

From the Department of Public Health Sciences,  
Division of Social Medicine,  
Karolinska Institutet, Stockholm, Sweden

**Primary Prevention of Mental Health  
Problems among Children and Adolescents  
through Social and Emotional Training in  
School**

Birgitta Kimber



**Karolinska  
Institutet**

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

© Birgitta Kimber, 2011

ISBN 978-91-7457-372-5

Printed by



[www.reproprint.se](http://www.reproprint.se)

Gårdsvägen 4, 169 70 Solna

## ABSTRACT

Among younger people in many high-income countries, mental ill-health, which includes depression, aggressive behavior, feeling down, and alcohol and drug abuse, is one of the greatest health problems. Since most young people attend school, there are grounds for pursuing the prevention of ill-health in the educational arena. A set of techniques, named social and emotional learning (SEL), based on cognitive and behavioral methods, is available to teachers to train students to improve self-control, social competence, empathy, motivation and self-awareness. SEL programs have their underpinnings in the theories of cognitive development and social learning, and in application of the ideas of risk and protective factors. The primary aim of this dissertation is to describe and evaluate, in a real-life setting, the impacts of a Swedish program derived from SEL, called social and emotional training (SET), on various mental-health outcomes. Such programs have been shown to have favorable effects in the international literature, but have not been tested before in Sweden. Sub-aims were to investigate whether there were outcome differences between subgroups, and to assess the development of an instrument for the measurement of social emotional maturity. The evaluation was performed in two experimental and two control schools (41 and 20 classes, respectively) in Botkyrka Municipality in Greater Stockholm. A variety of statistical analyses were applied to the data collected: two repeated-measures cohort analyses, with rather different designs, to measure changes over two and five years; latent-class analysis to examine variability and substance use; and, latent growth curve modeling with full maximum likelihood estimation to scrutinize our earlier findings. On the social and emotional variables, the impact of SET was found to be generally favorable. After five years, the impact of SET was found to be greater for internalizing than for externalizing problems, but no impact on social skills was detected until a quadratic (curvilinear) model was fitted to the data. Weaknesses in SET implementation and in our research approach are highlighted and discussed under certain themes. Project experiences indicate needs for wide community involvement, and greater discipline in administration, and the benefit of using a variety of study designs and statistical approaches in the interpretation of results.

Key words: mental health, prevention, schools, social and emotional learning.



## LIST OF PUBLICATIONS

This dissertation is based on the following five papers, which are referred to in the text by their Roman numerals:

- I Kimber, B., Sandell R., & Bremberg S. (2008a). Social and emotional training in Swedish classrooms for the promotion of mental health: results from an effectiveness study in Sweden. *Health Promotion International* 23, 134-143.
- II Kimber, B., Sandell R., & Bremberg S. (2008b). Social and emotional training in Swedish schools for the promotion of mental health: an effectiveness study of 5 years of intervention. *Health Education Research* 23, 931-940.
- III Kimber, B. & Sandell R. (2009). Prevention of substance use among adolescents through social and emotional training in school: A latent-class analysis of a five-year intervention in Sweden. *Journal of Adolescence* 32, 1403-1413.
- IV Sandell, R. & Kimber, B. (submitted). Heterogeneity of response to a universal prevention program.
- V Sandell, R., Andersson, M., Elg, M., Fhärm. L., Gustafsson, N., Kimber, B., & Söderbaum, W. (submitted). A psychometric analysis of a measure of socio-emotional development in adolescents.

Papers 1-III are reprinted with the permission of the journals concerned.

# CONTENTS

1	Introduction.....	1
2	Background.....	2
2.1	Life Skills and SEL programs .....	3
2.2	The SET program .....	6
2.3	Some evaluation aspects.....	8
3	Aims.....	10
4	Method .....	11
4.1	Population .....	11
4.2	Instruments.....	12
4.3	Occasions of questionnaire administration .....	14
4.4	Procedures.....	14
4.5	Study designs and statistical analyses .....	14
5	Results.....	23
6	Discussion.....	28
6.1	Summary of the SET studies in Sweden with some reflections.....	28
6.2	Key themes emerging from the implementation and evaluation of SET .....	30
6.2.1	Evidence and effectiveness with regard to SET .....	31
6.2.2	Attrition.....	31
6.2.3	Levels of analysis .....	32
6.2.4	Social skills.....	33
6.2.5	Delivery of SET: where and by whom? .....	35
6.3	Development of SET .....	36
6.4	Future research.....	36
6.4.1	What we could have done better.....	37
6.4.2	What we still can do in the SET project .....	38
6.4.3	Suggestions for future research.....	38
7	Concluding remarks .....	40
8	Appendix: validation analysis .....	41
8.1	The data set for the validation analysis .....	43
8.2	The variables considered in the validation analysis .....	44
8.3	Results of the validation analysis .....	46
8.4	Summary of the comparison .....	48

9 Acknowledgements .....49

10 References .....50

## LIST OF ABBREVIATIONS

ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
CAN	Centralförbundet för alkohol och narkotikaupplysning (the Swedish Council for Information on Alcohol and Other Drugs)
CFI	Comparative Fit Index
EI	Emotional Intelligence
ENSEC	European Network for Social and Emotional Competence
EQ-I	Emotional Quotient-Inventory
EU	European Union
FIML	Full Information Maximum Likelihood Estimation
HIF	How I Feel
IOM	Institute of Medicine
ITIA	I Think I Am
LC	Latent Class
LCRA	Latent Class Regression Analysis
LGM	Latent Growth Curve Modeling
MANCOVA	Multivariate Analysis of Covariance
MANOVA	Multivariate Analysis of Variance
MSCEIT	Mayer-Salovey-Caruso Emotional Intelligence Test
No SET	No Social and Emotional Training
PATHS	Promoting Alternative Thinking Strategies
PS	Propensity Score
RCT	Randomized Controlled Trial
RMSEA	Root Mean Square Error of Approximation
SAFE	Sequenced step-by-step training approach, use <b>A</b> ctive forms of learning, <b>F</b> ocus sufficient time on skill development, and have <b>E</b> xplicit learning goals
SEL	Social and Emotional Learning
SET	Social and Emotional Training
SFL	Skills for Life
SJT	Situational Judgment Test
SSRS	Social Skills Rating System



UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations International Children's Emergency Fund
WHO	World Health Organization
YSR	Youth Self Report



# **1 INTRODUCTION**

This dissertation is about the social and emotional training program that took place in a suburb of Stockholm between 2000 and 2005. The summary is organized as follows. First, the background and content of the intervention are described. Second, there is an account of the various evaluations that have been performed, and also of the development of a new instrument to measure emotional development or maturity. Third, there is a discussion of the various theoretical and practical issues that have arisen, and of how the organizers of the program and its evaluation have responded to them. This summary draws on the papers already published or submitted, but there are new sections concerned with statistical validation and SET practice. Also, recent research papers concerning social and emotional learning (SEL) within the relatively new but expanding field of prevention science (Stattin & Kerr, 2009) are referred to throughout. Finally, in view of the significance of the final validation analysis, a full account of it is attached as an appendix.

## 2 BACKGROUND

Among people in many high-income countries, aged 1-44, mental ill-health, which includes depression, aggressive behavior, feeling down, and alcohol and drug abuse, is the greatest health problem. Specifically, internalizing problems, such as depression, account for a larger proportion of mental ill-health than externalizing problems (Murray & Lopez, 1997). In Sweden, both in primary care and in hospitals, mental ill-health is one of the most prominent broad categories of illnesses (Allebeck, Diderichsen, & Theorell, 1998).

Given that the targeted resources of child guidance clinics and school health services are limited, there is a case for universal interventions for the prevention of ill-health among the young. Since virtually all children go to school, the school is an obvious arena for mental-health promotion. For example, the World Health Organization (WHO, 2003, p. 6) states: “The school is an appropriate place for the introduction of life skills education because of:

- the role of schools in the socialization of young people;
- access to children and adolescents on a large scale;
- economic efficiencies (uses existing infrastructure);
- experienced teachers already in place;
- high credibility with parents and community members;
- possibilities for short and long term evaluation.”

It is stated in the UN Convention on the Rights of the Child (UNICEF, 1989) that “education of the child should be directed to ... the development of the child’s personality, talents and mental and physical abilities to their fullest potential” (Article 29, 1a). It has been claimed that “the central tenet of [Article 29] is that education is not just a matter of fostering cognitive-academic development, but should be directed at the overall, i.e. physical, cognitive, social, emotional and moral, development of the child. Consequently, educational systems or institutions, such as schools, that exclusively or predominantly focus on academic development violate children’s rights” (Diekstra & Gravesteyn, 2008, p. 7). At the same time, it has been suggested that addressing social and emotional issues may counteract school failure: “Studies specifically examining the

causes of school failure have found that emotional and learning disorders are amongst the most important risk factors” (Patel, Flisher, Nikapota, & Malhotra, 2008, p. 315).

## **2.1 LIFE SKILLS AND SEL PROGRAMS**

A set of educational techniques, named social and emotional learning (SEL), based on cognitive and behavioral methods, is available to teachers to train students to improve self-control, social competence, empathy, motivation and self-awareness, and has shown promising results in the US (Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2002). SEL and its derivative in Sweden (SET) form a subset of Skills for Life (SFL) programs (WHO, 1997, 1999). Life skills are defined as “1) social and interpersonal skills (including communication, refusal skills, assertiveness, and empathy); 2) cognitive skills (including decision making, critical thinking and self-evaluation); and 3) emotional coping skills (including stress management and increasing an internal locus of control) (Mangrulkar, Whitman, & Posner, 2001, p. 5). Social as well as emotional aspects are important: “Children’s ability to develop positive peer relationships is critical to their wellbeing. Compared to children who are accepted by their peers, socially rejected children are at substantially elevated risk for later adjustment troubles, including academic underachievement, school dropout, criminal activity, and psychiatric problems” (McKown, Gumbiner, Russo, & Lipton, 2009, p. 2).

Life-skills programs in general and SEL programs, including the Swedish social and emotional training (SET) program, have their underpinnings in cognitive-development theories (Piaget, 1972; Vygotsky, 1978), social learning theory (Bandura, 1977), and application of the ideas of risk and protective factors (Arthur, Hawkins, Pollard, Catalano, & Baglioni JR., 2002; Durlak, 1998). For an overview, see Mangrulkar et al. (2001).

From a developmental perspective, during school age (ages 6-16) children develop the ability to think abstractly, to understand consequences, to relate to their peers in new ways, and to solve problems. Within this age span the skills of young people vary a lot, and activities therefore have to be developmentally appropriate. SEL programs teach social and emotional skills to different age groups in different ways that are designed to be age-appropriate. Relating to others in the social environment has a strong influence on the structure of young people’s thinking, and cognitive skills can be enhanced through interactions with others (Vygotsky, 1978).

With regard to social learning (social cognitive) theory, children learn to behave through both formal instruction and through observation. Teachers and parents are often involved in the instruction, and they are models for how to behave. In SEL programs teachers are trained and encouraged to use the skills that are taught in their everyday contact with pupils. One way of working together with parents is to prepare homework that encourages parents to take part in the teaching. Children also learn how to behave simply by observing adults and peers. This influences SEL programs in that teachers are looked upon as important role models, and also in that the teaching of social and emotional skills involves modeling, observation and social interaction.

In terms of risk and protective factors, there is an emphasis on the need to modify and promote children's healthy development. There are both internal factors (e.g. self-esteem, self-confidence, and sense of self-efficacy) and external factors (e.g. relationships with peers with positive behaviors, a non-violent home environment, strong bonds with the school, academic success) that can interact to help overcome problematic or difficult situations. Many of the skills taught in SEL programs are designed to enhance children's self-esteem, mastery and self-confidence, and also to help children bond with the school. To know how to manage emotions is viewed as a key skill in a SEL setting (Hawkins et al., 1992). By enhancing children's protective factors, they can resist the ill-health that often results from stressors or risks. Social, emotional and cognitive skills may serve as mediators for behavior. Life skills build competencies rather than address behavior directly. Via active learning, e.g. role play, problem-solving and situational analysis, young people can be engaged in their own development process. Teaching interpersonal cognitive problem-solving skills to children can prevent and reduce serious problems later in life (Spivack & Shure, 1994), and are therefore a critical part of life-skills programs. "By teaching young people *how* to think rather than *what* to think, by providing them with the tools for solving problems, making decisions and managing emotions, and by engaging them through participative methodologies, skills development can become a means of empowerment" (Mangrulkar, et al., 2001, p. 20).

