of these programs (Kling, Forster, Sundell, & Melin, 2010; Webster-Stratton, 1984). Both programs aim at promoting positive parenting behaviors, increasing consistent rule setting, and decreasing negative parenting behaviors.

The version of the Incredible Years parenting program used in this project is aimed at parents of children aged 2–8 with conduct problems. Parents of 6–8 children meet for 12 weekly 2-hour sessions in which several video vignettes of everyday interactions between parents and children are used as a base for parents to discuss the interactions. The parents also role play interactions which are then discussed in the parenting group (Kling, et al. 2010; Webster-Stratton, 1984). After each session the parents receive homework for the next week’s session.

The Comet program builds upon Incredible Years to a large extent, and shares some common features and theoretical orientation (Kling, et al. 2010). However, Comet was developed for a Swedish context. It also includes an individual session (the 6th), in contrast to the other programs. Comet is aimed at parents of children aged 3–11. A group consists of a maximum of parents to 6 children, who meet for 11 weekly 2-hour sessions. As in the Incredible Years programs, the parents receive homework, but sometimes the homework is tailored to meet a specific need of a specific family (Kling, et al. 2010).
The blended program

Cope is based on family system theory (Bowen, 1978) and group theory in addition to behavioral theory (Cunningham, Bremner, & Boyle, 1995). Family system theory is a systemic perspective where each family member is seen as intertwined with the other members. Hence, in order to bring about change, the whole family and the family context must be considered. In the present study, we categorized Cope as a non-behavioral program. Cope does not actively provide specific techniques as Comet and Incredible Years do, and the parenting group is quite large; up to 30 parents of children aged 3–12 can attend the 10 weekly 2-hour sessions. However, the larger group of parents is divided into sub-groups of 3–4 parents. These smaller groups then discuss problems and provide their own solutions via a group problem solving process (Cunningham, Davies, Bremner, Dunn, & Rzasa, 1993). The Cope program also uses homework.

The attachment-based program

Connect is based on attachment theory and aims to enhance secure parent-child relationship. The parenting groups in the project consisted of a maximum of 12 parents to children aged 8–12. The groups meet for 10 weekly 1-hour sessions. Parents are encouraged to “take a step back” and reflect on different aspects of their relationship with their child as well as on the meaning of the child’s behaviors (Obsuth, Moretti, Holland, Braber, & Cross, 2006). The intention is to enhance parents’ ability to understand and use developmentally appropriate ways to interpret and address child behaviors.

In contrast to the other programs in the project, Connect does not provide homework for the parents. Instead, at the end of each session, the parents receive a handout containing some key aspects of the current session and its main messages.

The evidence base of the included programs

At the time of the SBU report (2010), only Incredible Years was judged to have a base of evidence from both efficacy and effectiveness trials, though this was not in a Swedish context. The evidence base for the three remaining programs, Comet, Connect, and Cope, was insufficient overall, but encouraging. Since the SBU report (2010), several articles evaluating the programs have been published, including a meta-analysis specifically summarizing findings from evaluations of Incredible Years (Menting, de Castro, & Matthys, 2013). This meta-analysis included 39 articles published up to April
2010 with, in total, 50 efficacy and effectiveness trials where Incredible Years was compared to a comparison group (Menting, et al., 2013). The main aim was to investigate overall effectiveness, and so the meta-analysis did not differentiate between efficacy (i.e., under ideal circumstances) and effectiveness trials (i.e., real life circumstances) or between dependent and independent trials (i.e., with or without the involvement of the program developer/s). The results from parental reports showed small-to-medium effects (Cohen’s $d$) on children’s disruptive behaviors in comparison to the comparison group. The findings are interesting but must be interpreted with caution, since the included studies represent both efficacy trials and effectiveness trials. Further, the design (e.g., randomization, quasi-randomization, and non-randomization), and the comparison groups (e.g., receiving other treatment, waitlist) differed across evaluations. Hence, even though there is an existing meta-analysis on the effectiveness of Incredible Years, this does not provide actual information on what to expect of the program when it is implemented within regular care.

Short-term results have recently been published from the quasi-randomized control trial, the NJF project, with 908 participating parents of children aged 3–12 (Stattin, Enebrink, Özdemir, & Giannotta, 2015). The results showed that all four programs were effective immediately after program end when compared to the waitlist control condition (a thorough description of the procedure, participants, and so forth is provided in the Methods section below). The primary outcome measures of children’s externalizing behaviors decreased significantly, with short-term effect sizes ranging from small to medium (ECBI, Cohen’s $d$ = .17- .63; SNAP IV (ADHD-symptoms), $d$ = .01-.26). In comparison to the more behaviorally-oriented programs (Comet, Incredible Years, and Cope), the children in the attachment-based program Connect showed slightly smaller decreases in problem behaviors, and Connect did not show significant short-term reductions of ADHD symptoms. Hence, even though the programs varied slightly, all four decreased children’s externalizing behaviors in the short term in comparison to the waitlist controls.

Parents’ reported improvements in parental behaviors and well-being were generally small and not consistent across the four programs. The parents’ attempted understanding did not change significantly across any program, but parents in all four programs increased their sense of competence and decreased their depressive symptoms. The parents participating in Cope did not change their use of rewards, but the parents in Comet, Connect, and Incredible Years increased their use. Also, only parents in Comet showed
immediate significant decreases in negative parenting behaviors. In sum, even though parenting behaviors and well-being were affected by the programs, the immediate results were less consistent across the programs in comparison to the effects on children.

The short-term evaluation also tested for possible moderating effects of socio-demographic variables, but none of the tested variables (e.g., socio-economic status [SES], age, gender, immigrant status) influenced the immediate results of the programs (Stattin, et al., 2015). In sum, then, the short-term evaluation comparing the effectiveness of Comet, Connect, Cope, and Incredible Years with a waitlist control as well as with each other provides evidence of program effectiveness when the programs are delivered in regular care.

Adding to the evidence are a number of recently published program-specific randomized evaluations: one of Comet (Kling, et al., 2010) and several of Incredible Years (e.g., Axberg & Broberg, 2012; Charles, Bywater, Edwards, Hutchings, & Zou, 2013; McGilloway, et al. 2014; Perrin, Sheldrick, McMenamy, Henson, & Carter, 2014). A recent effectiveness evaluation of Comet, conducted in a clinical setting with 159 children aged 3–10, showed large effect sizes on children’s externalizing behavior and parents’ parenting practices in comparison to the waitlist controls immediately after program end (Kling, et al., 2010). Parents’ sense of competence also improved, but only slightly. Further, the age of the mother seemingly moderated the effects, with children of younger mothers improving to a greater extent compared to other children. Investigation of the mediating role of harsh and inconsistent parenting practices and positive parenting practices revealed that larger reductions in parents’ harsh and inconsistent parenting and larger increase in positive parenting had a positive influence on the program’s effects on children’s externalizing problems. The effects remained at the 6-month follow-up (Kling, et al., 2010). Thus, as expected, improved parenting behaviors influenced child behavior positively, and the overall improvements were larger for children with younger mothers. Apart from the short-term NJF evaluation (Stattin, et al., 2015), Connect and Cope have not been evaluated in randomized trials published after 2010, and Comet has only one additional randomized control trial. Hence, even though the results are promising, additional randomized effectiveness studies of these programs, in varied settings, would be beneficial. Until then, the evidence base for Comet, Connect, and Cope must be considered small, particularly in contrast and comparison to the constantly growing body of evidence concerning Incredible Years.
Implementation aspects are rarely included in efficacy and effectiveness evaluations even though aspects of program fidelity (e.g., integrity, adherence), adaptation, and attendance are commonly reported and discussed. This was also the case with the NJF evaluation, as it was designed as an effectiveness evaluation. The adherence ratings in the national evaluation were relatively high, with an average of 7.93 (range: 6.86–9.04) on a 10-point rating scale, and fidelity was satisfactory. Incredible Years had the lowest ratings and Cope had the highest (Stattin, et al. 2015). However, to date neither the reasons for the different ratings nor the impact of fidelity on program outcomes have been further evaluated. This is not an uncommon way to report, and as a consequence the potential influence that fidelity, adaptation, attendance, and other implementation aspects (e.g., adoption, feasibility, cost) might have on program effectiveness has yet not been evaluated with enough scientific rigor. Hence, the evidence base concerning the influence of implementation on the effectiveness of Comet, Connect, Cope, and Incredible Years is currently insufficient.

The sectors

The sectors included in this dissertation are the sectors that primarily deliver group-based parenting programs within the Swedish welfare system: child and adolescent psychiatric care, social care, and schools (including preschools). The three sectors function at different levels within this welfare system, and have different mandates and ways to organize their work. Child and adolescent psychiatric care is administered by the county council’s health and medical care department (the National Board of Health and Welfare, 2009), whereas social care and schools are administered by the local municipalities. Thus, the three welfare sectors carry out their work at different organizational levels. Further, as the most specialized sector, child and adolescent psychiatric care delivers the programs to those with problems of clinical magnitude, and these children need to be registered as clients at the care unit. Hence, in child and adolescent psychiatric care the programs are mainly given as indicated prevention.

In contrast, registration is not a requirement when parents attend parenting programs at either social care or schools. The main missions of both social care and schools are broader compared to child and adolescent psychiatric care. Social care provides individual support and family support, such as economic support and family counseling (The Social Services Act 2001:453). Within the educational system, the school health organization works with the aim of sustaining and improving child mental and physical
health at individual schools (The National Board of Health and Welfare, 2004). Within social care, the programs can be given as universal, selective, and/or indicated prevention, while schools commonly offer the programs universally, for instance to all parents in a particular school. In sum, the three sectors offer the programs at different levels within the welfare system, and their preventive approaches and missions with the programs differ to some extent.

Theoretical Framework

Prevention
Prevention refers to efforts to prevent the occurrence of mental health problems (or disorders). In this dissertation I will use the term “preventive interventions” when describing such efforts. In order to reduce the risk of future mental health problems, preventive interventions can aim to alter the development of mental health problems so that the level of the existing problems decreases, or does not continue to escalate (NRC/IOM, 2009). The preventive interventions can be given universally (to everyone in a population), selectively (to sub-group/s with markers of heightened risk), and/or as indicated prevention (to individuals with identified symptoms of mental health problems). In order to enhance the impact of preventive interventions, previous research suggests that it is beneficial to combine the different approaches in various ways (universal, selective, and indicated) and also to target multiple contexts or levels simultaneously (e.g., family, peers, school; Nation, et al. 2003; National Institute of Health, 2004; Wagener, Tubman, & Gil, 2004; Winters 2004). Hence, the prevention of mental health problems can involve preventive interventions at various levels and in various contexts.

An ecological approach to prevention
Prevention interventions can involve multiple levels and contexts, and to illustrate this previous research has adapted an ecological systems approach to prevention (e.g., NRC/IOM, 2009); see Figure 1. All levels within the ecological system (society, organization, family, and child) influence each other as indicated by the arrows. The development, implementation, and delivery of preventive interventions are influenced by, for instance, the current societal culture, how problems are identified and defined, the characteristics of the organizations that might deliver the program, and the targeted recipients (parents and/or children). As a consequence, prevention is
a complicated process with many considerations to be made before a program is developed and effectively implemented in regular care. The ecological circle is the broader, outer framework of my dissertation which helps me to consider all levels and their intricate influence on each other.

Figure 1. Ecological approach to preventive intervention and implementation

![Ecological approach to preventive intervention and implementation](image)

Adapted from Bronfenbrenner, 1979; Fixsen et al. 2005

The prevention research cycle
A narrower theoretical frame that I will use in this dissertation is the established theoretical approach to the development and implementation of preventive intervention presented in Figure 2: the prevention research cycle (Mrazek & Haggerty, 1994). In short, this theoretical approach defines the iterative process of program development through five stages: 1) problem
identification, 2) program development, 3) efficacy trials (under ideal scientifically controlled circumstances), 4) effectiveness trials (in real life settings), and 5) dissemination and implementation of the programs. As my aim is the effectiveness and implementation of group-based parenting programs, my studies focus on the fourth and fifth stages (see Figure 2). However, in order to explain the preventive research process, the whole sequence of the so-called preventive research circle will provide the narrower framework for this dissertation. Therefore, I give a more detailed description of the different stages in the following sections.

Figure 2. The preventive intervention research cycle\(^1\)

1 Adapted from Mrazek & Haggerty, 1994

Problem identification
When does a child’s behavior become an “externalizing problem behavior”? The answer to this lies within the child’s culture, which decides what is considered normal behavior and what is not. Despite cultural differences, normative child development and developmental psychopathology can provide some background and answers.
Developmental psychopathology

Developmental psychopathology (Achenbach, 1974, 1990; Cicchetti, 2010) represents an integrated approach to development, both normal and abnormal. Over the past century, several developmental theories (e.g., behavioral and social learning models, Bandura, 1986; psychoanalytic models, Erikson, 1950; Freud, 1905; Greenberg & Fisher, 1996; cognitive models, Piaget, 1981; and family systems models, Minuchin, 1974) have tried both to explain child development and to provide reasons why children develop mental health problems, or psychopathology. These different developmental models are somewhat complementary; they share some common features and differ on others. To describe them is beyond the scope of this dissertation, but none of them, separately or together, provide an exhaustive explanatory model of child development. Developmental psychopathology, however, integrates aspects of these psychological models of child development together with perspectives from such diverse disciplines as neuroscience and anthropology (Cicchetti, 2010; Wenar & Kerig, 2000). In sum, developmental psychopathology provides an over-arching meta-theoretical approach to child development. As such, it also gives a holistic perspective which helps to paint the “bigger picture” of how child development might go wrong.

Developmental psychopathology has a lifelong perspective. It is guided by a holistic approach in which development is integrated within a dynamic system, with all developmental domains (e.g., cognitive, spatial, and emotional) interacting with each other (Cicchetti, 2010; Wenar & Kerig, 2000). Within the same framework, developmental psychopathology recognizes normative and abnormal development as opposite ends of a continuum, where abnormal development is seen as normative development gone wrong. Development is considered to be hierarchical, which means it passes through stages, or developmental milestones. This approach takes the emphasis away from age. Instead, each stage has developmental tasks, stage-salient issues, which need to be mastered (Cicchetti, 2010; Wenar & Kerig, 2000). Hence, the lifelong perspective of child development as a hierarchical and stage-wise process provides a holistic perspective to children’s healthy (and unhealthy) development.

Healthy and successful development in the earlier stages (childhood) is viewed as a precondition for smooth and successful development in the later ones (Cicchetti et al., 1988). Experiences or effects at earlier stages of development are carried forward and influence the next stage of development. If an experience is negative and has a negative effect on how a child masters a
developmental task at a specific stage, it might lead to increased risk for abnormal development or mental health problems (Dodge, et al. 2008). However, the consequence of an unresolved developmental stage is not deterministic; rather, the probability of future problems increases (Cicchetti, 2010; Wenar & Kerig, 2000). Also, a successful resolution in the earlier stages does not guarantee that future developmental problems will not occur. A healthy and successful accomplishment of significant developmental tasks (e.g., emotion regulation, socio-emotional competencies, perspective taking) at the early stages in a child’s life provides a solid base that is essential for future development (NRC/IOM, 2009). In conclusion, then, success in accomplishing early developmental tasks does not guarantee later healthy development, just as failure does not by definition lead to the development of mental health problems or disorders.

Fortunately, failure to meet the demands is not deterministic, but instead it is possible to positively influence later development. Each developmental stage is seen as a period in which specific developmental tasks or competencies are the most open to change, at the same time as they are the most vulnerable if exposed to potential risk factors such as poverty or poor parenting (Cicchetti, 1993). Nevertheless, given life span development, it is possible later to alter or set right development that has gone wrong at earlier stages (Wicks-Nelson, & Israel, 1997). Therefore, emphasis is put on the understanding of both normal and abnormal development as well as risk and protective factors related to development (Cicchetti, 2010; Wenar & Kerig, 2000). When we understand these we can both treat mental health problems and preferably prevent them. Hence, when a problem is identified, such as externalizing problem behaviors, it is time to investigate theoretical explanations, to ask which risk and protective factors might be involved or need to be targeted, and to develop a preventive program.

Program development
In the second step, theoretical developmental underpinnings and relevant risk and protective factors are considered and the foundation of a preventive program is forged.

Risk and protective factors
Contextual risk factors, such as low SES (low income, education, unstable or unsafe housing) and/or poor parenting, are examples of contextual factors that precede the mental health problems they are associated with (Kazdin, Kraemer, Kessler, Kupfer, & Offord, 1997). Risk factors can also
be internal; that is, residing within the individual themself, for instance personality traits such as dysfunctional emotion regulation and avoidance (suppression and/or worry/rumination), and/or the individual’s gender (Kraemer, Kazdin, Offord, Kessler, Jensen, & Kupfer, 1997; Mrazek & Haggerty, 1994). Risk factors pose a potential threat to anyone who experiences them in a specific situation or context (Kazdin, et al., 1997). Experiencing a single risk factor might not have a great impact on development, but the influence and effects of risk factors is often thought to be cumulative and lead to a cascading development of future problems (Burt, Obradovic, Long, & Masten, 2008; Dodge, et al., 2008). For instance, being a boy increases the risk of externalizing problems, but this in itself does not predispose every boy to externalizing problems, and it does not mean that girls cannot experience or express externalizing problems. It is the combination of this with other risk factors such as low SES, poor parenting, and poor emotion regulation that increases the risk for mental health problems including externalizing problems. In line with the previous example, a boy born in a low income family, which might enhance the risk of his experiencing poor parenting, could experience a negative effect on his development of emotion regulation and social competence. Not developing this developmental task appropriately and in a timely way might, in turn, lead to the boy showing externalizing behaviors such as aggression or defiance in his contact with other children and with adults. These behaviors might influence his academic achievements, and possibly make him drop out of school later. Hence, not only can risk factors accumulate for one individual, but they can also influence child development over time in a cascading fashion (Burt, et al., 2008; Dodge, et al., 2008). In sum, being exposed to one or more risk factors increases the predictive power of future mental health problems (Kazdin, et al., 1997). Hence, if multiple risk factors are or become present within children’s ecological contexts, these risk factors increase the probability of future problems.

In contrast to risk factors, protective factors such as coping skills, education, and health decrease the chance that mental health problems will occur (Wenar & Kerig, 2000). In addition, protective factors also decrease or buffer the influence of risk factors, and enhance the probability of desirable development. Similarly to risk factors, the influence of protective factors is cumulative (additive) and what is stated about risk factors (contextual/internal, predictive power) can commonly be extended to protective factors (Burt, et al., 2008; NRC/IOM, 2009; Rutter, 2013). Thus, the presence of risk does not automatically doom children to abnormal development. A
healthy parent-child relationship, for instance, can serve as a protective buffer for a child growing up in a high risk neighborhood (low SES, high crime and drug rates). Hence, the presence of protective factors can decrease, buffer, and potentially alter development.