Durlak and colleagues (2011) have published a meta-analysis of 213 school-based, universal social and emotional learning (SEL) programs covering 270,034 students from kindergarten through to high school, run by school or non-school personnel, or a mixture of the two, using six outcome criteria: SEL skills, attitudes,

positive social behavior, conduct problems, emotional distress, and academic performance. They state: “Classroom by Teacher programs were effective in all six outcome categories, and Multicomponent programs [with school-wide as well as classroom elements] (also conducted by school staff) were effective in four outcome categories. In contrast, classroom programs delivered by nonschool personnel produced only three significant outcomes (i.e., improved SEL skills and prosocial attitudes, and reduced conduct problems). Student academic performance significantly improved only when school personnel conducted the intervention” (Durlak, et al., 2011, p. 413).

Here, it is worth mentioning a recent assessment of research developments in the field, based on reports of the Institute of Medicine (IOM), which is the health branch of the government-independent National Academy of Sciences in the US: “Overall, research on school-based mental health and competence promotion has advanced greatly during the past 15 years. The Institute of Medicine’s (1994) first report on prevention concluded there was not enough evidence to consider mental health promotion as a preventive intervention. However, the new Institute of Medicine (2009) report on prevention represents a major shift in thinking about promotion efforts. Based on its examination of recent outcome studies, the new Institute of Medicine report indicated that the promotion of competence, self-esteem, mastery and social inclusion can serve as a foundation for both prevention and treatment of mental, emotional, and behavioral disorders” (Durlak, et al., 2011, p. 420).

SEL programs, which were formerly prevalent only in the US, have now spread to some extent in Europe, e.g. to Germany (von Marées & Petermann, 2010) and Portugal (Moreira, Crusellas, Sá, Gomes, & Matias, 2010). They have their underpinning in many academic studies (Durlak & Weissberg, 2005; Durlak & Wells, 1997; 2007; Greenberg, 2004; Greenberg, Domitrovich, & Bumbarger, 2001; Shochet et al., 2001). Also, they are recommended by international institutions, such as the World Health Organization (WHO, 1997), the United Nations Educational, Scientific and Cultural Organization (UNESCO, 2006), and the European Union (EU, 2005). The European Network for Social and Emotional Competence in Children (ENSEC), which was set up in 2007, describes its mission in terms of being “devoted to the development and promotion of evidence-based practice in relation to socio-emotional competence and resilience amongst school students in Europe” (ENSEC, 2007).

It was in light of the research findings on SEL that SET was first developed for application in a Swedish setting.

## **2.2 THE SET PROGRAM**

The SET program was implemented in Sweden between 2000 and 2005. It was designed by the author of this dissertation (see papers I and II), and was delivered by regular class teachers during scheduled hours. The teachers taught SET to junior and intermediate students (grades 1-5) twice a week, each session with a duration of 45 minutes, and senior students (grades 6-9) one 45-minute session a week over the total school year. The program is guided by detailed manuals for the teacher, one volume for each grade. It also includes a workbook for students of each grade. Altogether, the program consists of 399 concrete exercises, as specified in the manuals and workbooks. Some of the tasks are inspired by programs in the US, in particular Promoting Alternative Thinking Strategies, known as PATHS (Greenberg, 1996). As a further example, the self-control unit in SET is a modified version of the Stoplight Model used in the Yale-New Haven Middle School Social Problem-Solving Program (Weissberg, Caplan, & Bennetto, 1988).

SET focuses on helping to develop the following five functions of the students:

1. Self-awareness – being aware of what one is feeling and thereby being able to use one’s feelings when taking decisions, making realistic assessments of one’s own capacities, and having a sound self-confidence.
2. Managing one’s emotions – knowing why one is feeling a certain way, and how to handle one’s feelings so that – instead of being destructive – they may aid coping with tasks, and enable control of feelings and waiting for rewards in order to achieve a goal.
3. Empathy – understanding how others feel and seeing things from their perspective, recognizing that others feel differently, and being able to cope with and understand the differences between oneself and them.
4. Motivation – using one’s own internal “engine” for goal achievement, learning to take the initiative and strive for improvement, managing setbacks and frustrations on the path to goal achievement, and being able to put up with any reward having to come later.
5. Social competence – being able to handle emotions in relation to others, to recognize social situations, and to manage in different social environments. This entails

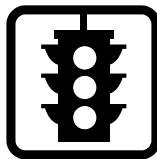


being capable of utilizing one's feelings for cooperation, negotiation and conflict resolution, for the handling of other people's feelings, and for utilizing various tools in conflict and problem situations.

Accordingly, there were five separate, albeit overlapping, components to the program. Typically, the components merge into one another, and therefore an exercise according to the manual may address several functions. The following themes recur in the tasks: responsible decision-making, problem-solving, coping with strong emotions, appreciating similarities and differences, clarification of values, conflict management, interpretation of pictures and narratives, doing more of what makes one feel good, resisting peer pressure and being able to say "No", knowing what one is feeling, recognizing people and situations, cooperation, communications skills, setting goals and working to attain them, giving and receiving positive feedback, and stress management.

For example, when the children are 6-7 years old, they use a traffic light as a symbol for problem-solving and handling strong emotions. They are presented with fictitious situations, but can also use the symbol when they have a real problem or conflict.

The red light symbolizes stopping and calming down. It is explained to the children that just like cars driving against a red light, they can hurt themselves or others unless they calm down before they act when they have been upset. The yellow light symbolizes thinking about possible solutions to the problem, and also about the consequences of different solutions. Children are encouraged to consider what they want, their goal in this situation, and how to achieve it. Examples of goals are playing with somebody, or being able to borrow a toy. The green light stands for "Go", try your best solution. If it does not work, try one of the others you have.



When the students get older they use a metaphorical model for resolving problems or conflicts. Just like the younger students they are presented with fictitious situations, and are also encouraged to use the model in real-life situations. The model is called [www.solutions.com](http://www.solutions.com), where www stands for the three "w"s below. It is not a website,

but serves as a reminder of the steps involved in problem-solving.

This is how they are invited to think about the **problem** they have.

**w** – Who is involved and what are their feelings?

**w** – What is the problem?

**w** – What is the goal?

**Solutions** – find as many as you can

**c** – Consequences of each solution

**o** – One solution is selected

**m** – Make sure to evaluate and learn

Teachers are instructed to use modeling and role-play as key elements in the exercises, and students should not only practice in school but also outside school (including the home). The desirability of interaction between school and parents is emphasized.

The author trained the teachers in SET in the school year 1999/2000. During this school year they had an opportunity to try out the relevant exercises themselves, and test them in their classes. They were encouraged to raise methodological and technical issues, and discuss remaining problems. The teachers were supervised once a month during the school year 2000/01 and offered supervision on a voluntary basis during 2001/02. Several independent ratings were performed of each SET teacher during the first two years of the program; an interview survey of a random sample of the teachers was conducted in 2003 after two years of program implementation (Gadd, 2003).

### **2.3 SOME EVALUATION ASPECTS**

The published literature reveals three recurrent weaknesses to the evaluation of school-based intervention studies.

First, as stated by Greenberg about evaluations of social and emotional learning (SEL) programs in general, “[m]ost evaluations have assessed programs that have lasted one school year or less. ...In contrast it has been well-recognized that educators perceive the need for multiyear programs that are of sufficient duration and are integrated with other multigrade curricula” (Greenberg, 2010, p. 157).

Second, most peer-reviewed studies published so far have been conducted in the US, and the generalizability of their results to other cultures and countries cannot be taken for granted.

Third, most of the studies report on efficacy trials, undertaken with a research team in charge, rather than effectiveness trials conducted in a community setting (Greenberg, 2004; Marlowe, 2004). The internal validity of efficacy studies may be satisfactory, but there is much to be asked about their external validity.

This study of the Swedish SEL project (with acronym SET, for Social and Emotional Training) attempts to address these relative shortcomings. First, being a multi-year program, SET covers mandatory preschool and all grades of compulsory school (1-9). Second, the project was conducted in a European country, namely Sweden. Third, the evaluation was implemented in a real-life community setting.

### **3 AIMS**

The primary aim of the studies was to describe and evaluate, in a real-life setting, the impacts of a Swedish social and emotional learning program (SET) on various mental-health outcomes. Sub-aims were to investigate whether there were differences between subgroups with regard to outcomes, and to assess the development of an instrument for the measurement of social and emotional development and maturity.

## **4 METHOD**

### **4.1 POPULATION**

In Sweden compulsory schooling encompasses preschool at age 6 and then school at grades 1-9. Children begin school at age 7 at grade 1 and end at grade 9 at age 16 (before going on to high school). Most children go to schools close to where they live.

The SET evaluation (papers I, II, III, IV) was carried out in Botkyrka Municipality, located in the Stockholm metropolitan area. In Botkyrka there are eight schools that cover all compulsory schooling, i.e. preschool and all grades (1-9). The study participants attended grades 1-7 in four of these schools, and responded to the questionnaires at baseline (t0) in August 2000, and then in May of each year from 2001 (the first year of intervention = t1) through to 2005 (the final year of intervention = t5). Students attending grades 1-3 at baseline were named juniors, while those attending grades 4-7 at baseline were called seniors. Two of the eight schools in Botkyrka were chosen as intervention (SET) schools. For comparative purposes, a control (No-SET) school of similar size serving a socioeconomically similar population was selected for each SET school.

There were a total of 110 classes in the two SET schools taken together; one had six classes per grade, the other five. Three classes at each of the first seven grades (1-7) within the two SET schools were then chosen on an organizational basis, i.e. from the same building or from among the particular classes for which a deputy head-teacher had responsibility, thus making 42 experimental classes in total. One class dropped out for administrative reasons, giving a final total of 41 experimental classes. The No-SET classes were chosen by the head-teachers of these schools, one for each grade (14 in total). The population was defined by those who responded to the questionnaire at t0. One junior student and two senior students did not obtain parental permission to respond in the SET schools; there were no such cases in the No-SET schools.

For testing the How I Feel (HIF) instrument, various versions were administered alongside the SET instruments in May each year between 2001 and 2005 (Paper V). Paper V, however, is based solely on data from 2005, although it uses additional data from 2004 to estimate the stability of the HIF. Re-test reliability was explored in a separate study, with a different population, where the HIF was administered, together with a self-rating questionnaire, on two occasions at an interval of three weeks. The

special re-test study was carried out on 119 students in grades 4, 8 and 9 in three schools in Botkyrka. One of these schools used the SET program but did not participate in the main study; the other two neither used SET nor participated (as controls) in the SET study.

## 4.2 INSTRUMENTS

All the instruments employed in the evaluation of SET (papers I, II, III and IV) were well-established and had documented reliability and validity.

I Think I Am (ITIA) is the Swedish self-rating instrument, “Jag tycker jag är” (Ouvinen-Birgerstam, 1985), which has roots in previous American research (Coopersmith, 1967). It is intended to map the young person’s self-image and self-esteem, and has subscales for body image, family relations, relations with others, talent/abilities, and psychological well-being.

There are two versions of the instrument: ITIA-I for younger students (grades 1-3) and ITIA-II for older ones (grades 4-9). In ITIA-I students are instructed to answer “Yes” or “No” to 32 questions in total. In ITIA-II students respond to statements on a four-point scale, “Exactly like me”, “Almost like me”, “Very little like me”, “Not at all like me” (72 items in total). Examples of ITIA items are: “I have a nice face”, “I like myself”, “I am often sad”, “My parents trust me”. Higher scores indicate more positive self-image.