It is not always clear cut whether a factor is a risk factor or a protective factor. A specific factor can pose a risk in some contexts and/or under some circumstances and act as a protective factor in others. Gender can serve as a good example. On the one hand, being a boy is related to heightened risk for externalizing problems (e.g., Leadbeater, Kuperminc, Blatt, & Hertzog, 1999), and, as a consequence, being a girl can be seen as a protective factor. On the other hand, in terms of internalizing problems, the relationship is the opposite; being a girl is a potential risk factor (Leadbeater, et al. 1999) while being a boy is a potential protective factor. Another example is shyness, which is related to feelings of loneliness (e.g., Neto, 1992) and of not being socially accepted (Leary & Buckley, 2000). These are known risk factors for later mental health problems (Benzies & Mychasiuk, 2009; Lösel & Farrington, 2012). However, being shy protects youth from engaging in problem behaviors (van Zalk, Kerr, & Tilton-Weaver, 2011). Hence, whether a factor (contextual or internal) acts as a risk or as a protective factor is determined by contextual and circumstantial factors at a specific moment in time (Kazdin, et al., 1997; Lösel & Farrington, 2012). Thus, the same factor can pose a risk under some conditions but be protective in another context or at another time. The relationship is complex.

Risk and protective factors are often seen as opposite ends of a continuum (Luthar, 2003; Rutter, 2003). As a consequence, they ought to pose different degrees of risk and protection depending on where they lie between these two ends. Unfortunately, the children exposed to cumulative or cascading (high) risks often lack the necessary protective factors (NRC/IOM, 2009). This is understandable, given that risk and protective factors ought to be mutually exclusive (as opposite ends of the same phenomenon). A child exposed to risk is not automatically doomed to developmental problems, but similarly a child who experiences several protective factors is not necessarily safe. Hence, it is important to remember that if risk and/or protective factors are present, they increase the probability of either unhealthy or healthy development, but they are not deterministic. Therefore, one focus of program development is commonly to increase the influence/impact of protective factors and, if possible, to decrease or prevent the escalation of one or more risk factors.
Also important, particularly in terms of prevention, is that some risk and protective factors are changeable. In contrast to fixed factors such as gender and ethnicity, it is possible to influence and change factors such as parenting, education, and SES (on a societal level). Parents and parenting, for instance, are recognized as one of the most influential factors on development during childhood (add supporting citation). Hence, one way to decrease the development of children’s externalizing (and internalizing) problems during this age period could be to improve parenting through education and/or training.

The influence of parents and their parenting
It is important to remember that parents and their parenting vary depending on the individual, family, society, and overall culture. These influences can be illustrated in Bronfenbrenner’s ecological model and the family systems approach.
Bronfenbrenner’s (1979) ecological theory offers a broad framework of how child development, and children themselves, are affected by and affect their surrounding contexts. As seen in Bronfenbrenner’s ecological model (Figure 3), four significant components are taken into account: the developmental process, the child (or person), time (sequences/temporality/socio-historical), and context. The context in which development occurs is nested around the child at four levels, or systems, with the fifth system, the chronosystem (time), crossing and connecting the others. The chronosystem constitutes of all the experiences the child has over the life span, such as school transitions, birth of siblings, graduation, or environmental events. The fourth and the broadest contextual level, the macrosystem, is defined by the overall culture in which the child and their family function. Its influence is

---

1 Adapted from Bronfenbrenner, 1979
indirect. The third level, the exosystem, constitutes the closest context that indirectly (distally) influences the child, for instance the parents’ workplace or the management at school. Next, the mesosystem constitutes all the sets of one-to-one microsystems the child has with others and their mutual interactions. For instance, the family constitutes the child’s one-to-one interactions with their parents and siblings. Another example could be the child’s class at school, which constitutes classmates and the teacher. Together these form the child’s mesosystem. The contextual level closest (proximal) to the child is the microsystem. This is the level where the child is at a specific moment in life with their direct one-to-one individual interactions with others. The parent-child relationship is an example of a microsystem.

As an example of the bi-directionality of the ecological model, the child’s development is influenced by the macrosystem through, for instance, school policies, community culture, and government. In turn, the macrosystem is influenced by the child’s mesosystem when, for instance, parents attend parent meetings or vote in elections. Hence, child development occurs through the mutual influence the levels exert on each other, with the children and their microsystems at the center of the model (Bronfenbrenner & Morris, 2006). Though never really tested in its whole, Bronfenbrenner’s ecological system theory provides a framework through which development can be conceptualized.

As seen in the ecological model, the child is directly influenced by the closest contexts, the microsystems and the mesosystem. Family systems theory (Belsky, 1981) provides a framework which helps explain the important mesosystem (and microsystems) of the family. The theory suggests how individuals within a family affect and are affected by each other, as spouses, parents, children, and siblings. It provides a model for how families as a system “assimilate” or “accommodate” to change. Although the focus has been on the understanding of adult functioning, family systems theory provides a framework for the interdependence within the family context. It specifically focuses on communication, power, control, and relationship structures within the family (Cox & Paley, 1997; Minuchin, 1985). Family systems theory and Bronfenbrenner’s ecological model complement each other, and both models will help us conceptualize how the context influences child development and ultimately how it relates to the prevention of mental health problems.
In some families, parents provide their children with genes, but it is not only parents’ genes and characteristics that influence children and their development. The parents’ relationship with each other and their individual life experiences are carried forward into both the family context and the parent-child relationship. For instance, parents who have high levels of marital discord are not as warm and supportive toward their children as parents whose interactions with each other are less dramatic (Erel & Burman, 1995; Park & Buriel, 1998). What parents bring with them has consequences; for instance, marital discord has been shown to be a risk factor for child depression, anxiety, and delinquent behaviors (Harold & Conger, 1997). On the other hand, the opposite, low marital discord, serves as a protective factor.

Overall, children in warm, responsive, nurturing families with well-functioning communication patterns and consistent behaviors (e.g., rules and regulations) seem to experience the best conditions for child development (e.g., Minuchin, 1985; Parpal & Maccoby, 1985). However, these family characteristics are not either/or; instead, each of them might be present in the family and the parents to “a degree”. In order to prevent children’s mental health problems, it is therefore important to recognize, first, what signifies poor/healthy parenting; second, what are the consequences of poor/healthy parenting; and third, how poor parenting can be improved.

Two closely related theories and concepts that greatly influence the overall understanding of parenting and its influence on child development are attachment (Ainsworth, Blehar, Waters, & Wall, 1978; Bowlby, 1969/1982) and parenting styles (Baumrind, 1973; Maccoby & Martin, 1983). These concepts have been studied extensively, and provide an established theoretical base for parenting and how parenting can influence child development. “Attachment” is a term used to describe the relationship that exists between a child and its parent. It is first developed in infancy, normally when the child is approximately 6–8 months old (Bowlby, 1969/1982). Attachment can be either secure or insecure (Ainsworth, et al. 1978). Secure attachment is defined by the child using the parent as a “safe base”. That is, the parent responds to the child in a predictable way, and the child knows that the parent will be there for support and comfort. This basic trust enables the infant to leave the safety of the parent and explore the environment, and to also seek out the parents if separated from them (Ainsworth, et al. 1978; Carlson & Sroufe, 1995).

Insecure attachment, on the other hand, is often divided into three categories: avoidant, ambivalent, and disorganized (Ainsworth, et al. 1978;
This can be seen in a situation where the caregiver leaves the child alone in a room and a stranger enters (Ainsworth et al. 1978). The child, when reunited with the caregiver, behaves in different ways, either avoiding contact with the caregiver (avoidant), not being reassured by the return of the caregiver (ambivalent), or seeming confused when approaching the caregiver (disorganized) (Ainsworth et al. 1978; Carlson & Sroufe, 1995). However, the common denominator to the different categories of insecure attachment is the low level of responsiveness from the parent/caregiver in the parent-child relationship. As the child develops, so does the parent-child relationship, and the characteristic (age-related) signs of attachment (secure and insecure) change. However, the bond or relationship is still important, and influences life span development. Secure attachment is related to, for instance, better social skills, more emotional maturity, less aggression, lower depression levels, and lower levels of externalizing behaviors (e.g., Allen, Porter, McFarland, Marsh, & McElhaney 2005; Carlson & Sroufe, 1995) compared to children with insecure attachment. Hence, growing up in a healthy and responsive parent-child relationship characterized by secure attachment is a protective factor for later mental health problems, whereas the opposite is a known risk factor. Fortunately, attachment is malleable and can be improved. Therefore, attachment often serves as a foundation and active ingredient (overtly or covertly) in the development of parent training and parenting programs. For instance, parenting programs based on behavioral theories (e.g., Kumpfer, DeMarsh, & Child, 1988; Molgaard & Spoth, 2001; Patterson, 2005; Sanders, 1999; Webster-Stratton, 1984) encourage parents to spend more time with their children and also encourage and teach parents proper techniques to play and interact with their children, thus strengthening the parent-child relationship and improving attachment. However, improving attachment is not an outspoken aim or component of these programs.

Attachment and parenting style are closely related. Parenting style consists of the patterns of higher or lower levels of parental responsiveness, warmth, and control. Combinations of these patterns can be used to categorize parents into four different parenting styles: authoritative, authoritarian, permissive indulgent, and permissive neglectful (Baumrind, 1973; MacCoby & Martin, 1983). However, these parenting styles are commonly conceptualized as dimensions labeled “control/demand” and “acceptance/responsiveness”.

- The authoritative parent is high on nurturance, communication, maturity demands (level of expectation), and control (clear and
consistent rule setting). Hence, these parents are high on both the control/demand and the acceptance/responsiveness dimensions. As one might expect, research shows that children growing up in families with authoritative parents show the most consistent positive outcomes compared to children reared in families with parents characterized by any of the other parenting styles (e.g., Luyckx, Tildesley, Soenens, & Andrews, 2011; Steinberg, Elmen, & Mounts, 1989).

- The authoritarian parent is high on control/demand but low on acceptance/responsiveness; these parents are power-assertive. Children of authoritarian parents have lower self-esteem and do less well in school compared to the other parenting styles (Baumrind, 1991; Maccoby & Martin, 1983). They might also show higher levels of aggression (Caputo, 2004).

- The permissive indulgent parents are low on control/demand but high on acceptance/responsiveness. Hence, they are tolerant and warm but let their children do much as they please. Children reared by these parents tend to be more aggressive, do slightly less well in school, and appear to be somewhat less mature compared to children of parents with different parenting styles (Maccoby & Martin, 1983).

- The permissive neglectful parent is low on both dimensions, being uninvolved or unavailable. Availability is an important prerequisite for the development of attachment between parent and child, and an unavailable parent enhances the risk for the development of insecure attachment. A disturbed and insecure parent-child relationship might account for research showing that these children are less achievement-oriented, more impulsive, and antisocial compared to others (Caputo, 2004; Lamborn, Mounts, Steinberg, & Dornbusch, 1991, Pulkkinen, 1982). Also, the lack of disclosure and therefore knowledge about children’s whereabouts (Kerr & Stattin, 2000) seems to be crucial.

In conclusion then, children who have warm, responsive, authoritative parents have advantages compared to other children. One of these advantages is that they are commonly securely attached. As mentioned, secure attachment serves as a major protective factor for child development whereas the opposite, insecure attachment, serves as a major risk factor. In addition, many disorders seem to stem from disturbances in the parent-child
relationship. Given that parents and their parenting have such a great impact on child development, it is not surprising that many prevention programs either include parenting components (e.g., parenting training) in for instance school-based prevention (e.g., Kratochwill, McDonald, McDonald, Scalia, & Coover, 2009), or directly target families in their efforts to prevent maladaptive child development.

Reciprocity in the parent-child relationship

Children are not only affected by their context or their parents. They are also active agents, influencing their environment and their parents, unintentionally and/or intentionally. All models (i.e., the ecological model, attachment theory, as well as the family systems approach) emphasize mutual influence and reciprocity between parents and their children (Belsky, 1981; Bowlby, 1969/1982; Bronfenbrenner, 1979). As an example, Patterson’s coercion circle, which emanates from family systems theory, is one influential explanation for disruptions in the parent-child relationship (Patterson, 1982). According to this theory, a parent and child might get caught up in a vicious circle where actions by, for instance, the parent (e.g., taking away privileges) spur reactions/actions from the child. The child might perceive the parent’s decision as unfair, become very upset, and throw a tantrum. In response to the child’s reaction, the parent starts screaming at the child. Actions and reactions like these might lead to an escalation of problematic interactions between the parent and the child. If not broken, the behaviors created within the parent-child coercive circle might generate, for instance, externalizing problems such as defiance and aggression (Andrews & Dishion, 1994; Patterson, 1982; Patterson, Reid, & Dishion, 1992). The theoretical explanation provided by Patterson’s coercive circle for how parents and children mutually influence the development of children’s externalizing problems generally serves as another theoretical foundation when parent training and parenting programs are developed. However, the focus of research and program development is mostly on changing the attitudes and behaviors of the parents. The contributions of the child are still, to a large extent, overlooked.

Program impact theory

Program impact theory (Rossi, Lipsey, & Freeman, 2004) can serve as a means to define, develop, and structure the components of a new program. According to this theory, the building blocks or components of prevention programs are included under the assumption that they can alter, in our case,
parents’ behaviors, perceptions, and/or feelings. Influenced by theory and research evidence of, for instance, risk and protective factors, different components (e.g., information, activities) are included in the programs. These components, which might include techniques of positive parenting behaviors (e.g., praise and awards) or reflective thinking for attitude change, are assumed to affect and change the participating parents and their parenting in specific ways (Rossi, et al. 2004). The components are then combined in a manner aimed to enhance their impact. For instance, behavioral parenting programs commonly start by introducing parents to child play and positive parenting behaviors, and then move on to techniques for consistent rule setting, which is considered a pre-requisite for the change of negative parenting. Hence, the essence of program impact theory is that each element included in a program is either a cause or an effect of improved parenting. Also, the elements build on each other. Ultimately, these processes of change are supposed to lead to desired outcomes; in our case, decreased child externalizing problem behaviors.

In terms of program development and in relation to program impact theory, it is also important to consider the level and setting in which the preventive intervention is supposed to be delivered. Concerning parenting programs aimed at children’s externalizing problems, for instance, children’s development is influenced by several factors (e.g., parents, peers, school) and therefore preventive interventions can target the individual, family, community, or society at large (NRC/IOM, 2009). In fact, previous research has shown that one aspect that can result in an effective preventive intervention is that it targets multiple levels or factors simultaneously (e.g., Nation, et al. 2003; National Institute of Health, 2004; Wagener, et al. 2004; Winters 2004). Thus, program developers need to consider more than just the components during program development; the format is also important. Child prevention has commonly been given as individual, group, or combined mixed efforts (e.g., child, parent, and school – either individual and/or group). This is important to remember, even though my focus is on free-standing group-based parenting programs.

In sum then, when a problem has been identified and the theoretical approaches established, taking necessary risk and protective factors into consideration, it is time to develop the specific preventive interventions. This is commonly done by combining theoretically informed components and strategies of change while considering at what level and in what format the intervention is going to be introduced. When this process is completed and
Before we move on in the prevention research cycle, I would like to present some information about group-based parenting programs and their theoretical approaches and structures. First, it is worth noting that parenting programs are often said to be targeted or aimed at children’s externalizing problems (I have not found any group-based parenting program aimed at internalizing problems). Hence, the aim of these programs is to decrease children’s problems. It is rarely, if ever, stated that the actual aim of these programs is to improve parenting alone. However, in line with program impact theory, the desired outcome of the programs is typically changed and improved parenting which, in turn, is supposed to lead to decreases in children’s externalizing behaviors.

**Parenting programs**

As previously mentioned, there are several theories and/or models concerning child development and the development of mental illnesses. These theoretical models provide the bases from which programs are developed. Concerning group-based parenting programs, the behavioral and social learning theories (i.e., behavioral) are predominant (e.g., Lundahl, Risser, & Lovejoy, 2006). Only a few programs have been developed based on other theories/models (i.e., non-behavioral). In addition, the elements included in the programs target important risk and protective factors recognized by research as moderators and mediators of child development (e.g., insecure attachment, harsh parenting, and inconsistent rule setting). In short, the program developer/s choose the theoretical approach and target those risk and protective factors that they perceive ought to influence parenting most effectively. From this foundation, the program components are developed.

**Behavioral and social learning approaches.** As mentioned, the majority of group-based parenting programs are based on behavior and social learning theories (e.g., PMTO, Patterson, 2005; Triple P, Sanders, 1999; Incredible Years, Webster-Stratton, 1984). Behavioral theories are guided by distinct characteristics: observable behavior and learning principles. The learning principles used in the programs are operant conditioning, and imitation (also called modeling or observational learning). Imitation or modeling is mostly used to teach the parents, for instance, new parenting techniques, and is not commonly used as a means for parents to influence their children. However, if receptive, the child might learn new behaviors by observing and
imitating others. According to social learning theory (Bandura, 1986), children learn new behaviors by watching others perform (model) them and then automatically imitate them, but whether the child will actually make use of the behavior depends on whether they perceive the behavior as rewarding. Another important aspect is that the child also considers whether they can actually perform the behavior successfully (Bandura, 1986).

Operant conditioning, on the other hand, is the main base for the parenting techniques parents are taught in the program. Operant conditioning is when a behavior becomes associated with a certain response. Whether the child will continue with a certain behavior, for instance aggressive behavior, is dependent on the reinforcement. The reinforcement can be either rewarding (getting hold of a toy) or non-rewarding (being punished). Hence, the child learns to associate a specific behavior with a specific response. In the program, the parents are taught techniques to reinforce their child’s “good” behaviors by providing rewards (praise, receiving a perceived benefit, such as a star or a treat) and to ignore or correctly non-reward “bad” behaviors (e.g., the behavior being ignored or receiving a consequence).

Active elements in these programs often include:

- Teaching parents to spend time with their child and appropriate ways to interact with them. Hence, the programs encourage parents to interact with their children. Even though not explicitly recognized in the programs, such activities may strengthen the parent-child relationship and enhance attachment.

- Teaching parents to reward children’s “good” behaviors while ignoring their “bad” behaviors. Thus, reinforcement of the desired behaviors should enhance the incitement to behave in a “good” way, and punishment of the undesired behaviors should lead to their extinction.

- Teaching parents consistent rule setting, thus providing consistency in what the child can expect when they behave in a certain way.

The blended and non-behavioral approach. Some programs include other theoretical approaches besides behavioral and social learning. For instance, Cope includes both a family systems approach and a problem solving approach in addition to a behavioral approach (Cunningham, et al., 1995). This program includes the basic elements of the behavioral approach in which positive parenting behaviors are modeled but parents are not taught specific parenting techniques. Instead the program encourages a group problem solving approach, where parents come up with their own solutions.
to encountered or suggested problems. Hence, in addition to the behavioral elements, this blended program also has the element of active problem solving.

**Attachment.** Attachment theory is an approach adapted for other programs (e.g., Connect, Obsuth, et al., 2006; Tuning in to Kids, Havighurst, Wilson, Harley, & Prior, 2009). These programs focus on fostering secure attachment through elements such as:

- Awareness of child behaviors and emotions
- Changing perceptions about child behavior and emotions
- Appropriate communication (understanding, acceptance, and availability)

### Efficacy trials

The third step of the prevention research cycle involves pilot testing of the program or efficacy trials. Here, the developed program and its structure are tested, and often re-tested, until the program developer/s perceive that they have the ultimate structure of the components included in the program (NRC/IOM, 2009). In efficacy trials, the focus is on internal validity; that is, ensuring that the results produced are due to program effects (NRC/IOM, 2009). In order to do this, efficacy trials are conducted under ideal and strictly monitored circumstances, and are often supervised and/or delivered by the program developer/s. Thus, at this stage, the composition of the program is thoroughly evaluated and adjustments to the program are often made.

There is a large body of research on the efficacy of parenting programs aimed at children’s externalizing problems (e.g., Cedar & Levant, 1990; Lundahl, et al. 2006; Maughan, Christiansen, Jenson, Olympia, & Clark, 2005; Serketich & Dumas, 1996; Thomas & Zimmer-Gembeck, 2007). The results from efficacy trials show that some programs involving parent training are efficacious, with moderate-to-large effect sizes immediately after the programs end (e.g., Cedar & Levant, 1990; Lundahl, et al. 2006; Maughan, et al. & Clark, 2005; Serketich & Dumas, 1996; Thomas & Zimmer-Gembeck, 2007). Hence, programs have been found to decrease children’s externalizing behaviors, particularly in the short term.

Because these programs focus on children’s externalizing behaviors, prior research has focused less on changes in parental behaviors and attitudes. Overall, there is surprisingly little efficacy research on how parenting programs work, and how they influence and create change in children’s externalizing behaviors. This is surprising, because it is the changes in parents’
parenting and/or attitudes that are supposed to bring on (mediate) the decreases in the children’s externalizing behaviors. It would be beneficial and informative if future efficacy research analyzed potential mediators on a general basis.

Nonetheless, efficacy research shows significant improvement in parenting behaviors and parental competence and well-being (i.e., depression and stress). The effect sizes are generally small to moderate (e.g., Cedar & Levant, 1990; Lundahl, et al. 2006; Maughan, et al. 2005; Serketich & Dumas, 1996; Thomas & Zimmer-Gembeck, 2007). In conclusion, then, the results from previous efficacy research on group-based parenting programs indicate that the programs can be effective. The children’s externalizing problems decrease, and parenting and well-being can be improved.

However, there is reason for caution because there are some limitations to previous research on the efficacy of group-based parenting programs. The main focus of this dissertation is on the fourth and fifth stages, effectiveness and implementation. However, it is still important here to highlight the limitations of stage three, efficacy, since they spill over to the later stages. First, many of the tested parenting trials have evaluated program effects on children with problem levels within clinical range (e.g., Kling, et al., 2010; Webster-Stratton & Hammond, 1997) or on high-risk samples (e.g., Dishion & Andrews, 1995; Gross, Garvey, Julion, Fogg, Tucker, & Mokros, 2009). With high problem levels and/or enhanced risk comes the possibility of greater change, which can influence the size of program effects (e.g., Hautmann, Eichelberger, Hanish, Plück, Walter, & Döpfner, 2010; Menting, et al., 2013). Further, the current literature is dominated by a few well-established programs (e.g., Furlong et al. 2012), as well as by results from efficacy trials, meaning that it might be difficult to properly draw conclusions about the effectiveness of group-based parenting programs in general. Finally, there is a lack of long-term results; that is, evidence that the immediate effects remain or are maintained over time (e.g., Furlong et al. 2012; Lundahl, et al., 2006; Sandler, Schoenfelder, Wolchik, & MacKinnon, 2011; Smedler, et al., 2015). This is a particularly serious limitation in prevention research, because the purpose of preventive interventions is to prevent future mental health problems. Without the knowledge and information gained from long-term evaluations, it is not possible to conclude whether the immediate effects provide the intended protection against future problems. In sum, previous efficacy research on group-based parenting programs shows that parents and children improve immediately after a program has been delivered. However, the results need to be interpreted with caution, and it
is concerning that there is a lack of long-term knowledge of the preventive abilities of such programs.

Nonetheless, when a program provides significant and sufficient outcomes in efficacy evaluations, it is time to move to the fourth stage in the process: effectiveness trials.

Effectiveness trials

In effectiveness trials, the program is delivered and evaluated in regular service with regular staff. Initially, this is commonly done or supervised by the program developer/s. However, the ideal is to also conduct independent effectiveness trials (Flay, et al. 2005). In order to ensure that the content of preventive interventions is delivered as intended, the programs are often highly structured and manualized, and delivered by specially trained group leaders from the regular staff. In effectiveness trials, the focus is more on external validity, compared to the internal validity focus of the efficacy trial. Studying a program in a real life setting allows for “normal” disruption and interference in the delivery of the program, such as illness, vacations, and group composition, that may not be as present when the program is delivered under the ideal circumstances of efficacy trials. Therefore, the fourth step in the prevention research cycle is important. Like the previous step, efficacy trials, the effectiveness trials provide much-needed information about how the program works and its effects, in this case under real life circumstances. Also, at this stage the program is still commonly adjusted based on both feedback from the effectiveness of the program itself and feedback provided by the group leaders delivering the program and the parents participating in it. Hence, effectiveness trials are important and provide knowledge about how the program functions in the settings it was designed for.

If the program still shows significant positive effects; that is, a decrease in children’s problematic behaviors, in a regular setting with ordinary staff as group leaders, it is considered an effective preventive intervention. Previous research on the effectiveness of group-based parenting programs is, in general, scarce. As was the case with the abovementioned efficacy evaluations, the focus has mainly been on effects on children’s problem behaviors. Measures of child outcomes are usually both the primary and secondary measures (Michelson, et al. 2013) measuring children’s externalizing problems, such as impulsivity, hyperactivity, aggression, and defiance. The outcome measures most commonly used in previous evaluations (both efficacy and effectiveness) are Eyberg’s Child Behavior Inventory (ECBI; Eyberg &
Ross, 1978), the Child Behavior Check List (CBCL; Achenbach & Edelbrock, 1981), and the Strength and Difficulty Questionnaire (SDQ; Goodman, 1997). The use of similar measures enables comparisons across evaluations. Several meta-analyses (Cedar & Levant, 1990; Lundahl, et al. 2006; Maughan, et al. 2005; Serketich & Dumas, 1996; Thomas & Zimmer-Gembeck, 2007) and reviews (e.g., NRC/IOM, 2009; SBU, 2010; Smedler, Hjern, Wiklund, Anttila, & Pettersson, 2015) have shown that the programs decrease children’s externalizing behaviors in the short term. In addition, a recently-published randomized effectiveness study (Stattin, et al., 2015) evaluated the group-based parenting programs Comet, Connect, Cope, and Incredible Years, with 908 participating parents with children aged 2–12, and showed improved child behaviors. Hence, the programs seemingly also improve children’s externalizing behaviors when delivered in regular care. Nonetheless, as discussed below there are some uncertainties considering the child effects of the programs, and little is known about the long-term effects.

In parenting programs, changes or improvements in the risk and protective factors related to parents and their parenting are assumed to function as mediators through which the parenting programs alter child behavior and, ultimately, prevent abnormal development. Clearly, knowledge about how parents are affected is vital, but previous research has put little focus on consistent and coherent evaluations of the parental outcomes (Sandler, et al. 2011). Compared to child outcomes, there is less consensus on parental measures across evaluations, and so the measures vary considerably (e.g., Axberg & Broberg, 2012; Gardner, Burton, & Klimes, 2006; Moretti & Obsuth, 2009; Zubrick et al, 2005). Parental outcome measures are rarely set as primary or even secondary outcomes (Michelson, et al. 2013). Also, it has been common to evaluate the programs’ immediate effects on parents but less common to follow these effects over time. That said, parenting programs seemingly have beneficial effects on parents’ parenting behaviors, feelings, and well-being in the short term (Litschge, et al. 2010; Maughan et al., 2005; McCart, et al. 2006; Serketich & Dumas, 1996; Stattin, et al. 2015). However, as is the case with child effects, less is known about whether the immediate effects on parents change or sustain over time. Furthermore, the mediating role of the parenting components included in the programs has received even less attention, and so the question “How did the programs work?” remains mostly unanswered (Sandler, et al., 2011). It seems plausible to assume that one of the first aspects to investigate is the question of which component/s and their effects on parental behaviors and
emotions influence the change in children’s externalizing behaviors in a parenting program, at least according to program impact theory (Rossi, et al. 2004). In sum, there have been effectiveness evaluations of some of the parenting programs, proving the effectiveness of the programs on both parents’ behaviors and well-being, and children’s behaviors, at least in the short term. Unfortunately, there is a clear knowledge gap concerning program functioning in real life settings, largely because program developers and researchers commonly skip this fourth step of the prevention research cycle and disseminate programs in what might be considered a premature fashion.

Dissemination and Implementation
In theory, then, if the program is effective the last and fifth phase, dissemination and implementation of the program, is initiated. Dissemination refers to the process where knowledge and information about the program are spread to the general public, or practice. Implementation, in turn, refers to the process through which both knowledge about the programs and the programs themselves are spread and put into regular and sustained use in everyday practice (e.g., Fixsen Naoom, Blasé, Friedman, & Wallace, 2005). The implementation activities move back and forth during the implementation process, and so it would probably be beneficial if implementation aspects were considered earlier in program development and the prevention research cycle.

Sustainable implementation is recognized among implementation researchers as a process that evolves over time (Fixsen, et al.; Ogden & Fixsen, 2014). There are several theoretical approaches to implementation. They all share the view of the implementation process being performed in an ecological context in a similar vein as prevention (Figure 1; e.g., Aarons, Hurlburt, & Horowitz, 2011; Durlak & DuPree, 2008; Fixsen, et al.; Wandersman, 2003). As an example, the implementation process is influenced by funding from society, the stakeholders at the different levels (e.g., developers, policy and decision makers, practitioners, and the clients), the program, the agency mission, and so forth. Hence, the implementation of a preventive intervention is a complex, iterative, and multi-level process. Further, it takes several years before an intervention is set and sustainably implemented in an organization’s everyday practice.

Overall, research on the implementation of parenting programs is scarce. Implementation research has taken a broad approach within the prevention field, discussing implementation of preventive interventions more in general.
Child prevention and group based parenting programs (e.g., Aarons, et al. 2011; Dane & Schneider, 1998; Durlak & DuPre, 2008; Fixsen, et al. 2005). Nonetheless, previous research generally concludes that the outcomes of a preventive intervention are significantly influenced by how well the intervention is implemented. As an example, a recent review concluded that intervention outcomes were between two and three times as large where the implementation process was monitored (Durlak & DuPre, 2008). Hence, a well-conducted implementation positively influences the outcomes of an intervention.

To sum up then, I use the prevention research cycle (Mrazek & Haggerty, 1994) embedded within an ecological approach (e.g., NRC/IOM, 2009) as the theoretical framework of this dissertation. The stage-wise iterative process of the prevention research cycle takes the surrounding ecological contexts into consideration in the first stage: identifying a problem, such as children’s externalizing problem behaviors, and its consequences. The second stage involves, for instance, taking culturally relevant risk and protective factors into account when developing a preventive intervention targeting externalizing problems. Parents’ parenting is influential on children’s development, and is consequently often included in interventions aimed at reducing children’s externalizing problems. Third, when an intervention, such as a group-based parenting program, has been developed, it is tested out in efficacy trials. Efficacy trials are conducted under controlled and close to ideal circumstances. In a controlled environment, the program developers may be able to ensure high internal validity when the program is delivered to parents and be largely confident that the outcomes produced are a result of the program. When the intervention has been rigorously tested (and adjusted) it is ready for the fourth stage, effectiveness trials. In effectiveness trials, the intervention is evaluated (and adjusted) under real life conditions to ensure external validity; that is, to ensure that it produces similar outcomes in real life settings as under the ideal circumstances from the efficacy trials. Ideally, in effectiveness trials, the parenting program is delivered by regular staff at local care settings (e.g., social care, school) to parents from their ordinary target population. Finally, when the program has been extensively evaluated both on efficacy and on effectiveness, providing a solid evidence base of overall effectiveness, it is ready for full dissemination and implementation in the final and fifth stage of the prevention research cycle. The prevention research cycle is the established theoretical approach to development and implementation of preventive interventions (NRC/IOM,
2009). Hence, this is, in theory, how a new program is supposed to be developed and implemented. However, this process is not always realized in common research practice.

**Gap between preventive theory and research practice**

Up until the fourth stage, the progress through the prevention research cycle proceeds much according to plan. However, it seems as if research has become somewhat stuck at the third step, namely efficacy trials. The majority of the existing parenting programs have been extensively evaluated in efficacy trials (e.g., Cedar & Levant, 1990; Lundahl, et al. 2006; Maughan, et al. 2005; Serketich & Dumas, 1996; Thomas & Zimmer-Gembeck, 2007), but program developers and other stakeholders (e.g., researchers and representatives of practice) have then appeared to move directly to the fifth stage, dissemination and implementation, without passing through the fourth stage, effectiveness trials. This is of major concern, because the program’s ecological validity has not been tested, and potential problems, for instance related to program implementation, have not been discovered or solved. The theoretical approach of the preventive research circle is as an iterative process, where one stage informs and provides a foundation for the next stage (Mrazek & Haggerty, 1994). Currently, however, the dissemination and implementation seems to be done before its time. Effectiveness trials provide much valuable information. First, if the programs fail to be effective in studies which replicate the efficacy trials in regular care, it is possible that they should not be implemented in the first place (Flay, et al. 2005; Gottfredson, et al. 2015). The reason for failure to replicate outcomes needs to be considered: does it depend on the program or on the implementation of the program? Additional efficacy trials and/or effectiveness trials would provide valuable information. Effectiveness trials in particular would provide information about the practicalities of the program when it is delivered in a regular (intended) setting. Failing to evaluate program effectiveness could potentially mean that issues regarding program implementation remain unknown and unsolved. A majority of the programs that have been disseminated and currently implemented have not been tested in real life settings (NRC/IOM, 2009; SBU, 2010; Smedler, et al., 2015), and so there is a serious lack of evidence of program effectiveness and possible implementation aspects.
Limitations of previous research

Program effectiveness

The question of whether the results of efficacy studies can be generalized into regular care is answered through effectiveness trials. As mentioned above, one serious limitation to current research lies in the overall lack of effectiveness trials, and evaluations with rigorous scientific criteria (e.g., randomization, control condition) are very scarce (Furlong, et al. 2012; NRC/IOM, 2009; SBU, 2010; Smedler, et al., 2015). Even though several reviews and meta-analyses have made comparisons between existing studies, and have estimated the effects of the components of the programs, these comparisons have been based on studies that differ in study population size, settings, design, instruments, and data quality. The comparisons have been made between “apples and oranges” (Furlong, et al. 2012; Sandler, et al. 2011). As examples, prior meta-analyses (e.g., Lundahl, et al. 2006; Maughan et al. 2006) and reviews (Farmer, Compton, Burns, & Robertson, 2002; Nixon, 2002) have included programs with other components in addition to group-based parent training, for instance child intervention (e.g., Lochman & Wells, 2003; 2004) and/or teacher intervention (e.g., Reid, Eddy, Fetrow, & Stoolmiller, 1999; Webster-Stratton, 1998). Targeting several contexts at the same time is known to enhance intervention effectiveness (Nation, et al. 2003; National Institute of Health, 2004; Wagener, et al., 2004; Winters, 2004). Hence, in respect to these study and program factors, the actual effects of the parent training component in these interventions could be overestimated. A similar interpretation difficulty is also present in the recent program-specific meta-analysis including 50 evaluations of Incredible Years, where both efficacy and effectiveness trials with varied scientific rigor were summarized (Menting, et al., 2013). The meta-analysis provided an overall mean effect of the program without differentiating between whether the trials were designed as efficacy or effectiveness trials. This makes it hard to generalize the findings to regular care, even though the meta-analysis was specifically on Incredible Years. As a consequence, the conclusions drawn on program effectiveness in these past overviews are not readily generalizable to specific programs or formats, such as group-based parenting programs.

The limitations identified concerning efficacy trials (high problem levels/high-risk samples, few programs, and lack of long-term follow-ups) are even more prominent in regard to effectiveness trials (Furlong, et al. 2012; NRC/IOM, 2009; SBU, 2010; Smedler, et al., 2015). In addition, the results
from previous research are somewhat disparate. For instance, the Triple P parenting program has failed to show evidence of effectiveness when implemented in regular care, whereas the Incredible Years program shows similar outcomes in effectiveness and efficacy trials (Little et al. 2012). In an attempt to address these disparate results, a recent meta-analysis evaluated eight parenting programs across 28 RCTs (Michelson, Davenport, Dretzke, Barlow, & Day, 2013). The results showed that the studies applying strict real world criteria, and hence close to being effectiveness trials (i.e., clinic-referred sample, routine setting, routine service, and non-specialist therapists) did not show significantly lower effect sizes than those applying less stringent real world criteria and hence close to being efficacy trials (Michelson, et al.). Given the current official recommendation to implement parenting programs in clinical contexts, these findings are encouraging (Flay, 2005). In conclusion, even though recent research shows promising results concerning the transportability of the programs into the “real world”, there are serious limitations to previous research. The research base is limited, with rigorous effectiveness trials of only a few programs. Thus, more effectiveness trials are needed to answer the question of generalizability of the positive outcomes to real life settings, both immediately and, equally importantly, over time.

Research is also lacking regarding aspects that might influence (moderate) program outcomes. Several factors pertaining to the programs or their implementation have the potential to influence program outcomes. Nonetheless, it is largely unknown whether the programs are more effective for some parents and children than others. Again, research is scarce and somewhat contradictory. There is, for instance, a lack of gender comparisons concerning program effects on boys and girls (SBU, 2010), and it is unclear whether SES, age, and other demographic characteristics of parents and children affect the effectiveness of the parenting programs (e.g., Furlong, et al. 2012; Lundahl, et al. 2006; Kling, et al. 2010; Stattin, et al. 2015). For example, Lundahl and colleagues (2006) found in their review of 63 studies, all including a parent training component, that program effects were moderated by SES. In another recent program evaluation of Comet, the results were moderated by the age of the mother (Kling, et al. 2010). Further, in the abovementioned meta-analysis of Incredible Years (Menting, et al., 2013), initial child problem levels influenced the effects of the program; children with higher problem levels showed greater improvement than those with lower initial problem levels. These results spilled over to the level of
intervention, where the treatment (clinical) trials had significantly larger effect sizes ($d=.50$) compared to the selective ($d=.13$) and the indicated ($d=.20$) preventive trials (Menting, et al., 2015). However, in the recent short-term comparison of Comet, Connect, Cope, and Incredible Years, none of the investigated variables (e.g., SES, parent and child age, immigrant status) influenced program effectiveness (Stattin, et al. 2015), which is in line with the findings from a recent review of 13 effectiveness evaluations of behavioral parenting programs (Furlong, et al., 2012). Hence, initial problem levels and whether a program is delivered as a treatment or prevention effort can influence program effects. In sum, there are some indicators that group-based parenting programs might be more effective for some parents and children than others, but the results are contradictory. More in-depth analysis is needed before a full understanding is reached concerning whether some parents and children benefit more from the programs.

It is not only the effects of the programs per se that might differ. In regular care, the programs can be delivered in a variety of settings (e.g., clinical, community, school, religious organizations). Research on if and how the setting or place of care influences program outcomes is, at present, negligible. Therefore, it is unclear whether the programs have similar impact in different settings. There are differences between organizations that offer parenting programs. For instance, the different sectors within the welfare system, child and adolescent psychiatric care, social care, and school health, differ in their level within the welfare system. Parenting programs might also be offered by church congregations or other organizations outside the official welfare system, and these contexts probably differ in comparison to each other as well as in comparison to the welfare organizations. In turn, these differences between settings might affect program outcomes. However, there is a lack of research in this particular area. Investigations of whether the characteristics of the setting influence program effectiveness ought to provide additional and valuable information which can be used when implementing intervention distribution and considering possibilities of co-operation between sectors (Kazak, et al., 2010). Hence, it is unclear whether the setting matters; that is, whether intervention effectiveness depends on where or by whom the programs are delivered.