Students in grades 4-9 also responded to a second questionnaire with the following elements:

Youth Self-Report (YSR) (Achenbach & Edelbrock, 1987), used in an abbreviated Swedish version (Lindberg, Larsson, & Bremberg, 1999), measures mental-health symptoms and problems. The abbreviated version has been shown to have psychometric qualities comparable to the original. Questionnaire items are rated on a three-step response scale “Not true”, “Somewhat or sometimes true”, “Very true or often true”. Besides the two subscales, internalizing problems and externalizing problems, suggested by Lindberg and colleagues, four new subscales were derived on the basis of a principal-components factor analysis. These were named: anxiety, e.g. feeling worthless or inferior and feeling unhappy; aggressiveness, e.g. threatening to hurt people, destroying things that belong to others; assertiveness, e.g. stubborn, moods or feelings changing suddenly; and attention-seeking, e.g. trying to get a lot of attention, bragging. The lower the score, the better the outcome.

Mastery (Pearlin, Liebman, Menaghan, & Mullan, 1981), in one of its Swedish versions, is a nine-item four-step self-rating scale, with responses to statements ranging from “Strongly agree” to “Strongly disagree” (making 36 different responses possible). In their original article, which was about the stress process, Pearlin and colleagues suggested an instrument to measure feelings of self-efficacy or hopelessness, defined as the extent to which one regards one’s life chances as being under personal control. Examples include: “There is really no way I can solve problems I have”, and “I have little control over the things that happen to me”. Higher scores indicate higher sense of self-efficacy. It is worth emphasizing the conceptual affinity between self-efficacy, mastery and locus of control.

The Social Skills Rating System (SSRS) (Gresham & Elliott, 1990) consists of 34 items for grades 4-6, and 7 additional items for grades 7-9, all with four-point response scales, “Never” (0), “Sometimes” (1), “Often” (2), “Very often” (3). The ratings were also scored on four subscales: cooperation, assertion, self-control, and empathy. Higher scores indicate greater social skills.

Contentment in school, or school satisfaction by analogy with job satisfaction, refers to a single item, “How do you like it in school?”, taken from a Swedish health-behavior questionnaire administered annually by the Swedish Council for Information on Alcohol and Other Drugs (CAN) (Hibell et al., 1997). Contentment was rated on a five-step response scale, ranging from “Very good” to “Very bad”. Higher scores indicate greater satisfaction.

Bullying in three aspects (being insulted, physically assaulted, or frozen-out) was assessed on three-step response scales, ranging from “Yes, often” to “No, seldom or never” ranging from “Very good” to “Very bad”. Higher scores indicate fewer problems.

Drug (substance) use refers to the use by students (only those in grades 7-9) of tobacco (7-step scale, ranging from “Never” to “Every day”), alcohol (9-step scale ranging from “Do not drink” to “Every day”), volatile substances (3-step scale, ranging from “No” to “Yes, several times”), and illegal narcotics, or simply drugs (7-step scale, ranging from “Never” to “More than 50 times”). The lower the score, the less is use on each item.

The How I Feel (HIF) instrument is designed to measure emotional development or maturity, and has been developed in successive versions since 2001.

The HIF is a situational judgment test (SJT), based on brief vignettes, where the protagonist (in some vignettes “you,” in others “he” or “she”) is described in situations of intrapersonal or interpersonal dilemma. An example of a minor classroom incident, where a researcher responds to misbehavior in class, is given in Paper V (p. 7). Each vignette is followed by two questions, “What do you feel, and why?” (the Feel item) and “What do you do?” (the Do item), each with three response options. Initially, there were 15 vignettes, thus 30 items. After the psychometric analysis, there are now 14 vignettes, with 28 items.

### 4.3 OCCASIONS OF QUESTIONNAIRE ADMINISTRATION

The occasions of administration of the various instruments are shown in Table 1 below:

**Table 1.** Times of instrument administration by grade.

	<b>t0 May 2000 (baseline)</b>	<b>t1 May 2001</b>	<b>t2 May 2002</b>	<b>t3 May 2003</b>	<b>t4 May 2004</b>	<b>t5 May 2005</b>
<b>I Think I Am (ITIA I)</b>	Grades 1-3	Grades 1-3	Grades 2-3			
<b>I Think I Am (ITIA II)</b>	Grades 4-9	Grades 4-9	Grades 4-9	Grades 4-9	Grades 4-9	Grades 5-9
<b>Youth Self-Report (YSR) Mastery Social Skills (SSRS) Contentment in school Bullying</b>		Grades 4-9	Grades 4-9	Grades 4-9	Grades 4-9	Grades 5-9
<b>Substance use</b>		Grades 7-9	Grades 7-9	Grades 7-9	Grades 7-9	Grades 7-9
<b>How I Feel</b>		Grades 4-9	Grades 4-9	Grades 4-9	Grades 4-9	Grades 4-9

### 4.4 PROCEDURES

The questionnaires were handed out each May by deputy head-teachers, and administered during school hours by regular class teachers. The teachers were encouraged to follow the written instructions and to make efforts to ensure that the students understood the questions. The questionnaires were then relayed back to the deputy head-teachers and forwarded for data entry by an independent organization.

### 4.5 STUDY DESIGNS AND STATISTICAL ANALYSES

Four of the papers (I – IV) concerned the evaluation of the SET program, while Paper V focuses largely on the development of an instrument to measure socio-emotional



development and level of maturity, which was a by-product of the SET evaluation. A later validation analysis was conducted, which is presented in full in an appendix, as well as in the body of this summary.

Two of the papers (I and II) reported repeated-measures cohort analyses, with rather different designs, while Paper III on substance use and Paper IV on variability among the students, employed a form of latent-class analysis. Although Paper V was largely about the development of an easy-to-use measure of emotional development and maturity, it does bear on SET evaluation in some respects, particularly with regard to social skills. The validation analysis addressed various problems of interpretation with the earlier analyses, including attrition, intra-classroom dependencies, and the possibility of non-linear relationships. Latent growth curve modeling (LGM) was employed.

The study reported in Paper I had a quasi-experimental longitudinal design, covering students of all grades over the first two years of SET. The study was quasi-experimental in the sense that the schools were not chosen at random; the two intervention schools, one in a relatively poor area the other in an area of medium socio-economic status, were selected to match the intervention schools in terms of their size and socio-economic catchment area. It was longitudinal in that cohorts of students were compared at two points in time; only students with full data on both occasions were considered.

Differences between the groups (SET and No-SET) in their development from May 2001 (t1) to May 2002 (t2) on each scale or subscale were analyzed separately by running a repeated-measures ANCOVA (or MANCOVA). Note that the questionnaires administered to junior and senior students were different. SET or No SET and year (t1 and t2) were the independent variables, and the scale (or subscale) of each instrument the dependent variable(s). The five ITIA subscales at baseline (t0) were used as covariates after standardizing each scale within each school level, i.e. separately for ITIA-I and ITIA-II (see above). The GLM routine of SPSS, version 11, was used. Significance was set at  $\alpha=.05$ . Using Becker's (1988) approach, between-groups effect sizes were computed for each dependent variable from unadjusted (raw score) means and standard deviations at t1. In this study, no adjustments were made for intra-classroom or intra-school dependencies, but the issues concerned were addressed in our validation analysis (see below).

Paper II describes a study with a mixed design, in which there is “a mixture of between-group and repeated-measures variables” (Field, 2005, p. 483) to compare students in the SET and No-SET schools according to duration of SET or No SET (1 to 5 years), regardless of grade (5 to 9). All students in the data set were included, but questions on substance use were only posed to students in grades 7 to 9. Given a student’s grade at t1, t2, t3, etc., we formed a variable for duration of the SET program (number of years). We then compared the mean trajectories on each outcome measure between students in the SET schools and the No-SET schools as a function of the number of years that the program had been running.

Differences between the groups (SET and No-SET) in their development from t1 to t5 were tested in three different ways. SPSS version 12 was used for the statistical calculations.

For each of the outcome variables a linear regression analysis was performed for each student group, which provides measures of the linear trends as effects of the intervention. Adopting Becker’s (1988) approach, change effect size parameters and between-groups effect sizes (Becker’s  $\Delta$ ) were computed for each dependent variable. Cohen’s (1988) classification of effect sizes (small =.2, medium=.5, large=.8) was employed. ANOVAs (or MANOVAs, when we analyzed an instrument with subscales, such as the YSR and the ITIA) were run on the outcome scale (or subscales), with intervention (SET or No SET), number of years (t1, t2 ...t5), and student gender as independent variables. Given significantly different mean changes on the unstandardized regression coefficients, the critical effects were the differences between the intervention-by-years interaction in the two groups. The GLM routine of SPSS, version 12, was used.

The study reported in Paper III, of substance use among students in grades 7 to 9, had a quasi-experimental, i.e. non-randomized, five-year mixed longitudinal and cross-sectional design, which compared students receiving the SET intervention with those who did not. Nonparametric latent class regression modeling with repeated measures was employed to analyze the data. Given a student’s grade at t1, t2, t3, etc., we formed a variable for duration of receipt/non-receipt of the SET program (number of years).

Due to the natural turnover of students in schools, a complete repeated-measures design across the five years would not have generated a sufficiently large

sample to allow any meaningful analysis. We therefore decided to compare the trajectories on each outcome measure according to number of years (duration) of SET/No SET and grade. We wanted to test whether there was (a) a differential change in the use of specific substances according to number of years of SET/No SET, and (b) a differential change in the use of specific substances by grade between SET students and No-SET students. Such changes, which might indicate treatment effects, would be reflected in significant interactions between intervention and years (duration), and intervention and grade (age), respectively.

We performed nonparametric latent class (LC) regression analysis with repeated measures (Vermunt & Van Dijk, 2001) to identify classes (segments) and then analyze the substance-use variables. As pointed out in the Latent GOLD user's guide (Vermunt & Magidson, 2005), nonparametric LC analysis has the advantages of being applicable to ordinal-level data, and is less subject to biases due to violations of conventional assumptions about linearity, normality, homoscedasticity, independence, and homogeneity.

The model included the following independent variables (SET or No SET, a dichotomy; 2 or 1, respectively), the number of years of receipt of SET (or non-receipt, in the case of the control group), years (5 categories: 1 to 5), grade (3 categories: 7 to 9), and their interactions. Thus, we created three new variables by calculating the products of SET/No SET and years, SET/No SET and grades, and years and grades (Jaccard, Turrisi, & Choi, 1990). For each substance-use variable separately, we regressed the repeated scores for each student on these six variables.

Further, in light of the non-randomized design, we stripped the outcome variables of variance components that might have arisen from selection-based differences between the SET and No-SET students. This was achieved by estimating a propensity score (PS) (Bartak et al., 2009) for each student and adding propensity as another independent variable. It has been suggested that the PS procedure is a promising way of correcting for selection bias in quasi-experimental studies (Rosenbaum & Rubin, 1983). In this study, each PS was based on the students' baseline measurements with regard to five different aspects of well-being and adjustment (from the ITIA) and sex, and to socio-economic status of school-catchment/living area (not a measure at individual level).

The outcome parameters were the unstandardized regression coefficients (slopes) for years (duration) and grade (age), representing each student's estimated average rate of change according to number of years or grade, respectively, and the intercept, which is the student's estimated score at baseline (t0) following covariate control.

For the study reported in Paper IV we continued to adopt a latency approach. We utilized a mixed longitudinal and cross-sectional design to analyze outcome trajectories for different subgroups of students as a function of the duration of the SET program, students' grade each year, and intervention status (SET or No SET). The full data set was employed. The idea was that the general analyses reported in papers I and II would obscure heterogeneity in the samples. We subjected the data to latent class regression analyses (LCRA) with repeated measures to identify subgroups with differential trajectories on each outcome measure.

We used linear modeling to compare the trajectories in the two groups (SET and No-SET) on each outcome measure according to number of years of implementation of the SET program. We expected a gradual improvement with duration of exposure to the SET program, i.e. the number of years in the program, and no comparable improvement in the No-SET group. Thus, intervention effects would be reflected in significant interactions between SET or No SET and number of years (duration). Also, we expected that a general deterioration with age, i.e. across grades, would be mitigated in the SET group but not in the No-SET group, generating significant interactions between SET/No SET and grades. However, assuming that the students constituted a highly heterogeneous group, we expected that these outcome contrasts between the SET and the No-SET groups would vary between subgroups of students.