Another limitation of previous research concerns the types of programs that are most effective for parents and children. To date, programs with different orientation (e.g., behavioral and cognitive-behavioral) have only been compared in a few meta-analyses (e.g., Furlong et al. 2012; Lundahl, et al. 2006). However, in line with the recent argumentation above, the
comparison has often been between “apples and oranges”, which makes it difficult to draw conclusions about the effectiveness of group-based parenting programs (Sandler et al. 2011). A recent Cochrane review (Furlong, et al. 2012) of 13 trials of behavioral and cognitive-behavioral group-based parenting programs revealed that the programs were effective in the short term, but failed to provide support of long-term effectiveness. In this report, however, the programs were not compared with each other. Hence, it remains unclear whether different types of programs might influence program effectiveness in both the short and the long term.

In line with the previously-mentioned limitation is that, overall, previous research is dominated by behaviorally-oriented interventions (e.g., Lundahl, et al. 2006). Hence, there is a lack of program evaluations of non-behavioral programs. This seriously hampers comparability across programs based on theoretical underpinning. Whether and how the theoretical underpinning of a program influences program outcomes is yet another research area with an extensive knowledge gap. Parenting programs are commonly categorized as either behavioral or non-behavioral programs, depending on their theoretical orientation. In line with program impact theory (Rossi, et al. 2004), the theoretical orientation ought to exert a vast influence on how the program content is structured and delivered, and on how the outcomes are manifested. Still, there is very limited understanding of how the theoretical mechanisms and components of the behavioral and non-behavioral programs affect the parents and their children.

Yet another limitation is whether the programs are effective when the “real life” delivery is not monitored by the program developers. It is desirable to evaluate program effectiveness independently; that is, without the involvement of the program developer/s (Flay, et al. 2005; Gottfredson, et al. 2015). Independent effectiveness trials are important means to investigate whether an intervention works as expected in regular care when the program developer/s are not involved. Eisner (2009) argued that the results might be over-estimated when program developers are involved, although the opposite has also been found (Sherman & Strang, 2009). For instance, a recent meta-analytic review including 50 efficacy and effectiveness studies of the Incredible Years program found no effects due to the involvement of the program developer (Menting, et al. 2013). However, one recent independent effectiveness evaluation of Triple P (Malti, Ribeaud, & Eisner, 2011) failed to replicate the findings reported by the program developers.
Hence, independent research is scarce, and it is not possible to draw conclusions from existing research about whether dependent or independent trials differ in effectiveness because the findings are inconclusive.

Independent investigation of long-term effectiveness has, to the best of my knowledge, only been reported for the Incredible Years program (i.e., Axberg & Broberg, 2012; Scott, 2005). Those findings indicated equal and maintained outcomes of the program over time, but have not been contrasted with long-term effectiveness evaluations conducted by the program developer. In sum, even though the findings are promising, there is still not enough evidence from independent effectiveness trials to conclude with certainty that the programs deliver the anticipated outcomes when they are implemented and delivered in regular care.

In order to appropriately compare similar programs, such as group-based parenting programs, the programs should ultimately be evaluated and compared together in one and the same study. At present, to my knowledge, only two studies have reported effectiveness findings of several group-based programs implemented and delivered in regular care within the same study (Lindsay & Strand, 2013; Stattin, et al. 2015). The first study evaluated the behaviorally-oriented programs Incredible Years, Strengthening Families 10–14, Strengthening Families Strengthening Children, and Triple P for parents of children aged 8–13. The immediate effects on children’s externalizing problems, parental behaviors, and parental well-being were generally moderate to strong. The programs did not differ substantially, and the effects remained at the one-year follow-up (Lindsay & Strand, 2013). However, the evaluation was post hoc in character, and parents were not randomized to the programs or to any control condition.

The second study is the recent independent randomized waitlist control trial evaluating the effectiveness of four group-based parenting programs (Comet, Connect, Cope, and Incredible Years) within the same study design (Stattin, et al. 2015). The programs differ in their theoretical base; Comet and Incredible Years are based on behavioral theories, Cope is a blended program (behavioral and problem solving), and Connect is based on attachment theory. The programs were implemented and delivered at both clinical and community care, with 908 parents of children aged 3–12 years randomized to one of the four programs, a book condition, or a waitlist control. The short-term results show that the four programs were effective in decreasing children’s externalizing problems (small to moderate effect size) and improving parental behaviors and well-being (small effect size).
were some small but significant differences between the programs. The behaviorally-oriented programs, Comet and Incredible Years, were somewhat more effective concerning children’s externalizing problems compared to the non-behaviorally-oriented programs, Cope and Connect (Stattin, et al. 2015). Even though this was a theoretically-driven and planned comparison study, it only reported the immediate short-term effects. Given the difference in theoretical underpinnings, it is plausible that program effectiveness develops differently. The findings from this study clearly identify the need to monitor and follow the development of program effects over time in order to establish whether one program is more effective in comparison to another.

To sum up the limitations of previous investigations into the effectiveness of group-based parenting programs, there are currently no empirical studies in the literature that have independently analyzed differences between parenting programs (with different theoretical foundations) in one single study over time. It also has not been fully investigated whether parenting programs work better for some participants than for others over time, or whether the settings in which the programs are delivered can influence program outcomes. In conclusion, there is a lack of evidence concerning several important aspects in current effectiveness research on group-based parenting programs. This, in turn, hampers the overall ability to draw conclusions about the generalizability of efficacy outcomes into a real world delivery of the programs. Equally importantly, it is plausible that the lack of evidence negatively influences the dissemination and implementation of the programs into regular care.

Implementation research

Implementation research on prevention is a young and rapidly growing field. Overall, inconsistencies concerning measures, constructs, and how to measure outcomes has been recognized and recently addressed (Proctor, et al. 2011). Although there are several models of implementation (e.g., Aarons, Hurlburt, & Horowitz, 2011; Durlak & DuPree, 2008; Fixsen et al. 2005; Wandersman, 2003), they have generally not been properly used as theoretical bases in previous preventive implementation research. As a consequence, little is known about the implementation of preventive efforts. Unsurprisingly, the knowledge gap concerning the implementation of group-based parenting programs is substantial, and there are several limitations within the current literature. For example, intervention outcomes have rarely been evaluated in relation to implementation outcomes, other than
aspects concerning program fidelity (e.g., Cantu, Hill & Becker, 2010; Stat- 
tin, et al. 2015) or adaptation (e.g., Kumpfer, Xie, & O’Driscoll, 2012; Ma-
rek, Brock, & Sullivan, 2006). In addition, existing research has commonly 
been post hoc evaluations of implementation, when a program has already 
been implemented, and not theoretically driven or planned implementation 
evaluations (e.g., Ogden & Sørlie, 2009; Sanders, Prinz, & Shapiro, 2009). 
Hence, much remains unclear about the implementation of group-based 
parenting programs in general, and whether there might be intervention-
specific challenges regarding the implementation of, for instance, group-
based parenting programs in comparison to individual parenting programs. 
The current preventive research approach to implementation has not been 
able to provide an answer. Hence it seems clear that there is a substantial 
lack of preventive implementation research on group-based parenting pro-
grams. There is a need to get a comprehensive overview of the current im-
plementation research on group-based parenting programs before the exist-
ing gaps of knowledge can be properly identified and addressed.

This dissertation

The main aim of this dissertation is to expand the knowledge concerning 
group-based parenting programs targeting child mental health problems. 
Taking my stance from the conclusions in the Swedish SBU report (2010; 
see also Smedler, et al, 2015), I have evaluated several aspects relating to 
the effectiveness and influencing the implementation of group-based parent-
ing programs.

In studies I, II, and III we used data from a large national independent 
effectiveness evaluation of four group-based parenting programs (Comet, 
Connect, Cope, and Incredible Years) to investigate the effects on parents 
(study I) and children (studies I and II) over time. The sample from this 
national evaluation was also used in study III, where we investigated 
whether program effectiveness differs across settings (school, social care, or 
child and adolescent psychiatric care).

More specifically, study I was a one-year follow-up of the immediate ef-
ficacy of the four parenting programs (Stattin, et al. 2015). We inves-
tigated whether the improvements in children’s problem behaviors and par-
ents’ parenting and well-being were maintained or had changed one year 
after the program ended, and whether the programs worked better for some 
parents and children compared to others (e.g., those differing in age, gender,
and parental education). Hence, study I evaluated and compared the effectiveness of the four programs concerning both parental and child outcomes one year after the end of the programs.

In study II, we concentrated our investigation on the children and the two-year program effects. We took a special interest in studying the program effects on children with ADHD symptoms above clinical cut-offs. Thus, in comparison to study I, study II evaluated and compared the effectiveness of the four programs on the children’s problem behaviors in more depth, with a specific focus on the effects on children’s ADHD symptoms, two years after the end of the programs.

The main focus in study III was to investigate whether different sectors of care (schools, social care, and child and adolescent psychiatric care) differed in overall program effectiveness when they delivered the four programs included in the national effectiveness evaluation.

In study IV, we used a systematic search and a narrative approach to review current research concerning the implementation of group-based parenting programs.

In this dissertation, I address the following research questions concerning group-based parenting programs:

1. Are the programs effective over time (studies I and II)?
2. Do the programs work better for some parents and children than others over time (study I)?
3. Does program effectiveness differ across settings (study III)?
4. What is the current state of preventive implementation research field (study IV)?
5. Does the implementation of group-based parenting programs face format-specific challenges (study IV)?

Description of the included studies

Methods

Studies I, II, and III

Participants

A total of 1113 parents of children aged 3–12 attended the initial parenting meetings. Overall, the recruitment process had worked well, and only three children did not meet the inclusion criteria (i.e., age range 3-12, no
autism spectrum disorder). Two of the children did not meet the age criteria; one was younger than three years and one was older than thirteen years. The third child was not eligible for participation due to a diagnosed autism spectrum disorder. Although these parents were excluded from the evaluation, they were still offered the opportunity to participate in the parenting programs. In addition, six other children were excluded by the personnel at separate care units, because the personnel concluded that these children would not benefit from the kind of intervention offered by the parenting programs (manualized and group-based). As a consequence, 1104 parents were randomized into the six different alternatives: Comet \( n = 207 \), Connect \( n = 218 \), Cope \( n = 202 \), Incredible Years \( n = 122 \), book \( n = 196 \), and waitlist \( n = 159 \). However, in the long-term studies included in this dissertation, no comparison were made with either the book condition or the waitlist condition, because the parents in both these groups were invited to participate in the programs after post-test (see Figure 4; only primary reporters included). Hence, my initial study sample included 749 parents and children.

The participants in studies I–III were not compared to the control condition because it was a strict waitlist control condition, and these parents received the intervention after program end. Hence, the participants in studies I, II, and III are the parents who attended the four different parenting programs. In the majority of cases, one parent per child, usually the mother, attended the session. When both parents participated in a program, the parent with the highest attendance was selected as the primary reporter. The mother was chosen as the primary reporter if attendance was equal. In most cases (approximately 85%), the primary reporter was the mother.

**Participants in study I**

The analytic sample of this study consisted of the 635 primary reporters who started one of the parenting programs (of the 749 who were initially randomized to the parenting programs). At post-test, 598 (94.2%) responded to the questionnaire, and at the one-year follow-up, 556 (87.6%) responded. Complete pre-test, post-test, and one-year data were available for 556 (87.6%) of the participants.

On average, the primary reporters were 38.3 years old (range 20-60 years). The majority of the parents were married or cohabiting (73.2%), and a majority of the parents were born in Sweden or in another European country (86%). Most households reported their monthly income in a range of 30 000–40 000 SEK (pre-tax, roughly 3000–4000 €), and the majority
1113 families assessed for eligibility

1104 randomized to a parent training program

Of these, 749 randomized to a parent training program

196 randomized to a self-help book condition
139 randomized to a waitlist condition
(not reported here)

3 did not meet inclusion criteria
6 excluded for other reasons

110 provided data
187 provided data
184 provided data
182 provided data
189 provided data
171 provided data
177 provided data
174 provided data
152 provided data
149 provided data
155 provided data
147 provided data
100 provided data
144 provided data
149 provided data
155 provided data
159 provided data
148 provided data
174 provided data
178 provided data
172 provided data
170 provided data
168 provided data
166 provided data
164 provided data
162 provided data
160 provided data
158 provided data
156 provided data
154 provided data
152 provided data
150 provided data
148 provided data
146 provided data
144 provided data
142 provided data
140 provided data
138 provided data
136 provided data
134 provided data
132 provided data
130 provided data
128 provided data
126 provided data
124 provided data
122 provided data
120 provided data
118 provided data
116 provided data
114 provided data
112 provided data
110 provided data
108 provided data
106 provided data
104 provided data
102 provided data
100 provided data
98 provided data
96 provided data
94 provided data
92 provided data
90 provided data
88 provided data
86 provided data
84 provided data
82 provided data
80 provided data
78 provided data
76 provided data
74 provided data
72 provided data
70 provided data
68 provided data
66 provided data
64 provided data
62 provided data
60 provided data
58 provided data
56 provided data
54 provided data
52 provided data
50 provided data
48 provided data
46 provided data
44 provided data
42 provided data
40 provided data
38 provided data
36 provided data
34 provided data
32 provided data
30 provided data
28 provided data
26 provided data
24 provided data
22 provided data
20 provided data
18 provided data
16 provided data
14 provided data
12 provided data
10 provided data
8 provided data
6 provided data
4 provided data
2 provided data

Baseline
Post-test
1-year follow-up
2-year follow-up

155 declined
147 declined
100 declined
69 declined
55 declined
149 declined
152 declined
155 declined
159 declined
163 declined
167 declined
171 declined
174 declined
177 declined
184 declined
189 declined
196 declined
202 declined
207 declined
218 declined
184 declined
189 declined
196 declined
139 declined
147 declined
149 declined
102 declined
100 declined
100 declined
102 declined
104 declined
106 declined
108 declined
110 declined
112 declined
114 declined
116 declined
118 declined
120 declined
122 declined
124 declined
126 declined
128 declined
130 declined
132 declined
134 declined
136 declined
138 declined
140 declined
142 declined
144 declined
146 declined
148 declined
150 declined
152 declined
154 declined
156 declined
158 declined
160 declined
162 declined
164 declined
166 declined
168 declined
170 declined
172 declined
174 declined
176 declined
178 declined
180 declined
182 declined
184 declined
186 declined
188 declined
190 declined
192 declined
194 declined
196 declined
198 declined
200 declined
202 declined
204 declined
206 declined
208 declined
210 declined
212 declined
214 declined
216 declined
218 declined
220 declined
222 declined
224 declined
226 declined
228 declined
230 declined
232 declined
234 declined
236 declined
238 declined
240 declined
242 declined
244 declined
246 declined
248 declined
250 declined
252 declined
254 declined
256 declined
258 declined
260 declined
262 declined
264 declined
266 declined
268 declined
270 declined
272 declined
274 declined
276 declined
278 declined
280 declined
282 declined
284 declined
286 declined
288 declined
290 declined
292 declined
294 declined
296 declined
298 declined
300 declined
302 declined
304 declined
306 declined
308 declined
310 declined
312 declined
314 declined
316 declined
318 declined
320 declined
322 declined
324 declined
326 declined
328 declined
330 declined
332 declined
334 declined
336 declined
338 declined
340 declined
342 declined
344 declined
346 declined
348 declined
350 declined
352 declined
354 declined
356 declined
358 declined
360 declined
362 declined
364 declined
366 declined
368 declined
370 declined
372 declined
374 declined
376 declined
378 declined
380 declined
382 declined
384 declined
386 declined
388 declined
390 declined
392 declined
394 declined
396 declined
398 declined
400 declined
402 declined
404 declined
406 declined
408 declined
410 declined
412 declined
414 declined
416 declined
418 declined
420 declined
422 declined
424 declined
426 declined
428 declined
430 declined
432 declined
434 declined
436 declined
438 declined
440 declined
442 declined
444 declined
446 declined
448 declined
450 declined
452 declined
454 declined
456 declined
458 declined
460 declined
462 declined
464 declined
466 declined
468 declined
470 declined
472 declined
474 declined
476 declined
478 declined
480 declined
482 declined
484 declined
486 declined
488 declined
490 declined
492 declined
494 declined
496 declined
498 declined
500 declined
502 declined
504 declined
506 declined
508 declined
510 declined
512 declined
514 declined
516 declined
518 declined
520 declined
522 declined
524 declined
526 declined
528 declined
530 declined
532 declined
534 declined
536 declined
538 declined
540 declined
542 declined
544 declined
546 declined
548 declined
550 declined
552 declined
554 declined
556 declined
558 declined
560 declined
562 declined
564 declined
566 declined
568 declined
570 declined
572 declined
574 declined
576 declined
578 declined
580 declined
582 declined
584 declined
586 declined
588 declined
590 declined
592 declined
594 declined
596 declined
598 declined
600 declined
602 declined
604 declined
606 declined
608 declined
610 declined
612 declined
614 declined
616 declined
618 declined
620 declined
622 declined
624 declined
626 declined
628 declined
630 declined
632 declined

Figure 4. Flow diagram
perceived their income as sufficient (59.7%). Almost half of the primary reporters (45.0%) had a university degree. About two-thirds of the children were boys (62.8%), and the children’s mean age was 7.9 years (range 3–12).

Participants in study II

A strict intention-to-treat approach was used for this study, and so the analytic sample comprised all primary reporters who were randomized to one of the parenting programs (n=749). Of these, all (100%) answered the pre-test questionnaires, 661 (88.3%) responded at post-test, and 543 (72.5%) responded to the two-year follow-up questionnaire. Complete pre-test, post-test, and two-year follow-up data were available for 531 (70.9%) participants in the intention-to-treat sample. Also included in this study was the parent daily report (PDR) interview; the response rates for these interviews were 702 (93.7%) at pre-test, 680 (90.8%) at post-test, and 579 (77.3%) at the two-year follow-up. Complete PDR data for all three measurement points used in study II were available for 562 (75.0%) participants. Parents of 114 (15.2%) children never started the parenting program they were randomized to. Hence, of the 749 participants in the intention-to-treat sample, 635 parents (84.8%) started their program and 114 never attended.

The sample used in study II resembled the sample used in studies I and III to a large extent. The mean age of the primary reporters was 38.23 years (range 20-60), the majority of the parents were married or cohabiting (72.4%), and a minority of the parents (14.7%) had immigrated to Sweden from a country outside Europe. Most households reported their income in a range of 30 000–40 000 SEK (pre-tax, roughly 3000–4000 €). Approximately half of the parents (46.5%) had a university degree. On average the children were 7.82 years old (range 3-12), and almost two-thirds of them were boys (62.4%).

Participants in study III

Parents. The participants in this sample comprised the 553 parents who were randomized to and subsequently started a parenting program at child and adolescent psychiatric care (3 units), social care (18 units), or schools (5 units). Of these 553 parents, 97 (17.5%) were recruited by child and adolescent psychiatric care, 376 (68%) by social care, and 80 (14.5%) by schools.