We conducted nonparametric LCRA with repeated measures (Vermunt & Van Dijk, 2001) to identify the latent classes and then analyze the outcome variables. The model included the following independent variables: SET or No SET (a dichotomy), number of years of SET or No SET (5 categories), grades (5 categories), and their interactions. In order to test the hypothesized interactions we created three new variables by calculating the products of SET/No SET and years, SET/No SET and grades, and years and grades (Jaccard, et al., 1990). For each outcome variable separately, we regressed the repeated scores for each student on these six variables and years (1 to 5), grades (5 to 9), and SET/No SET (1 or 2).

As in our analysis of substance use (Paper III), we stripped the outcome variables of variance components that might have arisen from selection-based differences between the SET and No-SET students by estimating a propensity score (PS) (Bartak et al., 2009) for each student and including the PS as another independent variable.

The outcome parameters were the unstandardized regression coefficients (slopes) for years (duration) and grades (age), representing each student's estimated average rate of change according to number of years or grade, respectively, and the intercept, which is the student's estimated score at baseline ( $t_0$ ). We used the Latent GOLD 4.0 software (Vermunt & Magidson, 2005). Given our hypothesis that the outcome comparisons between the SET and the No-SET group would vary between classes, all significance tests for slopes were two-tailed.

The study that assessed properties of a new instrument to measure social and emotional maturity (SEM), the How I Feel (HIF) instrument (reported on in Paper V) was different in kind from the five evaluations of the SET program, and had only a peripheral bearing on the evaluation of SET per se. HIF was developed over five years, but the assessment was largely based on one year's data (2005), although data from 2004 were employed to examine the stability of the instrument. Essentially, HIF scores were compared with scores on the SET-evaluation instruments. The scoring of each item was based on expert judgments, using a Thurstone type of scaling procedure (Dawis, 2000; Edwards, 1983). A testee's total score was computed as his or her average score across all items.

The instruments with which the HIF was compared were all self-reporting and had been used for evaluation of the SET program. In these comparisons, particular emphasis was placed on the measure of social skills, given the conceptual affinity between SEM and social skills.

The analysis proceeded as follows. First, we computed some summary descriptive statistics: means, medians, standard deviations, skewness and kurtosis. Second, we analyzed the factor structure among the items, using principal-components factor analysis to evaluate the dimensionality of the instrument. Third, we investigated reliability, using Cronbach's  $\alpha$  and test-retest over a period of three weeks. Retest reliability was tested in a separate sample and separately in classes with and without SET. Fourth, we examined stability by computing product-moment

correlations between 2004 and 2005. Retest reliability was tested in the SET and No-SET schools separately. Fifth, validity was considered by exploring relations between the HIF and other variables that have been reported to be associated with indicators of SEM. These included sex, age/grade, substance use, bullying, as well as the relations between the HIF and the SET-evaluation instruments. These relations were examined in a principal-components factor analysis with oblique rotation (direct oblimin).

Finally, the ability of the HIF to detect treatment effects was tested. An ANOVA was performed with HIF score as the dependent variable and SET/No SET, grade and student sex as between-subjects independent variables. Only students who had continued to be in the SET program from its inception were included.

Following discussion of the results presented in papers I, II, III and IV, we performed a series of analyses using different statistical techniques in an attempt to validate our findings. We used the largest and most wide-ranging of our data subsets, namely the one employed for our five-year follow-up of the effects of the SET program on the social and emotional variables (Paper II), but we had to set inclusion criteria for the analysis. To allow for the possibility of a quadratic growth model, we had to exclude any student who had not filled in the same set of questionnaires on at least three of the five occasions of measurement. In effect, this meant that we were restricted to students in grades 4, 5 and 6 at t1, for whom we could compute both intercept and slope estimates (SET = 443 students; No SET = 101 students).

In essence, we used latent growth curve modeling (LGM) with full information maximum likelihood estimation (FIML). The Mplus software was used (Muthén & Muthén, 1998-2010). There were several advantages to utilizing these techniques:

1. The data had longitudinal attrition. In our previous analyses, we did not impute missing values, on the ground that the non-random distribution of the missing data made imputation unsuitable” (Paper I). In the validation analysis, we adopted the FIML approach, which “estimates model parameters and standard errors using all available raw data” (Enders, 2001, p. 715). FIML does not impute or fill in missing data values but estimates the model parameters and their standard errors based on the full data set. The computational algorithm of FIML is based on the assumption that missing values are related to observed values of other variables in the set. FIML in the MPlus software also provides adjusted standard error

estimates, which is a useful safeguard against inflation of the Type-I error rate. Enders (2001) points to evidence that the FIML estimator is superior to other techniques for dealing with missing data.

2. Given that the SET program was implemented by teachers within classrooms, its effects may have varied according to teachers/classrooms. In addition, and more fundamentally, students who were exposed to the same teacher/classroom environment may have shown greater similarities with each other than they would have with students in other classrooms. That is, the observations within a given classroom may not have been independent due to clustering, and in turn, the assumption of independence of observations may have been violated. Accordingly, we took into account the clustering of the data using the Type=Complex option in MPlus.
3. We used LGM (Duncan & Duncan, 2004) to compare trends in growth in the treatment (SET) and comparison (No-SET) groups in order to estimate the program effect for each major outcome variable. The LGM approach has advantages over the ANOVA approach for the analysis of change in longitudinal data. In addition to its flexibility in comparison with ANOVA, LGM – like other latent-variable approaches – accounts for measurement error. LGM also models group-level growth rates and patterns by taking into account the initial status of individuals and variability within groups.

In the current analysis, as a first step, we fitted a single group LGM model to the data to identify the overall growth pattern for each outcome variable. We first fitted a linear growth model. If the model revealed poor fit, then we fitted a quadratic growth model. As a second step, once the overall growth pattern had been identified, we fitted a conditional growth model, in which a group variable identifying the treatment (SET = 1) and comparison (No-SET = 0) groups was included as a time-invariant covariate.

A significant path coefficient from the covariate to the intercept factor would suggest that there is a significant initial difference between the groups. Similarly, a significant path coefficient from the covariate to the slope factor would suggest that the observed growth pattern is different across the control and the treatment conditions. We considered the following variables, which are described in detail in Paper I, and analyzed over five years in Paper II: Youth Self-Report (YSR),

internalizing; YSR, externalizing; mastery; I Think I Am (ITIA), total; contentment in school, bullying; and, social skills (SSRS), total.



## 5 RESULTS

After two years, social and emotional training was found to have some favorable small-to-medium effects on mental health and health-related behaviors (Paper I). The dropout rate was high. Among the SET students, 48% of the senior-level students measured at baseline remained after two years of the intervention, but only 26% of the junior-level students, although it has to be remembered that roughly one-third of junior students per year will disappear as a matter of course as they advance from junior to senior level (i.e. 66% over two years).

Considering effects as a whole, there were positive impacts – albeit not always statistically significant – on 4 out of 5 of the scales for the juniors (the exception being body image), and 18 out of 20 for the seniors (the exceptions being mastery and cooperation). For the junior sample, there was a large effect size for psychological well-being, although it was not statistically significant ( $p = .074$ ). For the senior sample, there were statistically significant ( $p < .05$ ) medium effect sizes for body image, relations with others, psychological well-being, aggressiveness, attention-seeking and bullying.

Surprisingly, given the program's focus on social as well as emotional aspects, there was virtually no recorded differential impact on the social skills scales (assertion, cooperation, empathy, and self-control). SET also appeared to have had no favorable impact on mastery, defined as the extent to which one regards one's life chances as being under personal control. If hopelessness and lack of self-efficacy are construed as internalizing problems, like YSR anxiety, it appears that the program had stronger effects on externalizing problems. The typical result pattern was not so much that the SET students improved, but that the No-SET students deteriorated with regard to the aspects of mental health considered.

After five years, the impact of SET was shown to be generally favorable (Paper II). Relating duration of social emotional training to various outcomes associated with mental health, significant positive associations were found on five out of the seven dependent variables considered: YSR internalizing, YSR externalizing, mastery, ITIA (total), and contentment in school. Effect sizes were medium.

In the SET schools bullying was at a continuously low level, whereas in the No-SET schools the level varied strongly from year to year, but – with regard to duration – it was found that there was no difference in trend between the SET and No-SET

groups. SET may offer a means of providing greater continuity in this arena in that peak incidences in the level of bullying are avoided.

The five-year follow-up revealed significant duration lags on some variables. It appeared that there was a greater beneficial effect of SET on internalizing than externalizing problems, but this only emerged after three to four years. In the case of mastery, three years of SET seem to have been needed before the program had a detectable impact, and in the case of the ITIA (which measures self-image and self-esteem) four years. Social skills remained an exception; there was no detectable effect on the SSRS (Gresham & Elliott, 1990).

Although the repeated-measures analyses were cross-sectional, the sample on which these analyses were performed was subject to attrition. Obviously, some SET participants and controls did not respond over five years, or even over two. In an attrition analysis we have, however, shown that the differential outcomes between the SET and No-SET groups cannot be explained away by selective attrition within the SET group, i.e. that students with poorer mental health were less likely to respond over longer periods.

With regard to substance use (Paper III), statistically significant intervention-by-duration interactions, with medium to large effect sizes to the advantage of the SET students were found for all substances in one or more, but not all, of the latent classes we had generated. Favorable trajectories were found for non-users/light users of drugs, moderate sniffers, non-users/light users of alcohol, and occasional smokers. Assuming that degree of substance use is an indicator of mental ill-health, programs like these, given a duration of two years or more, may dampen increases in use with grade/age and discourage early debut, even though they are not specifically targeted at use itself.

As might be expected from a universal primary-prevention program, the effects were found to be heterogeneous with regard to level and trajectory of use. It should be noted that the classes (non-users, light and moderate users, etc.) emerged naturally from the LCA, and were not created by the researchers in advance. Thus, for each substance, one, or sometimes two, of these outcome classes displayed the expected SET/No SET-by-years interaction, indicating a gradual divergence between the mean SET and No-SET trajectories over time. Medium to large positive effect sizes for SET were recorded for selected subgroups, including once-or-more drug users, once-

or-more users of volatile substances, and drinkers of alcohol six times a year or more. The picture with regard to smoking is less clear. It has to be emphasized, again, that the positive effects were limited to specific subgroups and emerged only gradually over the five years. Further, there is some evidence that, heavy smokers excepted, SET has a dampening effect on the increase of use with grade (age), especially on drug use, but again only in specific subgroups and then even less dramatically.

Taking all scales into account, outcomes were found to be systematically heterogeneous (Paper IV). On all the outcome variables at least two significantly different classes were distinguished, at different levels and with different change trajectories. As expected, individual students responded differently to SET, but there was some patterning that enabled them to be divided into classes. Latent class regression analysis (LCRA) provided for a great increase in outcome variance accounted for – from around 5% in the whole group to 50-60% when broken down into latent classes.

The intervention effects of the SET program varied both between classes and between outcome variables. Generally speaking, in all classes where there was a significant years-by-SET/No-SET interaction, it was in favor of the SET group. On mastery, the ITIA and the SSRS, the interactions were quite strong for one or two classes; on the YSR, they were more modest, revealing unstable developments in the No-SET group. Although there appeared to be floor or ceiling effects for some classes on some outcome variables, there were no indications that the SET intervention had a differential impact on low-risk or high-risk groups (i.e. groups with a more or less favorable initial level on a particular outcome variable).

On social skills, as measured by the SSRS, the LCRA offered a telling demonstration of the consequences of neglecting outcome heterogeneity. The non-significant intervention effect found in the undivided student group was found to conceal two opposite interactions that balanced each other out. For about one third of the students there was a negative interaction, such that the outcome development was less positive in the SET group, whereas for almost as many, themselves divided in two classes, there was a positive interaction. In one of these classes, the difference between the SET and the No-SET group was quite dramatic.

Our validation analysis of the evaluations of social and emotional training in Sweden, based on latent growth curve modeling (LGM), largely verified our previous

findings, albeit with some important modifications. The validation analysis is presented in full as an appendix to this dissertation.

Three model-fit estimates were employed: Chi squared ( $\chi^2$ ), the Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA). Models with linear growth patterns fitted the data for four of the seven outcome variables with good model fit indices. The exceptions were externalizing problems, social skills, and bullying. A quadratic growth model fitted the data well for externalizing problems and social skills. Neither a linear nor a quadratic growth model fitted the data for bullying. Accordingly, change in bullying was not further examined.

The results of the LGM analyses suggest that the students in the treatment condition (receipt of SET) had significantly higher internalizing problems and lower school contentment than the comparison group (No-SET) on the first occasion of measurement (t1). There were no other initial between-groups differences.

The results suggest consistent program effects on the outcome measures. In the treatment (SET) group internalizing problems decreased and externalizing problems remained stable, whereas both problems increased in the comparison (No-SET) group. In addition, externalizing problems in the No-SET group showed an accelerating increase over time. Also, feelings of mastery and contentment in school in the SET group remained stable, which can be compared with the significantly decreasing trends observed in the No-SET group. Next, we observed a significant decrease in ITIA scores in both groups, but the rate of decrease for the No-SET group was over three times greater than for the SET group. Finally, the students in the treatment group displayed no change in perceived social skills, by contrast with the quadratic decreasing trend observed for the control group students. This was the first of our analyses to suggest a favorable impact of SET on social skills.

Although the detailed statistics from the repeated-measures and latent-growth analyses are not directly comparable, and differences are difficult to quantify due to adjustments to both the scoring and the intercepts and slopes, the directions of the earlier findings are largely confirmed. This applies to internalizing, mastery, the I Think I Am instrument, and contentment in school. The relationship between SET and externalizing appears to be quadratic rather than linear. Further, the validation analysis suggested that the initial differences between the SET and No-SET groups were somewhat larger than we had supposed, and that there was indeed a significant effect – in the quadratic model – of SET on social skills.

We also performed a psychometric analysis of a measure of socio-emotional development and maturity in adolescents (Paper V). Our initial observations were of negative skewness and high mean scores for most items and, consequently, for the total score. The item factor solution was highly stable, judging from a comparison with corresponding results from administration of the instrument in 2004. Both the test-retest reliability and the internal consistency of the HIF were relatively high, comparable with measures of emotional intelligence (EI), such as the EQ-i (Bar-On, 2004) and the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) (MacCann, Matthews, Zeidner, & Roberts, 2004). Its stability across one year was also reasonably high in the No-SET schools, where no systematic intervention had influenced the natural development of the students.

The validity of the HIF was tested in a number of analyses. There was a between-sexes difference, which was in line with previous research (Mayer, Caruso, & Salovey, 1999) that was not detected the Social Skills Rating System (SSRS), as applied in the SET evaluation, which suggests that the HIF has incremental validity in this respect. There were significant negative, albeit low, correlations between the HIF and different forms of substance use, and there was a significant negative correlation between the HIF and bullying. These results support the construct validity of the HIF.

Significant correlations between the HIF and measures of mental health problems and self-efficacy were found. Together with the association with the SSRS, particularly on its empathy and cooperation subscales, these results provide evidence for the discriminant and convergent validity of the HIF as a measure of SEM. HIF scores did not increase with age; rather, they followed a rather strong negative trend in the No-SET group.

The ability of the HIF to detect intervention effects was supported by the relative offset of this negative age/grade trend in the SET group. Comparative analyses of the SSRS, mastery, the YSR, the ITIA, bullying and contentment in school resulted in similar, though weaker patterns, indicating that this may be an intervention-related phenomenon rather than an instrument-specific artifact. Further support for the sensitivity of the HIF was the attenuated test-retest reliability and stability findings in the SET group in comparison with the No-SET group.

## **6 DISCUSSION**

This discussion contains a brief summary of the overall research findings, reflections on some methodological issues, and some comments on substantive issues relevant to the delivery and development of social and emotional learning programs. The methodological and substantive issues are organized under themes that have emerged as important during the implementation and evaluation of SET. The strengths and weaknesses of the project, including its evaluation, are considered under the headings: Evidence and effectiveness with regard to SET; Attrition; Levels of analysis; Social skills; Delivery of SET where and by whom. Finally, some issues relevant to the development of SET and future research are considered.

### **6.1 SUMMARY OF THE SET STUDIES IN SWEDEN WITH SOME REFLECTIONS**

Evaluation of the SET intervention in Sweden has been presented in four papers in which comparisons are made between outcomes for SET and No-SET students, and also in a separate validation analysis (see Appendix). There is one main data set (plus a further supplementary set for testing the HIF instrument), but five analyses, covering different time periods and using different statistical techniques. The first study concerned all aspects of SET after two years among all students, except for substance use among grades 1-6 (Paper I); the second, emotional and social aspects after five years among senior students (Paper II); the third, substance use among grades 7-9, again after five years (Paper III); the fourth, heterogeneity of responses to the program (Paper IV). Finally, we performed a validation analysis, which is reported upon in this summary, and also appended in full.

On the social and emotional variables, we found the impact of SET to be generally favorable. After two years (Paper I), there were positive impacts, albeit not always statistically significant on 4 out of 5 of the scales, covering social and emotional aspects for the juniors, aged 7 to 10 (the exception being body image); and on 18 out of 20 of the scales for the seniors, aged 11 to 16 (the exceptions being mastery and cooperation). After five years (Paper II), the impacts of SET on the social and emotional variables, by contrast with our findings after two years, were found to be greater for internalizing than for externalizing problems. In the validation analysis, when we tested alternatives to a linear relationship, a better fit-to-data for externalizing was found in a quadratic (curvilinear) rather than in a linear model (see

Figure 4 in the Appendix). There was no evidence of any effect of gender or socioeconomic status (as defined by school catchment area).

Our findings are broadly in line with the view that the effects of early interventions appear later for internalizing problems. For example, in a study of self-report depressive symptomatology over six years among children from grade 2 to grade 8, Mazza and colleagues identified the trajectories with regard to depression of latent groups of individuals with both internalizing and externalizing problems. The authors make a strong case for early intervention on the basis of their findings: “To be proactive in preventing and reducing depressive symptomatology, universal intervention programs ... should be implemented in elementary and early middle school” (Mazza, Fleming, Abbott, Haggerty, & Catalano, 2010, p. 590).

There was virtually no recorded differential impact on the social skills scales (assertion, cooperation, empathy, and self-control) after either two years (Paper I) or five years (Paper II). However, when we adopted a latent-class approach to see whether our overall findings obscured systematic outcome heterogeneity (Paper IV), we found that the non-significant intervention effect found in the undivided student group concealed two opposite interactions. Also, in our validation analysis, we found a significant difference between the SET and No-SET students in relation to social skills after fitting a quadratic (curvilinear) model to the data.

With regard to the substance-use items, i.e. smoking, drinking, sniffing and consuming alcohol, the overall SET effect was non-significant after two years, although the MANCOVA showed a significant positive effect for alcohol, and a close-to-significant ( $p=0.051$ ) effect for narcotic drugs (Paper I). For five-year follow-up (Paper III), students in grades 7-9 were divided into latent classes. Favorable trajectories were found for non-users/light-users of drugs, moderate sniffers, non-users/light users of alcohol, and occasional smokers. Only in the case of heavy smokers was a detrimental effect of SET detectable. The weakness of our results on smoking is in line with the findings of a recent study of a school-based substance-abuse prevention program in which effects were found for alcohol and drugs, but not for smoking (Faggiano et al., 2010).

Assuming that degree of substance use is an indicator of mental ill-health, programs like SET, given a duration of two years or more, may lessen increases in use with grade/age and discourage early use, even though they are not specifically

targeted at use itself. In particular, with regard to alcohol, there is the growing view that if it “is initiated before age 13, the person is more likely to have school-performance problems and display delinquent behaviors, e.g. marijuana use” (Peleg-Oren, Saint-Jean, Cardenas, Tammara, & Pierre, 2009, p. 1966). In retrospect, we feel that we should have taken measurements at early grades, especially of alcohol and smoking, in order to capture very early use.

There was a significant effect of SET on bullying after two years, but no significant difference between the SET and No-SET groups after five. We noted, however, that bullying was at a fairly stable mean level in the SET group, whereas it was quite variable in the No-SET group (as reflected in a highly significant SET/No-SET by years interaction). In our validation analysis, we were unable to fit either a linear or quadratic model to the bullying data.

There was no effect of SET on contentment in school after two years, but a highly significant effect, with a medium effect size, after five years.

The How I Feel (HIF) instrument was developed as an easy-to-use measure of the process of socio-emotional development or the level of socio-emotional maturity (SEM). It proved to have limited discriminatory power among individuals at high levels of socio-emotional maturity. Internal consistency and retest reliability were satisfactory, as too was year-to-year stability. By contrast with previous research (Geher & Renstrom, 2004; Mayer, et al., 1999), HIF scores did not increase with age, implying that, on this instrument, they did not develop socially and emotionally over the years; rather, they followed a weak negative trend in the SET group and a strong one in the No-SET group. The same trend was found for the SSRS scores, for mastery, the YSR, the ITIA, and substance use. These results suggest a general age-related phenomenon, as has been described by Moffit (1993). It may be questioned whether one and the same set of items can accurately measure SEM across such a broad age range as the one covered by the SET program, i.e. students aged 7 to 16. Our attempt to create vignettes that were equally relevant and applicable throughout the school-age span was probably not entirely successful.

## **6.2 KEY THEMES EMERGING FROM THE IMPLEMENTATION AND EVALUATION OF SET**

Some issues that have emerged as important in the course of implementation, evaluation and discussion of the project are now addressed. The themes are both



scientific and practical.

### **6.2.1 Evidence and effectiveness with regard to SET**

The concept of evidence-based medicine, or more broadly evidence-based practice, has been strongly advocated in recent years. In the case of SET, it should first be stated that the family of life-skills programs, of which it forms a part, has a strong evidence base in the USA. For example, the Center for the Study and Prevention of Violence (Blueprints, 2009) concluded that PATHS, which is one of the major influences on SET, is “among 11 model programs certified by Blueprints, meaning that they have a high level of evidence supporting their effectiveness and should be replicated in other communities to prevent violence and drug abuse,” which is a view in line with a recent report of the Institute of Medicine in the US (Durlak, et al., 2011). SET has to be regarded in the light of a growing body of life-skills research (Diekstra & Gravestijn, 2008).

Of relevance here are the concepts of efficacy and effectiveness, which are important in field research. Although employed in somewhat different ways in the literature, they are used in the current summary to mark the difference between studies conducted in an experimental context and those performed in a real-life setting. The SET study program was explicitly an effectiveness (real-life) study, given that it involved teachers as program implementers and data-gatherers.

In a recent prevention-related article, Welsh and colleagues (2010) presented a schematized account of what they call “the implementation and evidentiary process in going to scale”. The steps involved are called efficacy, effectiveness, and dissemination. The idea is that, given some basic research, an efficacy study is performed “under optimal conditions”, followed by an effectiveness study, which is an “implementation of intervention and effect replication study in secondary sites, target populations,” and then by dissemination, alternatively called “going to scale” or “rolling out”. The points are that an efficacy study, which might be expected to meet the strictest criteria for having an evidence base, is not the end of a demonstration of the viability of an intervention method, and that an effectiveness study cannot be considered in isolation from the body of basic research findings that precedes it.

### **6.2.2 Attrition**

One of the problems in evaluating the SET program was the high rate of reported

attrition. The distinction between effectiveness and efficacy studies is important in this context, for it is almost certain that attrition will be greater in real-life than in experimental studies, as pointed out in the Discussion in Paper I. Approximately a third of junior students “drop-out” each year, as a matter of course, as they advance from junior to senior level, as do senior students as they complete their schooling. Thus, there was progressive sample attrition over the years due to normal turnover. Also, there was variable, temporary absence of students at time of testing, which in some cases resulted in more respondents in one year than the year before (Paper II).

Analyses of possible biases showed that our comparisons between the SET and No-SET groups were unlikely to have suffered from bias (papers I and II). Although we initially argued in Paper I that imputation of missing data was unsuitable, we decided to further increase the statistical power of our analyses by employing latent growth curve modeling (LGM) with full information maximum likelihood estimation (FIML).

FIML is an alternative to multiple imputation as a way of handling missing data. We could have used multiple imputation instead of FIML, but it would have been laborious for several reasons. First, our data are from a prevention study, where a treatment is given to a program group but not to a control group. So, imputations would have had to have been performed separately for the program and control conditions. Second, there are different cohorts in the validation data, which cover grades 4, 5, and 6 at t1 (see Appendix). There may be cohort differences, so imputations would have had to have been performed for each cohort. Third, we suspected a clustering effect, since the program was implemented in different classrooms by particular teachers. Accordingly, imputations would have had to have been performed for each classroom.