Again, the participants’ demographic information resembled the samples from studies I and II. The majority of the participants were female (85.2%)
with an average age of 38.56 years (range 20-60). Almost three quarters were married or cohabitating (72.8%), and the rest were single parents. Commonly both parents were born in Scandinavian countries (79.0%). A majority indicated that their monthly income was sufficient for their expenses (68.1%). Finally, almost half of the parents (46.6%) had university education, but 8.2% of the parents only had comprehensive school educations. Two-thirds of the children were boys. The average age of the children was 8.01 years (SD = 2.44; range: 3–12).

**Group leaders.** In order to assess group leader differences in each sector, the group leaders at the 26 care units who volunteered to participate in the evaluation also participated in study III. There were 105 of them in total, of whom 20 were employed in child and adolescent psychiatric care, 63 in social care, and 22 in schools.

**Procedure**

The data included in studies I, II, and III were drawn from parents who participated in an independent national effectiveness evaluation of parenting programs initiated and funded by the Swedish National Board of Health. Four established parenting programs were included in the project: Comet (Kling, et al., 2010), Cope (Cunningham, et al. 1995), Connect (Obssuth, et al. 2006), and Incredible Years (Webster-Stratton, 1984; Webster-Stratton, Reid, Hammond, 2004). The project was launched in the fall of 2009, and the final two-year follow-up data were collected in the winter of 2012.

Four research groups collaborated within the project: Gothenburg University; Karolinska Institute, Stockholm; the University of Lund; and Örebro University. Researchers identified and approached care units in their proximity who could offer or had implemented at least two parenting programs (inclusion criterion). All in all, 30 care units within different sectors of care (mainly schools, social care units, and child and adolescent psychiatric clinics) who fulfilled the inclusion criteria (i.e., able to deliver two group-based parenting programs, and perform randomization) were recruited and participated in the project.

The research personnel held information meetings for the staff at each care unit, and all group leaders signed consent forms and answered a short questionnaire. In the case where the units did not already offer two parenting programs, group leaders were offered training in accordance with each chosen program’s normal group leader training. The training was funded by the project, and some new group leaders were trained for all programs.
The different parenting groups were run by group leaders from the care unit’s ordinary staff members. A majority of the parents were already enrolled at the care units or contacted the units on their own. Some complementary advertisement and information to parents were given in order to fill up all the parenting groups, for instance at regular parent meetings, but overall the recruitment followed the usual protocol.

When the care units had recruited enough parents (n=20–30), the parents were invited to an information meeting at the care unit, held approximately two weeks prior to program start. This meeting was the only change from the normal recruitment process at the care units. At these meetings, research personnel gave information about the upcoming study and asked if parents were interested in participating. The parents were given thorough information about the project, including the information that their participation was voluntary and that they could leave the project at any time without losing their incentive of 500 kronor (approximately €50, or $70). Overall, 99% of the parents chose to participate and signed the consent form. The parents who chose to participate were eligible for randomization.

**Design**

A randomized three-group design was used where the 1113 parents were randomized to either one of two programs or a control condition. The control groups eligible for randomization varied between the first and second year of data collection; parents randomized to the control condition during the first year were offered a self-help book (Forster, 2009), whereas parents in the second year were randomized to a wait-list control group. Thus, in the first year of the project parents were randomized to: 1) program A given at the unit, 2) program B given at the unit, and 3) the book condition. Initial preliminary results from the first year’s data collection suggested that the parents in the book condition also showed improvements; in other words, the book condition did not operate as a control condition, but had to be seen as an additional intervention (self-instructed). As a consequence, after approval from the Ethical Board, the research group decided to introduce a waitlist control condition. Thus, in the second year the parents were randomized to: 1) program A given at the unit, 2) program B given at the unit, and 3) the waitlist control condition. Overall then, the project covered four parenting programs, one self-help book condition, and a waitlist control condition (see Figure 4). Hence, at each unit, the parents were randomized between three alternatives, but the overall project contained six alternatives.
While the programs were running, the parents in both the book and the waitlist condition were allowed, if needed, to receive the ordinary treatment given at the care units (i.e., treatment as usual; TAU). The treatment varied between the care units as well as between the sectors. Hence, no pre-agreed standard treatment was offered at the units. Approximately a quarter (25%) of the parents received additional help and support from their community or care unit at pre-test, and approximately one fifth (20%) had extra help at post-test, irrespective of being in a program or in the control conditions.

After post-test, the parents in the control conditions were invited to attend the programs. The parenting programs target children within different age ranges; Incredible Years targets children aged 3–8, Connect targets older children aged 9–12, and Comet and Cope target children within the whole age range of the project (3–12). As a consequence, age was taken into account at randomization. The parents of the younger children were randomized between the Incredible Years, Comet, and Cope programs, whereas the older children were randomized between the Comet, Cope, and Connect programs. Thus, there was no direct randomization between Incredible Years and Connect.

Most parents answered the pre-test questionnaire (T1) at the information meeting. At the last program session, approximately three to four months after pre-test, research personnel collected the post-test questionnaires (T2). If a parent was not present at the last session (a minority), the research personnel sent out the questionnaire with information and a pre-stamped return envelope. This postal procedure was followed for all questionnaires at the one-year (T3) and two-year (T4) follow-ups, which took place approximately 12 and 24 months after program end.

In addition to the questionnaires, research personnel who were blinded to randomization conducted structured telephone interviews with the parents. An adapted version of the Parent Daily Report (PDR) was used (for in-depth description, see the measure section below). The first interview (PDR1) was conducted approximately one week before program start. The second (PDR2) was conducted five weeks into the program and the third (PDR3) one week after program ending. PDR 4 and PDR5 were conducted approximately 12 and 24 months after the last session of the program. At the PDR4 and PDR5 interviews, parents were also reminded about the questionnaire that had been sent a week before, and asked to answer and return it at their convenience.
Analyses

We used latent change models in study I because we analyzed change in child and parental outcomes between two time points (pre-test and one year, and post-test and one year). In study II, we analyzed changes in child outcomes over three time points (pre-test, post-test, and two years) using latent growth models (LGM). To evaluate sector performance in study III, we used mixed linear models for repeated measurements with the programs as covariates.

Attrition analyses

The attrition analysis in the recently published short-term study including the waitlist control revealed that the parents who only answered the pre-test and those with complete pre- and post-test data differed on parental depression, sense of competence, and immigrant status. The received Nagelkerke $R^2$ was low (.07), and therefore the conclusion was that the differences would not substantially influence the immediate results (Stattin, et al. 2015).

In the studies included in this dissertation, those who answered at pre-test ($n=635$ in study I, $n=749$ in study II, and $n=554$ in study III) but did not have complete data at post-test, the one-year follow-up, or the two-year follow-up differed significantly on some variables. In study I, the 141 parents lost to attrition at the one-year follow-up had reported higher levels of hyperactivity/impulsivity, reported less parental emotional outbursts, and had a lower educational level than the parents who had complete data at the one-year follow-up. In study II, only parental education remained significant at the two-year follow-up, but the attrition analysis revealed that being born outside Europe (immigrant status in study II) predicted non-response at the two-year follow-up (parents lost to attrition $n=187$). In study III (attrition $n=62$), we found that responding and non-responding parents still differed on parental education but not on immigrant status, and the non-responding parents reported higher levels of defiance for their children and perceived more economic strain than the parents with complete pre-test, post-test, and one-year data. Hence, overall the attrition analysis revealed some differences between the parents with complete data in comparison to those who only answered at pre-test. Despite these differences, we concluded that the influence of these differences on our results should be minimal because the Nagelkerke $R^2$ was generally low (range =.04-.068) in studies I, II, and III.
Study IV

Procedure

Study IV comprised a narrative review of the implementation of group-based parenting programs. In this study, we systematically searched several electronic databases (i.e., Cochrane, ERIC, Medline, PubMed, and PsycInfo) for abstracts containing the search words: implementation*, dissemination*, parent*, child, and evaluation or prevention, between the 17th and 24th of November, 2014. There was no time limit put on the search, but language was restrained to English. In addition to this broad search we also searched specifically for implementation and dissemination for the programs that, in the initial search, appeared the most frequently evaluated (i.e., Triple P, Sanders, 1999; the Parent Management Training Oregon, Patterson, 2005; Incredible Years, Webster-Stratton, 1984; and the Strengthening Families Program, Kumpfer, DeMarsh, & Child, 1988; Mølgaard & Spoth, 2001). Later, the reference lists of retrieved and selected full-text articles were searched for relevant references on the subject of implementation and dissemination.

We conducted a step-wise selection of the articles included in the review. In steps one and two, we scanned the titles and abstracts of the retrieved articles (see Figure 5). The third step involved full-text examination of the remaining articles (and reference lists for references). The inclusion criteria for study IV were: 1) evaluation or investigation of any aspect concerning the implementation of group-based parenting programs, quantitatively (in numbers/scales) or qualitatively (interviews/case studies), with qualitative studies included because they could provide in-depth insight into the practitioners’ perspectives and experiences of implementation; 2) the evaluated implementations were parenting programs aimed at parents of children aged 2–12; and 3) the programs targeted children’s mental health problems broadly, both externalizing and internalizing mental health problems.

Articles were excluded if they evaluated parenting programs regarding purely physical and academic target areas (e.g., child obesity, diabetes, HIV, and literacy skills), were not group-based, and/or if they were aimed at parents of infants, toddlers, or adolescents.
Figure 5. Flow chart of the review process

Literature search in the Cochrane, ERIC, Medline, PubMed, and PsycInfo databases.
Search words: implementation*, dissemination*, parent*, child, evaluation and prevention. Published articles in English only.

n=1356
Duplicates n=112
Left for review n=1258

Program-specific search in all databases
IY=51, PMTO=108, SFP=99, Triple P=100
Duplicates n=48
Left for review n=310

Result of total search
n=1258+310=1568

Not on topic=1487

Screening of titles and abstracts of articles, selected for full text reading
n=81

54 (not relevant=21; not evaluating=33)

Selected for the final rating
n=35

Added from references n=3
(8, not relevant=5)

Insufficient study quality=11

Included in the review
n= 24
After the initial three steps of screening, 35 articles remained. In the fourth step, we graded these articles on study quality using the GRADE rating system (Atkins, et al. 2004) adjusted for our narrative approach. The unadjusted version of the GRADE system rates study quality on design, external and internal validity, data completeness, attrition, analysis, results, precision, and dependence on/independence of the program developer. In addition to these quality criteria, we added aspects concerning implementation outcomes and whether aspects of implementation had been evaluated in relation to intervention outcomes. The answer options for the quality criteria were “yes”, “no”, “not possible to answer”, or “not applicable”. When rating the overall strength of each study (its evidence base), we used a 4-point scale (1 = insufficient, 2 = limited, 3 = moderately strong, or 4 = strong).

Before the final step in the screening process, we rated three articles in collaboration from approximately 30% of the included articles (in total 10 studies, 3 early and 7 recent), using stratified randomization to ensure that all studies were proportionally included (two groups; publication year 1992–2009: 12 articles, and 2010-2014: 23 articles). These articles were then rated independently by both authors, to ensure inter-rater agreement. Next, we compared the individual ratings. If we differed in our ratings, we reached consensus through thorough discussion. We set a satisfactory inter-rater agreement to 85% or above. When this was established, I, as the first author, rated the remaining articles. We decided to include articles with overall study quality ratings between 2 and 4 (with 4 being the maximum), which led to the eventual inclusion of 24 articles in the review (study IV).

**Measures**

**Studies I, II, and III**

In the first three studies of this dissertation, we used parental reports on child behaviors (studies I, II, and III), parental behaviors, and parental well-being (study I).

**Child measures**

Concerning the measurements of child behaviors, parents were asked to recall the past 6 months at pre-test, and the past month in the subsequent questionnaires (post-test, and the one-year and two-year follow-ups). For
the structured telephone interview (PDR), parents were asked to think of the past 24 hours (study II).

**Externalizing behaviors (studies I, II, and III)**

**Behavior problems.** We used the Eyberg Child Behavior Inventory (ECBI; Eyberg & Ross, 1978) as our primary outcome measure to evaluate parents’ perceptions of their child’s externalizing behaviors. The ECBI is a 36-item scale which assesses both the frequency of the child’s externalizing behaviors (*the intensity subscale*) and whether parents perceive the behaviors as problematic (*the problem subscale*). Examples of items include: “Does not obey house rules”, “Whines”, and “Constantly seeks attention”. In study II, we used the 95th percentile cut-offs from Swedish norm data to indicate clinical relevance (Axberg, Hanse, & Broberg, 2008). Parents rated the intensity items on a seven-point scale from 1 (*never*) to 7 (*always*), and the problem items on a two-point scale: 0 (*not a problem*) and 1 (*a problem*).

**ADHD symptoms and disruptive behaviors.** As our secondary outcome measure we used the DSM-IV version of the Swanson, Nolan, and Pelham Questionnaire (SNAP-IV) (Swanson, 1992). The SNAP-IV consists of two subscales comprising 9 items, each targeting the DSM-IV criteria for ADHD symptoms: inattention (e.g., “often does not seem to listen when spoken to directly,” and “often is forgetful in daily activities”), and hyperactivity/impulsivity (e.g., “often is ‘on the go’ or often acts like ‘driven by a motor’”, and “often blurts out answers before questions have been completed”). In addition, we used the oppositional defiance SNAP-IV symptom subscale, which consists of 8 items (e.g., “often loses temper”, and “often argues with adults”). Parents rated the items on a four-point scale, ranging from 0 (*not at all*) to 3 (*very much*). In study II, we used the 95th percentile cut-off scores indicating clinical relevance found at [http://www.myadhd.com/snap-iv-6160-18sampl.html](http://www.myadhd.com/snap-iv-6160-18sampl.html).

**Child behavior in the past twenty-four hours (study II).** The Parent Daily Report (PDR checklist) is a structured telephone interview where parents report on their child’s behaviors during the past 24 hours. It is a brief, non-intrusive, and cost-effective way to measure child behaviors. The checklist has been extensively used as an outcome measure of child conduct problems; it was used in 9 out of 22 studies included in a recent review (Eyberg, Nelson, & Boggs 2008; highlighted by Nadler, & Roberts, 2013). The checklist was originally developed as a supplement or alternative to home observations (Chamberlain & Reid, 1987; Patterson, Reid, Jones, & Conger, 1975), and contained 33 items. However, for the present effectiveness
evaluation, the checklist was translated and shortened from 33 to 24 items by two senior researchers in the research group. Nine items that were not related to externalizing problem behaviors were excluded (e.g., bedwetting and sadness). The interviews were administered through a structured telephone interview conducted by a trained test leader who was blinded to the participants’ randomization. During the telephone interview, parents are asked to report whether a behavior (e.g., teased, fought, been afraid) had not occurred (=0), occurred but was not stressful (=1), or occurred and was stressful (=2) in the past 24 hours. As our focus in study II was on child outcomes, the response alternatives 1 and 2 were merged into a single alternative, 1 (occurred). We did this because the stressful aspect of the behavior was related to whether the parent perceived the behavior as stressful or not, and not the occurrence of the child’s behavior. Hence, for the purpose of study II the responses were 0 (did not occur) and 1 (occurred). Further, also for study II, a principal component factor analysis was conducted for the PDR measure, and a 13-item core conduct problem factor (PDR-CP) emerged. These items were: "fought", "argued or spoke back", "was destructive", "was defiant", "hit, bit or scratched", "whined", "interrupted", "pouted", "used foul language", "teased", "was uneasy or had trouble sitting still", "had tantrums", and "was loud". In study II, the 13-item PDR-CP was used as a measure of child conduct problems.

Parent measures (study I)

Parents’ negative reactions to their children

Angry outbursts. Parents’ angry outbursts were measured by asking parents “What do you do when your child does something you REALLY DO NOT LIKE?” (Stattin, Persson, Burk, & Kerr, 2011). Four items were used for this, including “My first reaction is anger, and I yell at the child” and “I get angry and have an emotional outburst”, each rated on a three-point scale ranging from 1 (never) to 3 (most often).

Harsh treatment. To give a broad measure of harsh treatment, parts of the Parents’ Practice Interview (PPI; Webster-Stratton, 1998; Webster-Stratton, Reid, and Hammond, 2001) were used. Seven items were used, including “Take away privileges (like TV, playing with friends)” and “Slap or hit your child (but not spanking)”. Responses were rated on a seven-point scale ranging from 1 (never) to 7 (always).
Parents’ positive reactions to their children

**Attempted understanding.** To assess parents’ *attempted understanding*, we again asked “What do you do when your child does something you REALLY DO NOT LIKE?” (Stattin, et al., 2011), this time with five items including “I try to talk it through without creating new conflicts” and “I try to understand how the child thought and felt”, rated on a three-point scale ranging from 1 (*never*) to 3 (*most often*).

**Rewards.** We used a subscale of the Parents’ Practice Interview (PPI; Webster-Stratton, 1998; Webster-Stratton, et al., 2001) to measure parents’ use of rewards. The *Rewarding the Child* subscale we used had two items, “Buy something for him/her (such as special food, a small toy) or give him/her money for good behavior” and “Give him/her an extra privilege (such as cake, going to the movies, special activity for good behavior)”, which were rated on a seven-point scale ranging from 1 (*never*) to 7 (*always*).

Parents’ sense of competence and emotions

**Parents’ sense of competence.** We used the 16-item version of the Parenting Sense of Competence scale (PSOC; Johnston & Mash, 1989), which measures parental *satisfaction* (9 items; e.g., “It is difficult to know if you as a parent do a good or a bad job” [reversed]) and parental *efficacy* (7 items; e.g., “I believe I am a good role model as a parent.”). Parents answered on a six-point scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*).

**Stress.** To measure caregiver *stress* we used a version of the original Caregiver Strain Questionnaire (CGQS; Brannan, Helflinger, & Bickman, 1997). Specifically, we used the 10-item subscale of *objective strain* to assess parents’ levels of stress (e.g., “How often were you interrupted in what you were doing”, “How often did you have to stay at home from work or neglect other duties?”). Parents were asked to look back and report on how they had been affected by their children’s problems. The responses were given on a five-point scale ranging from 1 (*not at all*) to 5 (*very often*).

**Depressive symptoms.** We used the Center of Epidemiological Studies – Depression scale (CES-D; Radloff 1977) to measure parent’s depressive symptoms (20 items; e.g., “During the past week I was bothered by things that normally don’t bother me”, and “During the past week my sleep was restless”). Parents rated their responses on a four-point scale ranging from 0 (*not at all*) to 3 (*often*).
Results

Because studies I and II were follow-up evaluations of the short-term (immediate) effects of the four programs (Comet, Cope, Connect, and Incredible Years), I provide a short summary of the immediate effects (Stattin, et al. 2015). The results showed that all programs were effective immediately after program end in comparison to the waitlist control. Children’s externalizing behaviors decreased significantly, with effect sizes ranging from small to medium (ECBI, Cohen’s $d= .17-63$; SNAP IV [ADHD symptoms], $d= .01-26$). The attachment-based program, Connect, showed slightly smaller effect sizes on disruptive behaviors compared to the other, more behaviorally-oriented, programs (Comet, Cope, and Incredible Years), and did not show significant reductions of ADHD symptoms. In contrast to child outcomes, the immediate results concerning parents were less consistent across the programs. Parents participating in Comet, Connect, and Incredible Years increased their use of rewards, in contrast to parents in Cope. Also, only parents in Comet showed immediate significant decreases in negative parenting behaviors. Parents’ attempted understanding did not change significantly across any program, but parents in all programs increased their sense of competence and decreased their depressive symptoms. No moderating effects of socio-demographic variables influenced the immediate results. Overall, the parents reported smaller changes of their behavior, in general small effect sizes, compared to the change they reported concerning their children’s externalizing behaviors. Further, all programs showed significant effects on both children and parental behaviors and parental well-being compared to the control condition. Nonetheless, the immediate program effects differed, which implies a difference in effectiveness. However, it is unclear whether the immediate effects represent the final program effects, since it is possible that the program outcomes either deteriorate, improve, or sustain over time.

Study I

The aim of study I

As mentioned, study I was a one-year follow-up of the immediate child and parental outcomes of the four group-based parenting programs. The question that remained unanswered was whether the programs helped and supported parents and children over time. However, the aim of study I was four-fold, intended to address four questions. First, did the children and parents who participated in the parenting programs improve between pre-
test and the one-year follow-up? Second, were there any differences across the programs on the child and parent outcome measures? Third, were the immediate short-term improvements in child and parental behaviors and parental well-being maintained over a one-year period? And finally, did child age, parent age, gender, marital status, parents’ immigrant status and education, and family income and economic strain moderate the effectiveness of the programs over time?

Findings from study I

Concerning children’s and parents’ further improvement over time, our results varied. Parents participating in Cope and Connect, the initially slightly “weaker” programs, reported further significant but small improvement in their children’s externalizing behaviors (Cope, \(d = .09\); Connect, \(d = .21\)) and ADHD symptoms (Connect \(d = .11-.17\)). In comparison, parents who participated in Comet and Incredible Years reported either maintained or slightly deteriorated effects on their children’s externalizing behaviors (Comet, \(d = -.14- -.19\); Incredible Years, \(d = -.16- -.21\)) and ADHD symptoms (Comet, \(d = -.14- -.16\); Incredible Years, \(d = -.29\)). As a consequence, and answering our second question, the program effects on children’s externalizing behaviors did not differ significantly across programs one year after program end. Equally important, overall the child effects were maintained over time. Thus, the programs were equally effective between pre-test and one year after program end.

We found similar results concerning parents. At the one-year follow-up, parents either improved slightly further, deteriorated slightly, or maintained the immediate levels of parental behaviors and well-being. As with the child effects, Connect showed small but significant improvements (\(d = .16\)), Cope remained stable, whereas the effects of Comet and Incredible Years were either slightly deteriorated (Comet, \(d = -.16- -.40\); Incredible Years, \(d = -.25\)) or maintained between post-test and the one-year follow-up. Surprisingly, the parents in Comet significantly decreased in angry outbursts \((d = .36)\) between post-test and the one-year follow-up. However, also here the changes in effectiveness evened out the immediate differences across the programs. Hence, at the one-year follow-up, the effects on parental behaviors and well-being did not differ across programs. The effects were not influenced by socio-demographic variables. All in all, then, one year after the programs ended, the effects on parents and children did not differ across programs. Finally, the socio-demographic variables (child age, parent age, gender, marital status, parents’ immigrant status and education, and family income
and economic strain) did not influence the effectiveness of the programs over time. In sum, even though the program effects on parents and children manifested themselves in slightly different ways, our results indicate that the programs are equally effective over time, irrespective of theoretical orientation and program structure.

**Study II**

The aim of study II

In study II, we concentrated the evaluation exclusively on the children. The specific aims were to examine, first, whether the levels of child externalizing behavior and ADHD problems differed between the programs two years after program end, and, second, whether the trajectories of change in child behavior problems differed across the programs.

Findings from study II

Our results showed that the levels of child externalizing behavior and ADHD problems did not differ between the programs two years after program end. This was also true for children scoring over the 95th percentile on ADHD symptoms. All four programs (Comet, Cope, Connect, and Incredible Years) had reduced children’s externalizing behaviors with large effect sizes (Cohen’s $d = 0.75$ - $1.07$), and their ADHD symptoms with small to moderate effect sizes ($d = .22$ - .65) from pre-test to the two-year follow-up.

Our second aim was to examine if the change trajectories of the children’s externalizing behaviors differed across the programs. Our findings showed that the behaviorally-based programs, Comet and Incredible Years, and the blended program, Cope, reached their effects immediately after program end and maintained them over time. In contrast, the attachment-based program, Connect, had slightly smaller immediate effects on children’s externalizing behaviors, and non-significant reductions in ADHD symptoms immediately at the end of the program. Nonetheless, at the two-year follow-up there were no remaining differences between the programs, and our findings showed that the parents in Connect reported further overall improvements in their children’s behaviors ($d = .35$) throughout the two-year period following program end.

The children’s externalizing behaviors, measured by the structured PDR interview, showed immediate small-to-moderate effect sizes after program end, with no significant differences across the programs ($d = 0.49$ - 0.62). Interestingly though, the parents did not report any significant change in
these child behaviors between post-test (program end) and the two-year follow-up. Hence, they remained stable.

Overall then, the programs were equally effective concerning children’s externalizing behaviors two years after program end, irrespective of what measure we used. In terms of child behaviors measured by ECBI and ADHD symptoms, the program effects seemingly evolved through different paths. However, parents’ perceptions of their children’s behavior over the past 24 hours showed an equal and rapid immediate decrease across the programs, which was then maintained over the two years. Thus, the parents did not perceive that the 24-hour behavioral changes in their children evolved differently across programs.

Study III

The aim of study III

In study III, our aim was to investigate the different settings or sectors of care (child and adolescent psychiatric care, social care, and schools) which deliver parenting programs in Sweden. We had three aims in this study. First, we were interested in whether the parents and children differed on socio-demographic characteristics across the sectors. Second, we examined whether there were differences in children’s problem levels between the sectors. Third, we compared the short-term and one-year reductions in child problem behaviors across the three sectors of care.

Findings of study III

First, we evaluated whether group leader professions influenced the change in child outcomes or affected child outcomes across the sectors. Our results showed that the group leader professions did not influence the outcomes significantly, even though the professions differed across sectors. Next, we investigated whether aspects related to fidelity, attendance, and satisfaction differed between the sectors. We found that parents did not differ in their attendance or satisfaction across sectors. Program fidelity, on the other hand, differed slightly; the group leaders at schools delivered their programs with somewhat lower fidelity compared to the other sectors. However, the difference was small ($\eta^2 = .035$) and overall fidelity was satisfactory, and so we concluded that the difference should influence the results only marginally.

We were also interested in whether the parents and children differed across the sectors. We found some age differences, with the children at child
and adolescent psychiatry being the oldest and the children at social care
being the youngest. There were additional differences concerning socio-de-
mographic aspects, but the magnitudes of effects were small, indicating only
a marginal influence.

In addition, we found that the children’s levels of externalizing problems
(ECBI) and ADHD symptoms differed significantly between the three sec-
tors at pre-test. The children at child and adolescent psychiatric care gener-
ally had the highest problem levels, and the children at schools had the low-
est.

Next, we examined the reductions of externalizing behaviors across the
sectors. We did this in two steps, controlling for program effects. First, we
compared the immediate short-term reductions, and second, we compared
the pre-test through one-year reductions across sectors. The results for the
three sectors were parallel. Thus, the interaction between time and sector
was not significant in either step one or step two, indicating that the reduc-
tions in children’s externalizing behaviors did not differ significantly across
the three sectors of care, immediately or from pre-test to the one-year fol-
low-up. Hence, despite initial differences, when a program is delivered with
satisfactory fidelity, it seems as if the improvement in children’s externaliz-
ing behaviors does not differ across sectors.

Study IV

We used an established narrative approach (Baumeister & Leary, 1997)
to select the 24 studies included in study IV. This allowed us to also consider
inclusion of studies with lower quality ratings, for instance qualitative stud-
ies. The rationale for this approach is that the available studies mirror the
research field.

The aim of study IV

The main aim of our review was to 1) identify aspects that specifically
relate to the implementation of group-based parenting programs targeting
children’s mental health (i.e., externalizing and internalizing) within regular
care, 2) summarize current findings, and 3) identify areas where more re-
search is needed or missing.

Findings of study IV

First, we set out to investigate group-based parenting programs aimed at
children’s mental health problems using a broad conceptualization, thus in-
cluding both externalizing and internalizing mental health problems. Despite an extensive database search (the Cochrane, ERIC, Medline, PubMed and PsycInfo databases), we found no evidence of group-based parenting programs targeting internalizing problems. The studies we retrieved only concerned the implementation of programs targeting externalizing problems. Therefore we were only able to report on or draw conclusions about programs aimed at children’s externalizing problems, and that is what is reported in the following sections. Hence, it seems as if there is a lack of studies evaluating implementation of group-based parenting programs aimed at internalizing problems, if such programs even exist.

Second, we investigated whether group-based parenting programs encountered format-specific implementation challenges, in comparison to other parent training interventions (e.g., individual, family). However, in our search we found no implementation comparison between programs with different kinds of formats, and so we could not conclude that there were specific implementation aspects relating to group-based parenting programs. Thus, the next step was to investigate whether the present research on implementation of group-based parenting programs implied that specific aspects of the implementation of group-based programs were perceived as particularly challenging. A majority of our studies \((n=16)\) identified aspects pertaining to, for instance, group leader workload, provision of services outside of intervention delivery (meals, child care, and transportation), and minimum recruitment as challenging when a group-based program was implemented. Thus, even though we did not find evidence of format-specific challenges, we still identified that specific aspects related to the group format were perceived as challenging, both on the practitioner and on the agency level.

The second aim of our review was to summarize current implementation research specifically related to group-based parenting programs. For the summary, we used a recently suggested terminology (Proctor, et al, 2011) which includes eight implementation factors (constructs of implementation) found at all levels involved in the implementation of an intervention: acceptability, appropriateness, feasibility, adoption, fidelity, cost, penetration, and sustainability. Of these eight, we only found studies relating to five of them, namely fidelity (and adherence), acceptability, appropriateness, cost, and sustainability. Hence, it is evident that current research does not cover the whole implementation spectrum.

Concerning the aspects that had been investigated, the research base was too scattered and scarce for us to draw any firm conclusions. Instead we
focused on the implications we perceived in the findings from current research. Concerning *fidelity* (and adherence), our findings suggested that it influences parental satisfaction and attendance. Further, fidelity might be influenced by the behavior of agency management. Thus, our findings indicate a relationship between the different levels within the ecological framework (i.e., practitioner/parents, and practitioner/manager (agency)). Further, adherence was one aspect of fidelity that had been studied in its own right. Our review indicated that adherence to the content and structure of the interventions was not stable, but it remains unclear whether adherence was low to begin with or if the drift has evolved over time. Also, it is unclear as to what extent and how these deviations may have affected the outcomes of an intervention. The present findings indicate that practitioners’ overall understanding of the importance of fidelity and adherence in particular is low.

Next, our findings concerning *acceptability and appropriateness* related only to two individual studies on the cultural aspects of the Triple P parenting program and the Parent Management Training Oregon (PMT-O) program, implemented in Panama and Mexico, respectively. Both studies found that the programs were perceived, by parents (PMT-O) and practitioner (Triple P), as acceptable and appropriate in the respective countries. Neither of the studies compared an adapted version of the program to the original program. Hence, it is promising that the cultural aspects of the programs are identified as important to evaluate. Nonetheless, the fact that only two studies, to date, have investigated these aspects implies a need for future studies on this implementation aspect at all levels in the ecological system.

Only one study investigated the *cost* of a program. The evaluation was in relation to a concept we defined as agency *readiness*, which is an aspect related to (following) the implementation concepts of acceptability, appropriateness, feasibility, and adoption. It refers to whether an agency, practitioner, or client is ready to implement or participate in an intervention. The findings indicate that if an agency is not ready; that is, that it has not allocated and/or provided the resources needed for the implementation of a parenting program, the program is less cost effective, compared to an agency high on readiness. The importance of intervention cost effectiveness is a subject often highlighted in the prevention literature in general, and the literature on group-based parenting programs is no exception. On the contrary, the group format is supposed to expand the reach of the preventive effort, compared to, for instance, individual interventions. Therefore, we were surprised that there was only one study on the subject. Nevertheless,
we identified what we perceive to be an additional implementation construct, *readiness*. It seems that the stakeholders involved in implementation must be ready (in terms of having the resources and motivation) to go ahead with the program, in addition to deeming it as acceptable, appropriate, and feasible, and being willing to try it out (adoption).

* Sustainability in relation to the practitioners (group leaders) was the focus of four studies emanating from the implementation of the Triple P program. The four studies were based on the same sample of group leaders (*N* = 69 - 611), and investigated aspects relating to group leader training, competence, and current and predicted use of the intervention (sustainability). Overall, our findings indicated that time elapsed since group leader training influenced the later use of the program. Our findings also suggested that practitioner characteristics such as self-efficacy, competence, and years of experience influence intervention sustainability. Even though the findings from these four studies converge, there were methodological concerns (e.g., same base sample, level of randomization, and inadequate measures) that influenced our ability to draw firm conclusions. As a consequence, our findings can only help identify these aspects as group leader characteristics that seemingly influence sustainability.

Our final aim was to identify areas in need of further research using our conceptual ecological model and concept terminology (Fixsen, et al. 2005; Proctor, et al. 2011). We identified several gaps in the present literature. For instance, no evaluation had been done on implementation outcomes in relation to program effectiveness. Further, the implementation aspects of feasibility, adoption, and penetration had not been evaluated, and neither had the reciprocal relationships between the different levels (e.g., society, agency, participants). Hence, there is a need to cover all aspects of both the ecological implementation models and all aspects relating to implementation outcomes.

In addition, there were several methodological concerns that future research need to address. First, there is a need to ensure that future studies have sufficient scientific quality. For instance, that they are theoretically driven implementation evaluations using randomization and a multi-level approach. Second, according to our findings, implementation evaluations have only been conducted on a few (*n* = 7) of the existing group-based parenting programs. The program developer/s were involved in the majority of these evaluations, which might influence the outcomes. Finally, the majority of evaluations were conducted in North America. All in all, these are aspects that influence the overall generalizability of the findings. Equally important,
though, this also highlights the need of future research (e.g., well-planned evaluations of other programs, in other countries, independent of program developers, and in relation to program effectiveness). In short, there is a severe gap in implementation knowledge concerning group-based parenting programs that future prevention research needs to fill.

**Discussion**

**Adding new knowledge to prevention research**

The main aim of this dissertation is to expand current knowledge about group-based parenting programs aimed at improving children’s mental health problems. I have argued that there are substantial gaps in knowledge concerning aspects related to program effectiveness and the implementation of group-based parenting programs, and this dissertation aims to fill some of these knowledge gaps.

More specifically, studies I and II answered questions relating to the long-term effectiveness of four group-based parenting programs (Comet, Cope, Connect, and Incredible Years) on parents (study I) and children (study II). Study I also investigated whether long-term effectiveness was influenced by the socio-demographic characteristics of the participating parents and their children. Another perspective on effectiveness was added in study III. The three sectors of care that provide group-based parenting programs in Sweden (child and adolescent psychiatric care, social care, and schools) were compared to investigate whether they differ in effectiveness when they deliver the programs. Study IV investigated aspects relating to the implementation of group-based parenting programs aimed at children’s mental health problems. Specifically, study IV reviews and summarizes the current state of the implementation field regarding these specific kinds of programs as well as examining whether these programs face format-specific challenges. Taken together, the four studies included in this dissertation provide much-needed knowledge for both research and practice, and add unique information to the current knowledge base.

**Long-term effectiveness**

Even though we were unable to compare our results with a control condition over time, and thus unable to draw firm conclusions about program effectiveness over time, studies I and II still provide evidence that the parents participating in the four group-based parenting programs (Comet, Cope,
Connect, and Incredible Years) and their children, on average, maintained their immediate behavioral and emotional improvements over time. Equally important, the differences between the programs immediately after program end (Stattin, et al. 2015) were evened out over time, and so no significant differences remained between the programs at the one-year or two-year follow-up. This was found for both parents and their children after one year (study I), and for children after two years (study II). Studies I and II further showed that the program improvements seemingly developed through different trajectories. The behaviorally-oriented programs seemed to result in a more rapid decrease in children’s problem behaviors compared to the non-behaviorally-oriented programs. However, over time the behavioral programs indicated slightly worsened child behaviors, whereas the effects of the non-behavioral programs continued to improve; and so, after one and two years, there were no remaining differences between the programs in terms of the children’s externalizing problem levels, or parental behaviors and well-being. This was also true for the children with symptoms of ADHD at the beginning of the evaluation. Studies I and II were based on a quasi-randomized control trial in which the effectiveness of four similar, yet theoretically different, parenting programs were compared over time. The short-term results showed that the behavioral programs were more effective than the non-behavioral programs (Stattin, et al. 2015). However, the long-term findings in studies I and II clearly show that it is not enough to conduct short-term evaluations of preventive interventions. The immediate results cannot be interpreted as the final results. Instead, the effects on parents and their children continue to develop over time. Hence, studies I and II provide both new insight about the importance of long-term evaluations to assess program effectiveness properly, and new information about long-term effects.

Our findings are important because, to date, it is problematic to state with any certainty that a parenting program would show lasting effects in a real life setting. The existing evidence that programs tried in efficacy trials would yield similar results when they are delivered in real life settings is promising (Michelson, et al., 2013) but insufficient (NRC/IOM, 2009; SBU, 2010; Smedler, et al, 2015). In addition, the lack of long-term follow-ups, both in efficacy and effectiveness trials (NRC/IOM, 2009; SBU, 2010; Smedler, et al., 2015), makes it impossible for any stakeholder (e.g., policy makers, researchers, managers, practitioners, and parents) to conclude that a parenting program will work as a preventive intervention. That is, current research has not answered the important question of whether the parenting
programs prevent children’s mental health problems from occurring or escalating over time, which preventive interventions are supposed to do. To date, as concluded in the introduction of this dissertation, only a few parenting programs have been evaluated with scientific rigor in terms of effectiveness immediately after program end. Even fewer evaluations have been done on parenting interventions’ effectiveness over time, and no one has been able to compare program outcomes with a randomized control (Furlong, et al., 2012). In studies I and II, we were able to conclude that the effects of the programs were sustained over time. However, we cannot state with certainty that the programs were mainly responsible for these changes, because we lacked a control condition for comparison. Thus, much still remains unanswered concerning program effectiveness.

Studies I and II made long-term comparisons of the effectiveness of group-based programs with different theoretical orientations in one and the same design. This has not previously been done with parenting prevention programs. Our results indicated that program effects continued to develop over time, and that the effects of behavioral and non-behavioral programs might follow different trajectories. The short-term results (Stattin, et al. 2015) showed that the strictly behavioral programs (Comet and Incredible Years) resulted in somewhat greater improvements in the short term compared to the non-behavioral programs (Cope and Connect). Hence, if one only considered the short-term results, the behavioral programs would seem to be more effective. However, over time the development of program outcomes indicated some deteriorating effects for the behavioral programs, whereas the outcomes of the non-behavioral programs continue to improve. Thus, the long-term developments found in studies I and II clearly showed that evaluations of the short-term results are not enough, and so they do not provide sufficient definitive evidence of program effectiveness.