Thus, we had a lot of small subgroups where multiple imputations would have had to have been performed separately. FIML does not require all this, and has been shown to be a very efficient method of handling missing data (Enders, 2001). We had good reason to prefer FIML to multiple imputation. As we have seen, our results, with the exception of social skills, were largely validated by our LGM and FIML analysis.

### **6.2.3 Levels of analysis**

The early analyses (papers I and II) used individual students as observation units, which

assumes that the responses of each individual are independent of those of others. We were aware that there were possible interdependencies in the data due to respondents being in the same school class and exposed to the same teacher and other classroom-specific factors. We tested this in our evaluation of outcomes over five years by repeating our analysis first by classes and then by schools as analytic units (Paper II). We found that the comparative analyses showed larger differences between the SET and No-SET groups at classroom and school level than between the SET and No-SET students as a whole. We concluded that within-group dependencies had not exaggerated the between-groups differences.

We went on to employ different statistical methods, in particular latent-class analysis, to identify subgroups of students with different sets of responses to the questionnaires (papers III and IV). Finally, in our validation analysis, we took into account any clustering of the data by using the Type=Complex option in MPlus.

#### **6.2.4 Social skills**

The findings on social skills clearly constitute a theme that should be examined in some detail. The Social Skills Rating System (SSRS) was first presented in a manual by Gresham and Elliot (1990). Since the system's items only apply to grades 4-9, it could only be administered to the senior students. A total score can be generated, but the items can also be divided into four subscales, namely assertion, cooperation, empathy and self-control.

At the outset of our project we expected the social and emotional aspects of the SET program to run together, at least in a loose sense that would not presume any particular relation, causal or otherwise, between the two. That is, if a positive impact of SET on student well-being was found, it would be reflected in enhancements to *both* social *and* emotional skills.

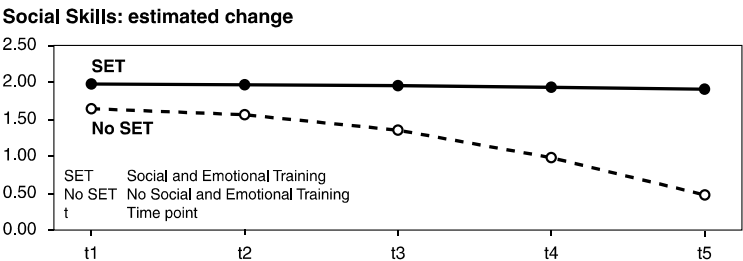
When we first examined our findings regarding senior students after two years of SET, we were surprised that differentials in favor of the SET students on some of the emotional scales (such as aspects of self-image, the hindering of aggressiveness, and attention-seeking) were not accompanied by any differentials at all on social skills, either on SSRS total or on any of its subscales. In our discussion (Paper I), we suggested the alternatives that SET is ineffective or that the SSRS, despite its proven reliability and sensitivity, did not pick up relevant changes.

We revisited the social-skills issue after five years of data were available (Paper II). We were able to report reasonable reliability for the SSRS for the first two years

of our study (Cronbach’s  $\alpha$  and test-retest). However, although we only reported on SRSS total (not the four subscales), we were again unable to show any difference between SET and No-SET students on social skills, by contrast with emotional skills, even after a duration of five years. We referred to an early suggestion that SEL programs have a greater impact on emotional than on social skills (Durlak & Wells, 1997).

In this context, we noted two aspects of the findings concerning heterogeneity (Paper IV) and the development of an instrument to measure social and emotional maturity (Paper V). First, there were separate groupings among both the SET and No-SET students who scored in opposite directions over five years. Second, there was, as expected, a stronger relation between the SSRS and the vignette-based HIF than between the SET emotional-skills ratings and the HIF (Paper V). All in all, this suggested that the SET/social-skills issue needed re-examination.

As described above, we performed a validation analysis using latent curve modeling (LCM) to address several issues: the possibility of a curvilinear relationship between SET/No SET and any one of the outcome variables, possible clustering of responses according to classroom, and adjustment for differential attrition. Although there were some nuances with regard to all the social and emotional variables, the only outcome for which the nature of the relation between SET and No-SET students changed in principle concerned the SSRS. The SET students were shown to have scored significantly better than their No-SET counterparts when a quadratic growth model was fitted to the data (see Figure 1, which is a replication of the social-skills component of Figure 4 in the Appendix).



**Figure 1.** Trend in social skills (the SSRS) over time from the LGM, with estimated latent scores on the vertical axis and repeated measurements on the horizontal axis.

The figure reveals that, starting from rather similar levels, there were minimal changes

in the scores of the SET students across the five points of measurement, whereas the scores of the No-SET students fell, and also fell at an increasing rate.

One possible interpretation of this finding lies in a theme that has appeared periodically in the literature on adolescent development over the years. The idea is that young people's mental health tends to deteriorate during the teenage years, in part because adolescent experiences tend to deflate what may be over-inflated conceptions of themselves and their capacities in all domains (academic and emotional, as well as social). Moffit (1993) considered this kind of trajectory in a study of persistent antisocial behaviour, Sampson and Laub (2003) with regard to delinquency and crime over the life course, and Özdemir (2010) in relation to adolescent perceptions of academic achievement. Indeed, the issue of what has been called calibration, which refers to the overlap between self-rating and performance, has been widely discussed since the 1970s in the context of mastery or self-efficacy (see, for example, Bandura, 1997). A recent comparative international study of adolescents in several European countries (Peetsma, Hascher, van der Veen, & Roede, 2005) consistently found a decline in sense of self-efficacy with age.

It is possible that programs like SET give young people tools of a social nature to handle the "real challenges and the need to cope with change" during "the teenage transitional period" (Rutter, 2007); with regard to alcohol, see also Brown et al. (2008). If that is the case, there is a clear argument for pursuing SEL programs in school.

#### **6.2.5 Delivery of SET: where and by whom?**

There are questions over whether the teaching of social and emotional skills should take place in schools, and over who is best at teaching them (see Durlak, et al., 2011). Are teachers and other school personnel really up to teaching these skills or should such teaching be left to outside experts? To the extent that SET has been demonstrated to be successful as an intervention, our findings indicate that teachers have indeed been successful in promoting mental health.

There is some earlier evidence that classroom teachers and other school staff are successful in promoting social and emotional skills, and also "at levels of fidelity ... nearly as high as those demonstrated by ... program specialists" (Rohrbach, Dent, Skara, Sun, & Sussman, 2007, p. 130). When SEL programs are delivered within an ordinary school setting, they seem to impact on students' academic performance (Durlak, et al., 2011; Hattie, 2009). Since the school's main objectives are to make

sure students learn and that the teaching is effective, it seems reasonable to say that social and emotional skills can play an important role in academic achievement.

Another issue is whether schools should implement a structured program or let it be up to the individual teacher to find ways to teach these skills. Accumulated research, including our own, shows that structured programs, like SET, have an effect: “In general, a school that chooses a standardized program, supervises the prevention effort, provides frequent high quality training to team members, and integrates the program into normal school operations can increase the implementation quality of the intervention, which can then increase its intended effectiveness” (Payne & Eckert, 2010, p. 139). To our knowledge, more or less systematic efforts made by individual teachers in the life-skills arena, inside or outside non-structured programs, have not been scientifically evaluated. We know, however, that efforts of various kinds, in particular with regard to bonding with the school and good peer relations, were made in the No-SET schools.

### **6.3 DEVELOPMENT OF SET**

In a recent meta-analysis of 213 school-based, universal social and emotional learning (SEL) programs, covering 270,034 students of all school ages, Durlak and colleagues (2011) found that universal programs have generally positive outcomes, in particular with regard to academic outcomes, and especially “if they use a [S]equenced step-by-step training approach, use [A]ctive forms of learning, [F]ocus sufficient time on skill development, and have [E]xplicit learning goals” (p. 408). Also, they point to how effective implementation influences outcomes and how problems with implementation can limit the benefits. Both SAFE (an acronym referring to these four outcomes) and implementation issues are important for the development of SET, and teachers and other stakeholders should be invited to give their views on the program, parts of the program, and how it is implemented.

For SET to be successful, head-teachers must not only be “on board”, but also actively support the teachers, in particular by ensuring that they receive training and supervision. School leaders are keys to the successful implementation of the program, and require knowledge of the entire process to be able to provide support of this kind.

### **6.4 FUTURE RESEARCH**

The studies aimed at finding out whether SET, implemented in ordinary schools with regular teachers, could promote mental health. While performing the study, a number

of research issues have arisen, which should be investigated further.

#### **6.4.1 What we could have done better**

In retrospect, there are several things we could have done better in the project. Most of them concern field research in a real-life context.

We have already considered aspects of implementation. Perhaps the most important concerns what might be called ownership of the project. Making this kind of project work requires involvement of the school, its leadership, and its commissioner (in the Swedish case, the municipality in charge). In a sense, we came into the project underprepared in that we did not perform a full prior analysis of the specific needs of the SET schools and how they were to integrate the SET program into their areas of educational responsibilities. Rather than having implemented our (the researchers') project, the schools would have been running a project of their own to fulfill their assignments in line with national stipulations, including the curriculum. In particular, the school leaders (the head-teachers) and their principal (the municipality) required support in the form of knowledge and ongoing training with regard to the content and implementation of prevention programs. We knew that preventive interventions worked, but we were not so well aware, during the first years, of the importance of implementation issues, as has recently been highlighted by Guldbrandsson (2008).

In terms of measurements, as mentioned previously, we could have measured academic achievement, and also considered alcohol use at earlier ages.

Administratively, the teachers proved poorer at data collection than we had hoped, possibly because their general workload was so heavy, and making sure that all questionnaires were responded to was naturally not their first priority. This implies that the number of missing data could be reduced in future studies. Here, greater involvement on the part of the research team would have been needed with regard to the delivery of questionnaires, and to following up students who failed, for various reasons, to fill them in. It should be remembered that, during the project period, resources for the digital collection of data were not available.

Further, we equipped the teachers with forms to record data on truancy, rule-breaking, reports to social services and the police, etc., but these were filled-in too seldom to be analyzable. Clearly, different routines are needed for this kind of information to be acquired, and there is a particular need for clarifying the roles of the different people involved in data collection.

#### **6.4.2 What we still can do in the SET project**

Here, given that the project has now been implemented and the data set is complete, we are concerned with possible follow-up.

Social and emotional learning programs have been shown to be related to academic success (Durlak, et al., 2011). This question could possibly be addressed by following up the SET and No-SET students with regard to how well they performed later in Swedish high school. The data are available, and record linkage would be possible subject to ethical approval.

Much school research (Hattie, 2009) shows how important the actual teacher is for a student's academic performance, and the same would be expected when it comes to SET outcomes. Since we have data on teachers' performances on various aspects of teaching SET (for the first two years of the program), it would be interesting to analyze these further.

Although the intervention period is over, the SET and No-SET students could still be followed up. For example, record linkage might be effected with various national and regional registers concerning mental ill-health, substance use, and so on, again subject to ethical approval. It might be particularly worthwhile to follow up patterns in substance use, since alcohol and drug problems are more common after the compulsory-school period.

#### **6.4.3 Suggestions for future research**

Doing research in "real life" rather than in an experimental setting has its advantages, and also its difficulties. In this context, head-teachers are the key to how well a program is set up and implemented so as to meet the needs of a particular school. Therefore, it is important for researchers continuously to meet with them to follow up what is going on in light of a prior analysis.

That the head-teachers in our schools changed time and time again showed us the importance of having support higher up in the school hierarchy, preferably from among those highest in the municipal administration. During our project time there were major organizational changes to the structure of leadership of the schools, which meant that many in top positions were suddenly no longer there.

We have learnt the importance of the structures there are in a municipality. To be able successfully to do this kind of research in Sweden, it has to be firmly rooted within the municipality concerned, down from the municipal chief executive, via the



municipal director of schools and school leaders, to individual schools' deputy heads and classroom teachers. There is considerable turnover of personnel in the school sector, but not all staff will change at the same time, which would allow a project to receive ongoing commitment. There is no way to safeguard against reorganizations, but a clear contract could be made with a municipality before embarking upon a research project of this kind

Since we know that implementation plays a crucial role, it would be interesting to guide and follow the implementation process in detail in a single school or municipality. In particular, attention should be paid to the monitoring of training and fidelity (Lee et al., 2008).

One important issue concerns the relationship between social competence and emotional competence in relation to mental health. Are they moderators, in which they act in concert, or does one mediate the other (act as an intervening variable), or might they act largely independently of each other? There are complicated aspects to this issue, which require both conceptual clarity and empirical investigation.

Qualitative studies to complement quantitative ones are required. Interviews with parents, students and teachers might clarify their views on students' social and emotional development, and also generate suggestions for program improvement. For example, parents could be asked about certain exercises and whether they notice any effects of the SET teaching in the home.

## **7 CONCLUDING REMARKS**

The primary aim of the study was to describe and evaluate, in a real-life setting, the impacts of a Swedish social and emotional learning program (SET) on various mental-health outcomes, and to draw out their implications for future interventions. Sub-aims were to differentiate between subgroups with regard to outcomes, and to develop an instrument for the measurement of social and emotional development and maturity. The outcomes of the project were generally favorable. In the context of a growing number of findings in the arena of social and emotional learning (SEL), there is evidence that SEL programs do make a contribution to the prevention of mental ill-health.

Weaknesses in the implementation of SET and also in our research approach have been highlighted. Experiences of the SET project indicate the necessity of wide community involvement, the need for greater discipline in administration, and the benefits of using a variety of study designs and statistical approaches in the interpretation of results. Life skills are essential to young people's everyday lives, and may help prevent school drop-out, and promote both contentment in school and mental health.

## 8 APPENDIX: VALIDATION ANALYSIS

Following discussion of the results presented in papers I, II, III and IV, we performed a series of analyses using different statistical techniques in an attempt to validate our findings. We used the largest and most wide-ranging of our data subsets, namely the one employed for our five-year follow-up of the effects of the SET program on the social and emotional variables. In essence, we used latent growth curve modeling (LGM) with full information maximum likelihood estimation (FIML). The Mplus software was used (Muthén & Muthén, 1998-2010). There were several advantages to utilizing these techniques:

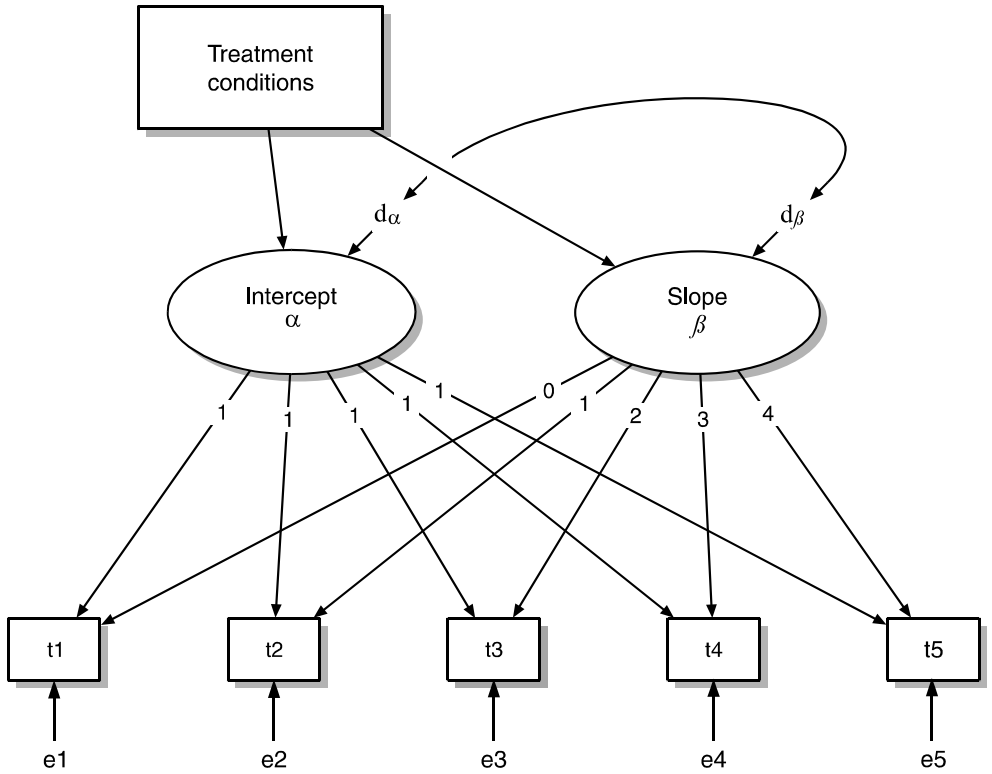
1. The data had longitudinal attrition. In our previous analyses, we did not impute missing values, on the ground that the non-random distribution of the missing data made imputation unsuitable (Paper I). In the validation analysis, we adopted the FIML approach, which “estimates model parameters and standard errors using all available raw data” (Enders, 2001, p. 715). FIML does not impute or fill in missing data values but estimates the model parameters and their standard errors based on the full data set. The computational algorithm of FIML is based on the assumption that missing values are related to observed values of other variables in the set. FIML in the MPlus software also provides adjusted standard error estimates, which is a useful safeguard against inflation of the Type-I error rate. Enders (2001) points to evidence that the FIML estimator is superior to other techniques for dealing with missing data, such as listwise deletion, pairwise deletion and mean imputation.
2. In the current data, students were clustered in classrooms. Given that the SET program was implemented by teachers within classrooms, its effects may have varied according to teachers/classrooms. In addition, and more fundamentally, students who were exposed to the same teacher/classroom environment may have shown greater similarities with each other than they would have with students in other classrooms. That is, the observations within a given classroom may not have been independent due to clustering, and in turn, the assumption of independence of observations may have been violated. Accordingly, we took into account the clustering of the data using the Type=Complex option in MPlus. This modeling feature computes robust standard-error estimates and adjusted-fit statistics to counteract clustering and non-independence in the data (Muthén & Muthén, 1998-2010). Simulation

studies have demonstrated the efficiency of this modeling approach in analyzing complex data structures (Asparouhov & Muthén, 2006).

3. We used LGM (Duncan & Duncan, 2004) to compare trends in growth in the treatment (SET) and comparison (No-SET) groups in order to estimate the program effect for each major outcome variable. The LGM approach has advantages over the ANOVA approach for the analysis of change in longitudinal data. In addition to its flexibility in comparison with ANOVA, LGM – like other latent-variable approaches – accounts for measurement error. LGM also models group-level growth rates and patterns by taking into account the initial status of individuals and variability within groups. In a specific case in a different arena, but directly relevant to the issue at hand, it has been claimed that such a procedure has the key advantage over repeated-measures ANOVA in its “ability to control for initial status and the ability to model missing data using full-information maximum likelihood” (Chamberlain, Leve, & DeGarmo, 2007, p. 189).

In the current analysis, as a first step, we fitted a single group LGM model to the data to identify the overall growth pattern for each outcome variable. We first fitted a linear growth model. If the model revealed poor fit, then we fitted a quadratic growth model. As a second step, once the overall growth pattern had been identified, we fitted a conditional growth model, in which a group variable identifying the treatment (SET = 1) and comparison (No-SET = 0) groups was included as a time-invariant covariate (see Figure 2).

A significant path coefficient from the covariate to the intercept factor would suggest that there is a significant initial difference between the groups. Similarly, a significant path coefficient from the covariate to the slope factor would suggest that the observed growth pattern is different across the control and the treatment conditions. Figure 2 depicts the fixed and estimated parameters of the conditional model.



**Figure 2.** The conceptual model for the analysis of program effects on the growth trajectories of the treatment (SET) and control (No-SET) groups, where  $d$  = disturbance or unexplained variance in the outcome variable (unmeasured error),  $t$  = time point,  $e$  = error in measurement (measurement error).

### 8.1 THE DATA SET FOR THE VALIDATION ANALYSIS

We started with the data set employed for our evaluation of the SET intervention over five years (Paper II), but we had to set inclusion criteria for the analysis. First, to allow for the possibility of a quadratic growth model, we had to exclude any student who had not filled in the same set of questionnaires on at least three occasions of measurement after the first time of measurement ( $t_1$ ). We did not have repeated measures on at least three occasions after  $t_1$  for grades 7 and upwards, while grades 1 through to 3 responded to different questionnaires during the period of the evaluation. Grades 2 and 3 were measured on three occasions, but there were no comparable measurements at  $t_1$ , meaning that data from these grades would have impacted on the slope estimates but not have been considered for the intercept estimates. In effect, this

meant that we were restricted to students in grades 4, 5 and 6 at t<sub>1</sub>, for whom we could compute both intercept and slope estimates (SET = 443 students; No SET = 101 students). See Table 2. The maximum number of occasions of measurement for any one student was five, and the minimum three.

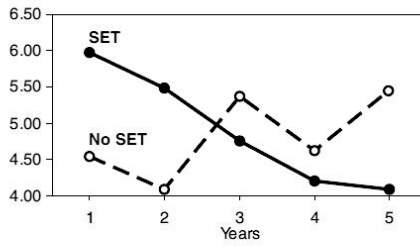
**Table 2.** Measurements according to time of questionnaire administration (t), and cohort and grade (C and G) with measurements meeting the criteria for the validation analysis marked in bold in strings of three or more along the diagonals. Cohorts are defined by their grade at baseline (t<sub>0</sub>). Measurements taken but not eligible for validation are shaded in gray.

t <sub>0</sub> (2000)	t <sub>1</sub> (2001)	t <sub>2</sub> (2002)	t <sub>3</sub> (2003)	t <sub>4</sub> (2004)	t <sub>5</sub> (2005)
<b>Cohort 1</b>	C1 @ G2				
<b>Cohort 2</b>	C2 @ G3	C1 @ G3			
<b>Cohort 3</b>	<b>C3 @ G4</b>	C2 @ G4	C1 @ G4		
<b>Cohort 4</b>	<b>C4 @ G5</b>	<b>C3 @ G5</b>	C2 @ G5	C1 @ G5	
<b>Cohort 5</b>	<b>C5 @ G6</b>	<b>C4 @ G6</b>	<b>C3 @ G6</b>	C2 @ G6	C1 @ G6
<b>Cohort 6</b>	C6 @ G7	<b>C5 @ G7</b>	<b>C4 @ G7</b>	<b>C3 @ G7</b>	C2 @ G7
<b>Cohort 7</b>	C7 @ G8	C6 @ G8	<b>C5 @ G8</b>	<b>C4 @ G8</b>	<b>C3 @ G8</b>
<b>Cohort 8</b>	C8 @ G9	C7 @ G9	C6 @ G9	<b>C5 @ G9</b>	<b>C4 @ G9</b>

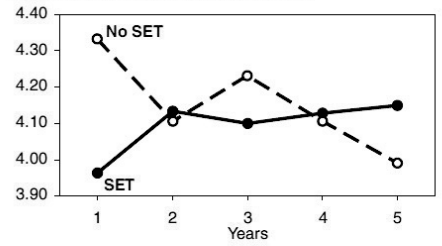
## 8.2 THE VARIABLES CONSIDERED IN THE VALIDATION ANALYSIS

We considered the following variables, which are described in detail in Paper I, and analyzed over five years in Paper II: Youth Self-Report (YSR), internalizing; YSR, externalizing; mastery; I Think I Am (ITIA), total; contentment in school, bullying; and, social skills, total. In Figure 3 we replicate the summary figure in Paper II to facilitate comparison between the results reported in Paper II and the findings of the validation analysis.