One plausible reason for the differences in outcomes of the behavioral and non-behavioral programs might reside in the differences in their primary focus to change either behavior (Comet and Incredible Years) or attitudes (Cope and Connect). The behavioral programs with their practical “toolbox” of parenting techniques concerning, for instance, spending quality time with the child, rewards, and consistent rule setting (Kling, et al. 2010; Webster-Stratton, 1984), provide parents with more direct means to improve their parenting. As a consequence, it is plausible that their children’s behavior improves more rapidly. Through positive reinforcement, their children’s improvements positively affect the parents’ stress and sense of competence. This might provide an explanation for the larger immediate
effects on parents’ and children outcomes received with the behavioral programs. Future mediation analysis will be needed in order to provide insight into why and how the outcomes of varied programs develop differently.

The non-behavioral programs, on the other hand, change parents’ parenting through altered perceptions and attitudes concerning the parent-child relationship, overall parenting, and child behavior through problem-solving processes (Cunningham, et al. 1995) and reflective exercises (Cunningham et al. 1995; Obsuth, et al. 2006). It is quite plausible that changes in reflection and attitude take more time to manifest themselves, both in parents’ parenting behaviors and in improved child behaviors. Thus, it is plausible that parents participating in the non-behavioral programs do not experience the same immediate positive reinforcement as the parents in the behavioral programs. Instead, the parents’ attitudes and perceptions about their parenting practices and emotions and their children’s behaviors continue to change over time, and the effects become visible later. Hence, change takes time to manifest. The possibility that a behavioral change might be more rapid than a change in attitude could be an explanation for why the parents in the non-behavioral programs reported continued improvements over time.

An explanation for why the parents in the behaviorally-oriented programs reported deteriorating effects could be that after the immediate success, these parents might have “put away the toolbox”. However, “putting away the toolbox” might be one reason that their children’s problem behaviors increased after program end. As a consequence, the parents might have recapped what they had learned and reflected more on the benefits of their changed parenting practices. Thus, reinstating and reinterpreting what they previously learned in the program might have hindered further problem escalation between the one-year and two-year follow-ups. However, further research relating to program impact theory and mediating processes are needed before we can answer these questions with certainty.

In sum, even though we lacked a control condition at the one-year and two-year follow-ups, studies I and II provide much-needed information about the long-term effectiveness of group-based parenting programs. Thus, our findings expand the knowledge and add evidence of long-term effectiveness and development to the existing knowledge base, and so they add significantly to current research.
Do socio-demographic characteristics influence improvements?

In study I, we also investigated whether the programs, overall, worked better for some parents and children compared to others. We found that socio-demographic characteristics such as income, age, child gender, and immigrant status did not influence the overall effects the programs had on the participating parents and their children. Thus, our findings suggested that many parents and their children may benefit from the four programs.

This finding is somewhat surprising, given that there have been indicators that socio-demographic characteristics influence the outcomes of a parenting program in the short term. Still, effectiveness evaluations assessing short-term moderating effects of parental and child characteristics are scarce and somewhat inconclusive (Lundahl, et al. 2006, Kling, et al. 2010; SBU, 2010; Smedler, et al., 2015; Stattin, et al. 2015), and research on long-term moderating effects of parental and child characteristics is, to my knowledge, almost non-existent. There are, however, two exceptions (Hautmann, et al. 2010; Scott, 2005). In a one-year follow-up from a clinical trial of the behaviorally-oriented Incredible Years program with 59 participating parents, the results indicated that children with heightened antisocial behaviors benefited from the program to a larger extent compared to the other children whose parents participated in the program, whereas none of the participants’ socio-demographic characteristics influenced program effectiveness (Scott, 2005). Similar results were found in a study of the Prevention Program for Externalizing Problem Behavior (PEP) with 101 participating parents. The results indicated that the children with the most severe problems and the parents experiencing the lowest levels of parental self-efficacy improved more (Hautmann, et al., 2010). In studies I and II, we only investigated the impact of strictly socio-demographic characteristics (e.g., age, gender, SES) within a Swedish context. It is possible that there are cultural aspects such as social and welfare policies that influence the impact of, for instance, SES on program outcomes. As a consequence, even though our findings add to the existing literature, I perceive that we have only scratched the surface relating to aspects that can potentially moderate program effectiveness. Hence, more research is needed on moderating effects overall.

Does it matter where parents participate?

In study III, our findings indicated that the three welfare sectors (child and adolescent psychiatric care, social care, and schools) did not differ in their ability to deliver the programs. The individual sectors showed parallel
improvements on children’s externalizing behaviors even though their pre-
conditions, such as initial child problem levels and group leader professions, 
differed slightly. We argued that one reason for our results might be due to 
high fidelity or adherence to the program manuals, which was deemed sat-
isfactory for all programs (Stattn, et al. 2015). However, one reason that 
group leader adherence was rated high in the national effectiveness trial 
might very well be due to the overall program evaluation and the fact that 
the group leaders knew that they and their performance were going to be 
monitored and evaluated. We found some support for this argument in 
study IV; studies investigating adherence after program implementation 
found that group leaders commonly adjusted parts of the program to fit 
them or their participants better (e.g., Moore, Bumbarger, & Cooper, 2013; 
Stern, Alaggia, Watson, & Morton, 2008). As a consequence, program con-
tent generally drifted, and it seems that some evaluation and supervision is 
needed to ensure adherence to program content over time (Flay, et al. 2005; 
Gottfredson, et al. 2015). Nonetheless, study III added insight and 
knowledge as to whether different settings or group leader differences may 
have influenced program outcomes. Our results are promising, but before 
we can state that one sector delivers the programs just as well as another, 
additional research is needed, for instance relating to sector equality, feasi-
bility, and cost.

Implementation research on group-based parenting programs

The findings in study IV provided little evidence of a priori planned or 
theoretically driven implementation research on group-based parenting pro-
grams. None of the studies included in study IV evaluated implementation 
aspects during the implementation of the programs. This was surprising, 
but understandable in a sense, given that implementation research is a new 
field within prevention research and so has not drawn much attention. I 
suggest two explanations. The first explanation is the formulation of the 
theoretical approach which is modeled by the established prevention re-
search cycle (Mazrek & Haggerty, 1994). This explanation can also serve 
as an explanation for the current lack of effectiveness trials. The prevention 
research cycle is a sequential model (Figure 2), and aspects of effectiveness 
and dissemination and implementation are entered as the two last compo-
nents in the modeled sequence. Hence, because prior research was focused 
on program development and efficacy trials, the model suggests that there 
is not timely to deal with effectiveness and implementation issues until after
the development and efficacy modules are finished. As a consequence, prevention research might have focused more on continued program development, refinement, and efficacy trials than on how to disseminate knowledge about the programs and implement them in real life settings. Hence, prior researchers were not expected to strictly measure implementation as early as stages two (program development) and three (efficacy trials) of the cycle.

In line with this reasoning I suggest a second explanation. The publication of promising results of the programs, with large decreases in children’s externalizing problems, might have triggered a demand from practice to use these programs. In addition, programs that were developed relatively early (i.e., Incredible Years, PMT-O, SFP, and Triple-P) have gathered a solid evidence base of program efficacy, and there is now time to evaluate them in real world settings. Thus, these programs enter the fourth phase, the evaluation of program effectiveness. As the prevention research cycle suggests, prior research might have been conducted under the assumption that experiences from the effectiveness evaluations will or should inform the fifth phase, large-scale program dissemination and implementation. This has occurred, but it has also brought forth the importance of implementation knowledge as a prerequisite for program effectiveness and sustainability, and the lack of research and knowledge concerning implementation aspects and outcomes. Hence, the second explanation suggests that there was both a demand from practice for preventive interventions as well as the existence of programs ready for evaluation under real life conditions; and, taken together, these circumstances led to the dissemination and implementation of some of the programs without consideration of implementation aspects and outcomes.

By identifying both the complexity of implementation and the influence it has on program efficacy, effectiveness, and sustainability, our findings move preventive implementation research forward. Further, by identifying the existing knowledge gaps regarding the effects of implementation aspects on program outcomes, this dissertation provides insights into important new areas for future prevention research.

**Were there format-specific implementation challenges?**

Another interesting finding from study IV was that the group format of the parenting programs created several implementation challenges which might hinder sustainable program implementation. For instance, recruitment of parents, provision of child care, and group leaders’ workload diffi-
culties (e.g., not receiving over-time) were aspects that group leaders mentioned as challenging when their agencies implemented group-based programs. Long-term commitment and possible invasion of privacy were examples of aspects influencing parents’ decisions to participate in the programs. Even though previous implementation research has not compared different prevention formats, our findings identified several implementation challenges that could relate specifically to the manualized group format of these programs. The parenting groups are commonly offered in the evenings, so they do not interfere with the parents’ work situation. As a result, there is commonly a need for both child care, because the programs are out of school hours, and provision of meals, because it is hard for the parents to attend after dinner time. There is also a possible need to provide transportation to and from the parenting groups. Child care, meals, and transportation are often stipulated or strongly encouraged in many of the programs to ensure parental participation and retention. Hence, the abovementioned challenges ought to be specifically related to many of the group-based parenting programs, and so it is surprising that their impact on the implementation and sustained use of the programs has not previously been considered for further evaluation.

**Limitations and strengths**

This dissertation is not without limitations. First, the national effectiveness evaluation was not conducted as a combined effectiveness-implementation trial. The care unit inclusion criterion, stating that the unit should deliver two parallel parenting programs per semester, meant that several of the participating units chose an additional group-based parenting program for implementation. As a consequence, this trial would have been a good opportunity to study many aspects and outcomes related to implementation, for instance how agency readiness influences recruitment, the group leader’s sense of competence, and/or program outcomes. It could also have been an opportunity to investigate aspects related to fidelity and/or group leader adherence over time. However, at the time of the project, the focus of funders and the research community was on the lack of effectiveness trials, and not on aspects relating to implementation. As a consequence, the aim of the project was to evaluate program effectiveness.

We were not able to keep the control condition past post-test, and therefore we cannot say with certainty that the programs are effective over time. We can only draw conclusions about the programs effects’ in relation to the
short-term results and each other. The results from the short-term evaluation showed that the parenting programs were effective in the short term in comparison to the waitlist control (Stattin, et al. 2015). This meant that, for instance, the children whose parents participated in any of the four programs improved their externalizing problems to a greater extent compared to the children whose parents were randomized to the waitlist control. Over time, though, it is still possible that the children in the waitlist also improved to the same extent as the program participants, and that they all could have had the same levels of externalizing behaviors one year and two years after program end. However, the children with parents randomized to the programs reached the lower levels of problem behaviors sooner, and so they and their parents experienced the elevated levels of externalizing problem behaviors for a shorter period of time. Consequently, it would have been more beneficial for the children in the control condition if their parents had participated in the programs instead of taking a chance and letting “life run its course”. To sum up, we were able to compare the long-term development and outcomes of the four parenting programs with each other even without a long-term control condition. The four programs showed greater improvements in parental behaviors and emotions as well as children’s externalizing behaviors immediately after program end when compared to the waitlist control (Stattin, et al. 2015). Thus, we know that the programs were more effective compared to the control at the end of the programs, and in addition, we can conclude that the programs more or less maintained their effects until two years after program end.

Relating to studies I, II, and III, the national effectiveness evaluation was a quasi-randomized control trial. Neither care units, parenting programs, nor group leaders were randomized in the project; instead, the randomization was conducted on the parental level. Moreover, the parents did not have a similar chance to be randomized to all conditions because they were not randomized to all conditions (four programs, book and waitlist) at the same time (i.e., the three-group design). This was one constraint put on the project due to the real life situation. None of the care units could possibly deliver all four programs simultaneously. Program age limits also hampered randomization to some extent, because Incredible Years and Connect targeted children of different ages (Connect: 8–12, Incredible Years: 3–8). As a consequence, parents were never randomized between Incredible Years and Connect because the programs were never delivered simultaneously. Thus, due to real life limitations, the national trial was a quasi-randomized
control trial, a fact that might lead to inflated type I error. This has to be
taken into account; otherwise, the results might be over-interpreted.

Conducting effectiveness trials under real life conditions and in regular
settings is difficult. On the one hand, research needs to be conducted with
satisfactory levels of scientific rigor, such as the use of randomized con-
trolled trials to exclude the influence of other variables besides the programs
and ensure internal validity. On the other hand, research has to work with
what is available and achievable, and keep research interference to a mini-
mum in order to ensure the external validity of the evaluation (Flay, et al.
2005; Gottfredson, et al. 2015). As a consequence, effectiveness trials are a
constant balancing act in which researchers try to keep both external and
internal validity as high as possible. Even though the national trial was a
quasi-randomized trial, it still provides evidence of how effective the in-
cluded parenting programs are in comparison to each other when they are
delivered by regular care under real life circumstances.

The care units chose the programs that they perceived either best fitted
their mission and values or served as a complement to the program they
already delivered. Internal validity would have been improved if we had
been able to randomize the programs to each care unit, because there might
have been selection biases relating to program choice. However, effective-
ness trials should mirror real life circumstances and delivery as much as
possible (Flay, et al. 2005; Gottfredson, et al. 2015). In everyday practice,
a program would be chosen by a care unit, and not randomized to it. In the
evaluation, a randomization of the programs could have resulted in a per-
ceived mismatch between the program and the care unit/group leader, which
could have influenced program outcomes negatively, for instance through
group leader performance. Hence, by letting the care units choose their pro-
grams, we mimicked the real life circumstances of program selection and
implementation, and thus enhanced the external validity of the national
evaluation even though this meant a reduction in internal validity.

Another limitation is that we relied on parent reports. Our results would
have been more reliable if they had been confirmed by other sources of in-
f ormation, such as independent observations and/or teacher reports. How-
ever, the measures we used to evaluate child and parental behaviors and
emotions are those most commonly used in previous research on group-
based parenting programs (Michelson, et al. 2013). This enhances compara-
brability across studies, and so our findings can be adequately compared with
prior and future research.
Despite the abovementioned limitations, this dissertation has several strengths. First, the findings in studies I, II, and III are based on a large (n=749), well-powered, long-term national sample of parents with high participation over time. The parents were randomized to one of four group-based parenting programs, which provided an opportunity to compare long-term outcomes across programs. As the programs differed in their theoretical orientation, we were also able to make comparison between behavioral and non-behavioral programs within the same design. In most cases, the care units already delivered a behavioral program and chose a non-behavioral program as a complement. Thus the parents were commonly randomized between a behavioral and a non-behavioral program; one exception was Incredible Years and Connect, which were never given at the same time at any of the care units, as mentioned in the limitations above. Nevertheless, the national comparison is a long-term, well powered quasi-randomized evaluation.

Second, the programs were delivered, two-by-two, at 30 care units representing all sectors within the Swedish welfare system (child and adolescent psychiatric care, social care, and schools). The care units used their normal routines to recruit parents, though they might have increased their normal efforts in order to fill the three groups (three-group design: program A, program B, and the control condition). The normal recruitment procedures ensured that a representative sample of parents were recruited as participants to the national trial. Hence, the participation of several care units in the three sectors helped ensure the participation of a representative sample of parents in need of group-based parenting programs. This strengthens our findings concerning sector performance as well as program outcomes.

Third, the trial was conducted by independent researchers without close relationships to program developers, which ensured that the results were not affected by the presence or influence of the developers (Flay, et al. 2005: Gottfredson, et al. 2015). This enhances both the generalizability and the reliability of our findings.

In sum, even though this national evaluation suffered from some well-recognized limitations which threaten the internal validity of our findings, these limitations are balanced by the real life considerations which instead enhanced the external validity of the evaluation. If our results are interpreted with the abovementioned limitations and strengths in mind, they ought to serve as valid indicators of what can be expected when these programs are delivered in regular care. All in all, we made unique independent
long-term comparisons across programs (studies I and II) and between sec-
tors (study III), and our results should be fairly reliable.

Finally, study IV also contributes to the overall strengths of this disserta-
tion as it is the first summary of implementation research specifically on
group-based parenting programs. We made a thorough search in several
electronic databases (the Cochrane, ERIC, Medline, PubMed and PsycInfo).
In addition, we used an established narrative approach (Baumeister &
Leary, 1997) and rated the scientific rigor of all studies using the established
GRADE rating system (Atkins, et al. 2004). As a consequence, we were able
to assess the current state of implementation research on group-based par-
enting programs, and so our findings concerning implementation of group-
based parenting programs and possible challenges should be reliable.

Implications for theory

The ecological model

In this dissertation, I used an ecological system approach as my theoreti-
cal framework (Figure 1); this was also used in study IV. In the discussion
of study IV, we suggested further adaptations to this model in relation to
the implementation of group-based parenting programs. Now, I propose
this adjusted model as a model for overall prevention research in this dis-
sertation (see Figure 6), and the proposed model adds to the present litera-
ture on prevention research. To illuminate how the different levels and their
intricate relationships are related, I will use my experiences from the na-
tional effectiveness trial to illustrate the ecological system. On the societal
level, around 2007, the Swedish government recognized and prioritized the
increase in children’s and adolescents’ mental health problems (problem
identification). The government stipulated that improved child mental
health was a political priority, and as a consequence, the prevention of chil-
dren’s mental health problems became a priority. The government commis-
sioned the Swedish Board of Health and Welfare (Socialstyrelsen) to inves-
tigate and provide a review of the existing preventive interventions delivered
in Sweden, together with a screening of the evidence base relating to each
individual interventions. Both representatives of the Swedish Board of
Health and Welfare and researchers participated in the review group (SBU,
2010), and so several different stakeholders at the same level were involved.

As mentioned in the introduction, the results from the review (SBU,
2010) identified a serious knowledge gap between research and practice.
For instance, the review concluded that of the approximately one hundred
Figure 6. Suggested ecological approach to preventive research
programs that were used to prevent children’s mental health problems at the time of the study, none had been evaluated in a randomized effectiveness evaluation, had long-term follow-ups past six months, or had been evaluated in a Swedish context. Hence, the knowledge gap concerning effectiveness trials was recognized; and as a consequence, the societal funding of the evaluation of preventive interventions increased substantially.

This increase in societal funding, in turn, encouraged practice (organization/practitioners) to implement evidence-based preventive interventions, such as group-based parenting programs, and evaluate them in collaboration with established researchers (society). Or, as in our case, the Swedish Board of Health and Welfare approved funding for a grant proposal regarding an effectiveness evaluation of four established parenting programs. A consequence of this funding was that we recruited 30 care units (agencies): communities (schools and social care) and councils (child and adolescent psychiatric care) who were willing to participate in the national effectiveness trial, thus involving the level of organizations/agencies (the specific sites). The agencies, in turn, recruited practitioners, and engaged other personnel (responsible for recruitment or referral of parents to the project) and group leaders in order to fulfill their commitment to the national trial, which was to run two group-based parenting programs in parallel each semester. As a consequence, there were several stakeholders at several levels involved in the evaluation at this stage.

At the different care units, parents (clients) of 1113 children participated in the project and the majority also attended an intervention, the parenting programs, some after participating in the waitlist control. The children, however, never participated in the programs or the project, but were, according to the parents’ reports in the evaluation, positively affected by the programs. Thus, the improvements in the children’s externalizing problem behaviors were based on parental perceptions, and these improvements were achieved through the parents’ participation in a program. However, it is still largely unknown how the program structure and the parent training influence child behaviors, and further research based on, for instance, program impact theory is needed to provide more knowledge regarding the mediating role of parental behaviors, parenting, and parents’ emotions. These aspects and others, for instance the prevention research cycle, relate to the intervention level of the ecological model.

The parents’ participation meant that the parents attended parent training sessions over several subsequent weeks (time), and also that the pro-
grams affected the children indirectly through the parents. All of these indirect effects, society->research->agency->practitioners->program->parent->child, serve as an illustration of the intricate relationships within the ecological system. Clearly, the shift in policy and recognition of the need for effectiveness trials paved the way for the national effectiveness trial. The national trial, in turn, affected all levels in the ecological system and their relationships.

The prevention research cycle

I have included the level of intervention in the ecological system model. At this level, all aspects of an intervention ought to be considered which include the prevention research cycle (Mrazek & Haggerty, 1994; NRC/IOM 2009). The prevention research cycle illustrates the intricate process that a preventive intervention should pass through, from problem identification, program development, efficacy trials, effectiveness trials, and finally dissemination and implementation. As illustrated above, this cycle is influenced by and influences all levels within the ecological system, for instance through the process of problem identification, funding of research on risk and protective factors, efficacy and effectiveness trials, and the demands from practice for effective interventions. The prevention research cycle illustrates the importance of effectiveness trials that replicate the findings from earlier efficacy trials before a program can be deemed effective and ready for implementation. Our findings in studies I, II, and III indicate that the programs are effective in the long term, and that they can be implemented at all of the three sectors within the Swedish welfare system. Hence, through the involvement of all levels in the ecological system, studies I, II, and III provide some of the information needed at the fourth stage in the prevention research cycle, thus paving the way for the fifth stage, sustained implementation.

Implementation matters (Durlak & DuPre, 2008). The importance of implementation aspects is not only supported by our findings from study IV, for instance that low agency readiness is related to programs’ cost- effectiveness (Romney, Israel, & Zlatevski, 2014). The importance is further highlighted as we know from program impact theory that programs need to be delivered as intended (Rossi, et al., 2004), and implementation research shows that the effectiveness of an intervention depends on the quality of its implementation (e.g., Dane & Schneider, 1998; Durlak & DuPre, 2008; Smith, Schneider, Smith, & Ananiadou, 2004). Still, implementation is com-
plicated (e.g., Fixsen, et al. 2005). In conclusion, it is evident that both theory and research identify implementation as problematic but crucial for program sustainability. As a consequence, implementation aspects and outcomes need to be included in future research, preferably at all stages in the program development and implementation process.

As we have concluded in studies I, II, and IV, prior research has largely failed to take the programs through the last two stages, effectiveness (4) and implementation (5). There seems to have been a leap from efficacy trials of programs showing promising results, and I have provided two explanations for this. The first is that knowledge about the positive outcomes from the efficacy trials of the programs has been spread (disseminated) to practice. Practice, in turn, has implemented the programs to meet these needs without knowledge about the need of confirmatory effectiveness trials providing the evidence base, and also without knowledge about the importance of well conducted implementations. This is a good illustration of the gap between research and practice, but here it also serves as an illustration of the gap between research and theory.

The need for effectiveness trials, both short-term and long-term, is currently well recognized by prevention research (Gottfredson, et al. 2015; NRC/IOM 2009, SBU, 2010; Smedler, et al., 2015; Society of Prevention Research (SPR), 2011). Hence, the need to move on from stage three, efficacy trials, to stage four, effectiveness trials, is underway, and so its importance does not need to be further highlighted in the prevention research cycle. In contrast, there seems to be a need to highlight implementation.

I therefore propose an adjusted version of the prevention research cycle, where implementation aspects are included at almost every phase of the cycle (see Figure 7 for illustration). In this model, the importance of implementation aspects is already present at the beginning, when a new program is being developed. In the originally established version of the prevention research cycle, as suggested by Mrazek and Haggarty (1994), implementation is supposed to influence the whole process. However, as I have argued, it is possible that the sequential structure of the cycle, where implementation come last, diminishes its importance (see Figures 2 and 7 for comparison). Adding implementation as an overarching component of the model, while keeping the fifth dissemination and implementation module intact, both strengthens and identifies the importance of the inclusion of implementation aspects from the beginning and throughout the cycle.
Figure 7. The preventive intervention research cycle

1 Adapted from Mrazek & Haggerty, 1994
In sum, I have proposed adjustment to the two established prevention models that I used as the theoretical framework of this dissertation (and study IV). The models are extended so that additional levels are included and relationships highlighted in the ecological model; and, relating to the level of interventions, an additional overarching implementation component is added to the prevention research cycle. In this way, this dissertation adds to the theoretical approach currently used by prevention research.

Implications for practice

In studies I and II we followed and compared parents who participated in the four parenting programs (Comet, Cope, Connect, and Incredible Years) and their children, one and two years after the programs had been delivered. Our results showed that the immediate short-term differences in child and parental improvements were not sustained over time. The immediate results indicated that the programs differed in effectiveness; Comet and Incredible Years improved, for instance, children’s behaviors to a slightly greater extent compared to Cope and Connect, but our longer-term comparison revealed that both children and parents benefited, with no significant differences across the programs. For practice, this is important knowledge because it indicates that it does not matter what program (of the four tested here) the care unit chooses to implement; one will probably be as good as another over time. As a consequence, the care units can more freely consider the programs that best fit their values and missions. However, because the children whose parents participated in the behaviorally-oriented programs showed more rapid immediate improvements in their behaviors, these programs could be used when the need for change is more urgent, as with indicated and selective preventive interventions.

Practice needs to recognize that it is not enough to select and implement an evidence-based parenting program. That a program builds on a solid theoretical base and/or has been evaluated, whether in efficacy trials or effectiveness trials, is not enough to guarantee that the program will render similar outcomes when delivered at their particular care unit. Instead, program effects or outcomes depend on how well the program is implemented, and there are important aspects of implementation that practice needs to take into account. Hence, knowledge about implementation aspects that might appear due to, for instance, a group-based parenting program, is important. As an example, our findings in study IV indicate that agency management needs to recognize the importance and need of adequate resources, such as training enough group leaders, allocating funds for meals and child
care, scheduling evening sessions, and allowing program supervision when implementing a group-based parenting program. After investing funds and resources in the initial steps of implementation, it is important to follow through with all of the aspects of the implementation of the program. Otherwise, an effective program might turn out to be a waste of monetary as well as human resources. Hence, stakeholders at all levels within the ecological system; that is, policy makers, managers, and practitioners, need to recognize and address implementation aspects when considering the dissemination and implementation of new parenting programs.

The findings of study III indicate that the programs can be delivered in several sectors within the Swedish welfare system, at least as long as the group leaders adhere to the program manuals. This is important information for practice because it can enhance service delivery through co-operation between sectors, for instance through shared program recruitment and delivery. Enhanced service delivery such as increased co-operation might make delivery more cost-effective and also increase the opportunities for parents to attend the programs. Hence, another important aspect of implementation is that the program is delivered with high program fidelity, which means that the care unit and their group leaders adhere to the program structure and content, or that the program is delivered as intended. It is then, and only then, that any program can be said to be evidence-based. A recent review of the effectiveness of group-based parenting programs (Furlong et al, 2012) concluded that the studies where content fidelity was rated high, in general, produced larger effects (moderate compared to small non-significant effects). Hence, adherence to the content of program manuals is important for program effectiveness. As mentioned in the introduction, the adherence ratings in the national evaluation were relatively high (Stattin, et al. 2015), and our findings suggest that the programs can be delivered in several settings within the welfare system as long as the group leaders adhere to the program manuals.

Changes in program structure, for instance, delivering a program to an extended group of parents, might be tempting in order to increase reach or because many parents want to participate, and might not seem to be a big alteration. However, group size; that is, the minimum and maximum number of participating parents, as well as the composition of the parenting groups, influences aspects such as group leader workload. As an example, according to the manual for the 11-session Comet program, the parenting groups should consist of parents to 3–6 children. Hence, the minimum number of participants is 3 and the maximum is 12, depending on whether the
participants are single parents or couples. Parents participating in the Comet program can, and should if needed, receive individually tailored homework assignments. Also, the sixth session of Comet is not group-based but an individual session. Adding one or two extra parents/couples to the program would clearly increase the workload for the group leaders, by increasing the number of individual sessions and possibly also the number of tailored home assignments. As an example of the impact of the evaluation of program delivery, we asked for and received permission to maximize the Comet groups to 8 participants under the condition that we provided any necessary extra resources, because the group size of the Comet program was limited in comparison to the other programs (maximum range = 12–30). However, given the provision of extra resources, this change should not have altered the results. Without the allocation of extra resources it might have been necessary for the group leaders to decrease the tailored homework or the time spent on the individual sessions, and these actions could have had a direct and negative effect on the program’s effectiveness.

In conclusion, then, our results suggest that practice should be able to choose the program (of the four evaluated programs) that best fits their mission and values. However, practice needs to acknowledge the influence of implementation aspects on the effectiveness of evidence-based parenting programs, for instance high adherence to program manuals. Our results indicate that the programs can be delivered at all sectors within the welfare system if adherence is high. Failure to consider implementation aspects might lead to a considerable waste of scarce resources, and ultimately mean that effective interventions are not delivered to the parents and children in need of support, thus possibly increasing children’s mental health problems instead of decreasing them.

**Future research directions**

The theoretical frameworks (the ecological model and the prevention research cycle) I have used in this dissertation are well established (NRC/IOM, 2009). Nonetheless, as we concluded in study IV, research theory and research practice seem to be far apart, which was surprising. Consequently, the first implication for future research directions that I suggest in this dissertation is that coming research should prioritize the use of a well-developed theoretical approach when designing future effectiveness and implementation evaluations of group-based parenting programs.

Overall, there is a lack of program evaluations. This applies to both to effectiveness and implementation research. As shown in studies I, II, and IV,
only a few programs have been evaluated in effectiveness or implementation trials. Hence, there is a great need for further evaluations of other programs, including all aspects of effectiveness, such as effectiveness of the programs, settings, and costs. Equally important is that these evaluations are combined with implementation aspects and outcomes. Ultimately, these programs are conducted over an extended amount of time. Given that some of the group-based programs target children from the age of three, it is important to follow the development of these children longer than one or two years, and, if possible, into young adulthood. These are important aspects which ought to influence the approach to funding and design of future evaluations.

Next, researchers need to recognize that there is a difference between evaluating parenting programs which are already implemented at the care units, and evaluating parenting programs which are implemented due to a current effectiveness evaluation. In the former case, one might think that the stage is already set and implementation cannot readily be investigated in full, though aspects of implementation should be incorporated in the evaluation nonetheless. Information on implementation aspects such as fidelity, adherence, recruitment, and management support can still be evaluated and put in relation to each other and to program effectiveness. In the latter case, when programs are implemented in regular care settings due to a trial, researchers should design the evaluations as combined long-term effectiveness and implementation trials. The ultimate goal of the implementation of a program is sustained use, and so we need to have knowledge of how to implement the programs successfully. These evaluations need to be conducted by the program developers as well as by independent researchers.

The need for randomized control effectiveness and implementation trials is another important practical aspect that future research needs to address. Both short-term and long-term randomized trials are needed. One suggestion to work around the problem with long-term untreated control groups could be to follow a randomized cohort of children and their parents over time, using measures that are appropriate to measure change due to preventive interventions. It would be beneficial to have normative long-term information regarding the behaviors and emotions that parents and their children experience. At present, there is a lack of long-term normative development relating to the measures most commonly used. Even though the ultimate comparison to program outcomes will always be randomized long-term control conditions, the proposed normative group of parents and their children could serve as a way around the problem surrounding randomized untreated control conditions. The data gathered would also have the benefit
of serving as a national comparison which could be used as a comparison condition by several preventive interventions with related outcomes.

Equally important is to recognize the lack of group-based parenting programs targeting internalizing problems. In study IV, we conducted a thorough and broad search for implementation studies on children’s mental health problems, both externalizing and internalizing, and did not come across any study evaluating programs targeting internalizing symptoms. Also, when reviewing prior long-term effectiveness trials in preparation for studies I and II, we found no studies evaluating program effectiveness on children’s externalizing and internalizing problems. Hence, internalizing problems seem to have drawn little attention in prior research. This is surprising given the well-recognized comorbidity between externalizing and internalizing problems (e.g., Biglan, Flay, Embry, & Sandler, 2012). Thus, a program targeting children’s externalizing problems through improved parenting (e.g., increased responsiveness, warmth) would probably affect the children’s emotional well-being as well. As a consequence, future research needs to address and measure both aspects of children’s mental health in order to provide a full coverage of the prevention of children’s mental health problems.

The development of measures appropriate to evaluate changes in children’s behaviors or emotions when the problem levels are beneath severe or clinical levels is also an important task for future prevention research. The currently used measures of externalizing behavior problems, such as ECBI (Eyberg & Ross, 1978), CBCL (Achenbach & Edelbrock, 1981), and SDQ (Goodman, 1997) measure what are perceived as problematic externalizing behaviors. However, they have commonly been used to measure change in clinical or close to clinical samples (e.g., Kling, et al., 2010; Scott, 2005; Webster-Stratton, Hollinsworth, & Kolpacoff, 1989). As such, they might not be sensitive enough to detect the small changes that can be expected in, for instance, a universal preventive approach to the programs. When used in regular care settings, the purpose of the programs is, in many cases, to prevent problems from occurring. Hence, the programs are not only delivered to parents with children where the externalizing behaviors are perceived as problematic; but, at least in Sweden, are also delivered to parents who do not yet perceive that their child experiences any problem behaviors. Many parents want to prevent externalizing problems from occurring by becoming better parents. In a sample of parents who currently do not perceive their children as problematic, the currently available measures would probably show the effects of a program to be lower than expected or even
nil. As a consequence, an otherwise effective program might be deemed ineffective when instead the lack of significant outcomes relates to inadequately sensitive measures. Consequently, research needs to investigate the appropriateness of current measures and, if needed, to develop new measures adequate for preventive purposes at all levels of prevention: indicated and selective as well as universal.

There is also a need to know more about how the programs work. Research needs to address questions about which components of the programs affect parents and how they are affected. It is important to understand how a program works and what leads to the desirable change in child behaviors and emotions (Kaminski, Valle, Filene, & Boyle, 2008; Kraemer, Stice, Kazdin, Offord, & Kupfer, 2001; Rossi, et al. 2004). Prior research has focused on the program’s effects on child behaviors. Thus, the primary outcomes of the parenting programs have been measures assessing change in the children’s behaviors; but the change is, theoretically, mediated by changes in parents’ parenting and possibly also in their emotions. As a consequence, little is currently known about the mediating processes, changes in parents, which are supposed to lead to the desired positive changes in the children’s behavior.

More research is also necessary concerning the influence of moderation. Our findings from study I, that parental and child characteristics do not influence the effectiveness of the programs, only serves as a first indicator and a scratch on the surface. More research is needed concerning other characteristics of parents and children which might influence program effectiveness, such as initial levels of parental stress, emotion regulation strategies, or children’s problem levels. In addition, further research is needed on combined characteristics, for instance whether young single mothers are affected differently compared to older single mothers or mothers with a partner. In line with this reasoning, we also need to know more about whether a specific program works better for some parents compared to other programs; for instance, whether a behaviorally-oriented program suits a certain group of parents better than an attachment-based program, and vice versa. Understanding what program works best for whom ought to make prevention delivery more effective. Hence, this is important knowledge for both research and practice.

**Summary and conclusion**

The purpose of this dissertation was to fill some of the existing knowledge gaps concerning the effectiveness and implementation of group-based parenting programs. As a consequence, an otherwise effective program might be deemed ineffective when instead the lack of significant outcomes relates to inadequately sensitive measures. Consequently, research needs to investigate the appropriateness of current measures and, if needed, to develop new measures adequate for preventive purposes at all levels of prevention: indicated and selective as well as universal.
based parenting programs. Studies I and II indicate that the outcomes of the four parenting programs did not differ over time, and they seem to be effective for all parents and their children over time. Because the three welfare sectors which delivered the programs did not differ in their overall delivery of the programs, the results from study III suggest that the setting is not particularly important. Finally, study IV identifies and highlights the lack of implementation research on group-based parenting programs, but suggests several implementation challenges which might affect the implementation of group-based parenting programs. In addition, I have proposed adjustments to the theoretical models I used as the theoretical framework in this dissertation, several directions for future research, and various implications for practice. Even though the need for future research is evident, this dissertation has filled some of the existing knowledge gaps.

Conclusions

- Both effectiveness and implementation research on group-based parenting programs are lacking. Ultimately, future research should incorporate theoretically driven randomized long-term trials combining both effectiveness and implementation evaluations.
- Long-term evaluations are crucial to properly establish program effectiveness. Program effects continue to change over time, and therefore short-term effects cannot provide evidence of effectiveness.
- The effects of programs developed on different theoretical bases seemingly develop through different trajectories. The behaviorally-oriented programs seem to result in a faster decrease in children’s problem behaviors compared to the non-behaviorally-oriented programs. However, even though the four programs (Comet, Cope, Connect, and Incredible Years) differed immediately after program end, these differences were evened out over time and there were no significant differences across the programs at either the one-year or the two-year follow-up.
- Socio-demographic characteristics such as income, age, child gender, and immigrant status do not influence the outcomes of the programs. This suggest that the programs are effective for any par-
ent and child. However, more research is needed on moderating effects in general but also concerning, for instance, parental stress, emotion regulation strategies, and children’s problem levels.

- There is a need to know more about how parental change influences child behaviors and emotions. Mediating questions or questions about which components of the programs affect parents, how they are affected, and how this relates to child outcomes need to be answered.

- The three welfare sectors (child and adolescent psychiatric care, social care, and schools) do not differ in their ability to deliver the programs. A practical implication could be enhanced intervention delivery through enhanced co-operation between sectors.

- It is unclear whether there are any group-based parenting programs targeting children’s internalizing problems. This needs to be investigated further.

- Externalizing and internalizing problems co-occur, and therefore internalizing outcomes should be included in future program evaluations.

- Recruitment of parents, provision of child care, and group leaders workload difficulties (e.g., not receiving over-time) were identified as challenges relating to the group format of the parenting programs. These challenges could hamper program implementation.

- Adjustments were made to both the ecological system model and the prevention research cycle used as the theoretical framework in this dissertation. The adjustments in the ecological model identify the importance of aspects relating to, for instance, the practitioners, the interventions, and the participants. On the intervention level, the adjustments to the prevention research cycle enhance the importance of implementation aspects at every stage of the prevention research cycle.
References


Achenbach, T. M. (1990). Conceptualization of developmental psychopathology. In M. lweis & S. M. Miller (Eds.), *Handbook of developmental psychopathology* (pp. 3-14). New York: Plenum


Lochman, J. E., & Wells, K. C. (2004). The Coping power program for preadolescent aggressive boys and their parents: Outcome effects at the


10. Persson, Andreas, Leisure in Adolescence: Youth’s activity choices and why they are linked to problems for some and not others. 2006.


12. Larsson, Mats, Human Iris Characteristics as Biomarkers for Personality. 2007.

* Finns sedan tidigare utgiven i serien ”Örebro Studies”.