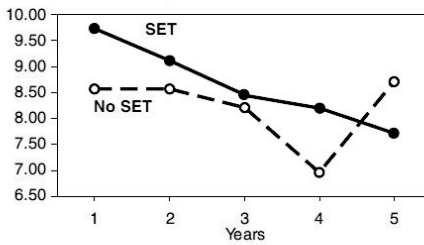
**YSR Internalizing Problems: raw scores**



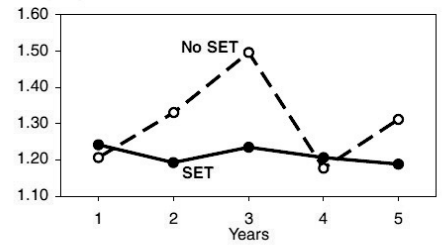
**School Contentment: raw scores**



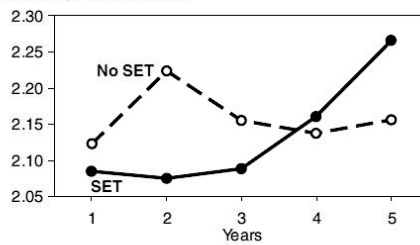
**YSR Externalizing Problems: raw scores**



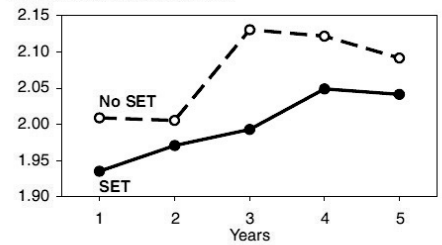
**Bullying: raw scores**



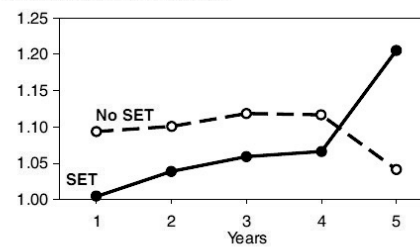
**Mastery: raw scores**



**Social Skills: raw scores**



**I Think I Am: raw scores**



#### Abbreviations

YSR Youth Self Report  
 SET Social and Emotional Training  
 No SET No Social and Emotional Training

**Figure 3.** Relations between duration of SET/No SET and the outcome variables from the repeated-measures analysis, with raw scores on the vertical axes and number of years on the horizontal axis (from Paper II).

### 8.3 RESULTS OF THE VALIDATION ANALYSIS

We present the results of the LGM analyses in Table 3. Three model-fit estimates were employed: Chi squared ( $\chi^2$ ), the Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA). Models with linear growth patterns fitted the data for four of the seven outcome variables with good model fit indices. The exceptions were externalizing problems, social skills, and bullying. A quadratic growth model fitted the data well for externalizing problems and social skills. Neither a linear nor a quadratic growth model fitted the data for bullying. Accordingly, we did not further examine change in bullying.

The results of the LGM analyses suggest that the students in the treatment condition (receipt of SET) had significantly higher internalizing problems and lower school contentment than the comparison group (No-SET) on the first occasion of measurement ( $t_1$ ). There were no other initial between-groups differences.

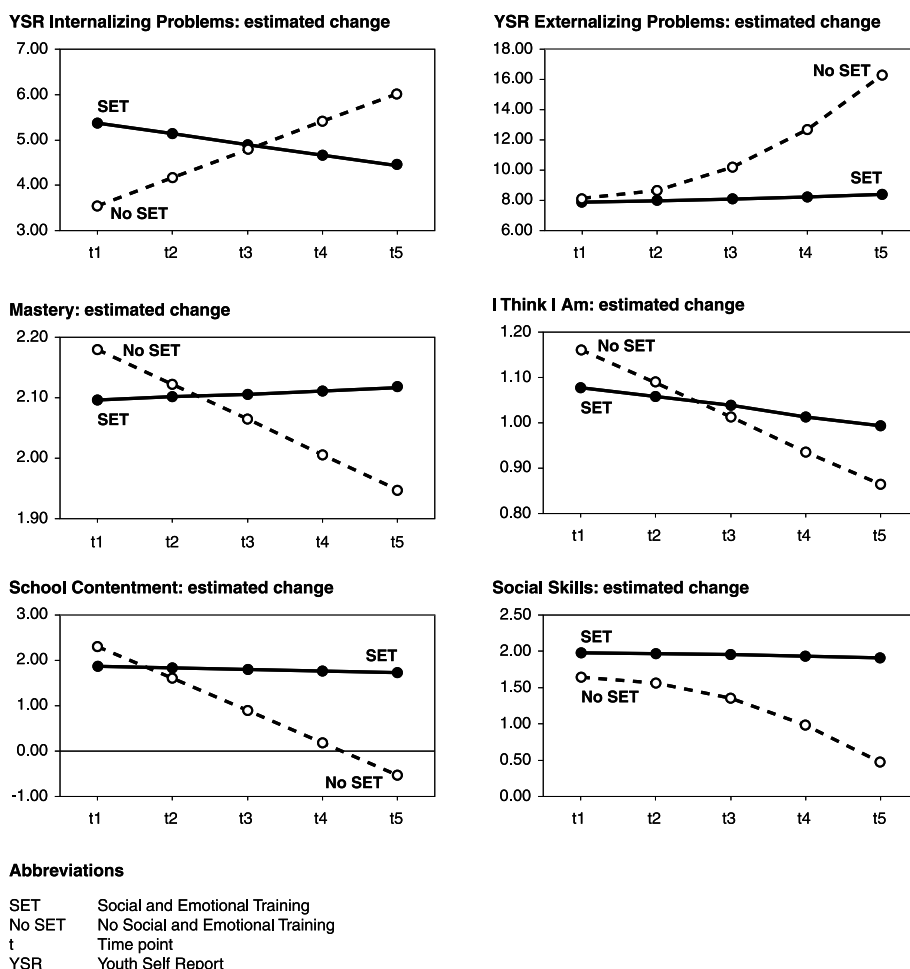
The results suggest consistent program effects on the outcome measures. In the treatment (SET) group internalizing problems decreased and externalizing problems remained stable, whereas both problems increased in the comparison (No-SET) group. In addition, externalizing problems in the No-SET group showed an accelerating increase over time. Also, feelings of mastery and contentment in school in the SET group remained stable, which can be compared with the significantly decreasing trends observed in the No-SET group. Next, we observed a significant decrease in ITIA scores in both groups, but the rate of decrease for the No-SET group was over three times greater than for the SET group. Finally, the students in the treatment group displayed no change in perceived social skills, by contrast with the quadratic decreasing trend observed for the control group students. This was the first of our analyses to suggest a favorable impact of SET on social skills. The results are summarized in Table 3 and Figure 4.



**Table 3.** Unstandardized coefficient estimates (B) and robust standard errors (SE) for random intercepts and random slopes regressed on treatment conditions (No SET = 0; SET = 1) and model-fit estimates ( $\chi^2$ , CFI and RMSEA) for the conditional models.

	Coefficient estimates		Model-fit estimates		
	B (SE)	p	$\chi^2$ (df)	CFI	RMSEA
Internalizing			18.41 (13)	0.96	0.03
Intercept	1.72 (0.67)	0.010			
Slope	-0.85 (0.28)	0.003			
Externalizing			13.96 (11)	0.99	0.02
Intercept	0.94 (0.60)	0.118			
Slope <sup>1</sup>	-0.50 (0.20)	0.015			
Mastery			17.06 (13)	0.97	0.02
Intercept	-0.08 (0.06)	0.161			
Slope	.06 (0.01)	0.009			
ITIA, total			25.46 (13)	0.92	0.05
Intercept	-0.08 (0.05)	0.120			
Slope	0.05 (0.02)	0.010			
Contentment			16.33 (13)	0.98	0.02
Intercept	-0.25 (0.13)	0.046			
Slope	0.12 (0.05)	0.016			
Bullying	Data not fitted by either a linear or a quadratic growth model				
Social Skills			28.62 (11)	0.89	0.05
Intercept	0.33 (0.18)	0.065			
Slope <sup>1</sup>	-0.37 (0.16)	0.025			

<sup>1</sup>In line with the results of the growth models, quadratic growth factors were regressed on the SET and No-SET conditions for externalizing problems and social skills.



**Figure 4.** Trends in the outcome variables over time from the latent growth curve modeling, with estimated latent scores on the vertical axes and repeated measurements on the horizontal axes.

## 8.4 SUMMARY OF THE COMPARISON

Although the detailed statistics from the repeated-measures and latent-growth analyses are not directly comparable, and differences are difficult to quantify due to adjustments to both the scoring and the intercepts and slopes, the directions of the earlier findings are largely confirmed. This applies to internalizing, mastery, the I Think I Am instrument, and contentment in school. The relation between SET and externalizing appears to be quadratic rather than linear. Further, the validation analysis suggested that the initial differences between the SET and No-SET groups were somewhat larger than we had supposed, and that there was indeed a significant effect – in the quadratic model – of SET on social skills.

## 9 ACKNOWLEDGEMENTS

I would like to express my gratitude to everyone who has supported me during the SET project and the completion of this dissertation. Special thanks to:

### **My supervisors**

Sven Bremberg, my supervisor and co-author, who has helped me a lot in my development as a researcher over the years. Thanks Sven for not giving in to me when I had some wild ideas about what I could express in academic papers.

Rolf Sandell, my co-supervisor and co-author, for supporting and believing in the project from the outset. I admire your way of being both strong-minded and humble at the same time, and thank you for all your help with the statistics.

### **People who have helped me in my research**

Special thanks to Håkan Stattin, who has acted as a mentor for me. He has been supportive all along, willing to discuss different parts of the SET project, and coming up with interesting and very useful ideas.

Metin Özdemir, who helped me greatly with the important validation analysis.

Therése Skoog, for her support and encouragement.

### **People who I have worked together with on the SET project**

Annika Magnusson, who worked alongside me during the first years of the SET project, organizing the data collection and also observing the teachers.

Ina Bergquist, who organized the data collection in later years and kept all the files in order.

Sune Lindholm and Ulrika Gigård, and all the staff at the schools where the SET project took place.

Jerrold Baldwin for doing the figures and some of the layout so nicely.

Ulla Britt Ahrens for observing the teachers.

Cecilia Bingert, José Castillo, Henrik Siljelid and Johannes Kimber for assistance with data management.

Last but not least, my husband, Jon Kimber, for all the help and support he has provided. Without him, I would have struggled a lot with the language, and without his support this dissertation might not have been completed.

## 10 REFERENCES

- Achenbach, T., & Edelbrock, C. (1987). *Manual for the youth self-report and profile*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Allebeck, P., Diderichsen, F., & Theorell, T. (1998). *Socialmedicin och psykosocialmedicin*. Lund: Studentlitteratur.
- Arthur, M., Hawkins, J., Pollard, J., Catalano, R., & Baglioni JR., A. (2002). Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors: the Communities That Care youth survey. *Evaluation Review*, 26(6), 575-601.
- Asparouhov, T., & Muthén, B. (2006). Comparison of estimation methods for complex survey data analysis
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman and Company.
- Bar-On, R. (2004). The Bar-On emotional quotient inventory: rationale, description and summary of psychometric properties. In G. Geher (Ed.), *Measuring emotional intelligence: common ground and controversy* (pp. 115-145). New York: Nova Science Publishers.
- Bartak, A., Spreeuwenberg, M., Andrea, H., Busschbach, J., Croon, M., Verheul, R., Stijnen, T. (2009). The use of propensity score methods in psychotherapy research. A practical application. *Psychotherapy and Psychosomatics*, 78, 26-34.
- Becker, J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, 41, 257-278.
- Blueprints. (2009). Blueprint model programs Retrieved 12 December 2010, from <http://www.colorado.edu/cspv/blueprints/modelprograms.html>
- Brown, S., McGue, M., Maggs, J., Schulenberg, J., Hingson, R., Schwartzwelder, S., Murphy, S. (2008). A developmental perspective on alcohol and youths 16 to 20 years of age. *Pediatrics*, 121(Suppl 4), S290-310.
- Catalano, R., Berglund, M., Ryan, J., Lonczak, H., & Hawkins, J. (2002). Positive youth development in the United States: research findings on evaluations of positive youth development programs. University of Washington. *Prevention and Treatment*, 5(ART. 15 APA).
- Chamberlain, P., Leve, L., & DeGarmo, D. (2007). Multidimensional treatment foster care for girls in the juvenile justice system: 2-year follow-up of a randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 75(1), 187-193.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*: Lawrence Erlbaum Associates.
- Coopersmith, S. (1967). *The antecedents of self-esteem*. San Francisco: WE Freedman and Co.
- Dawis, R. (2000). Scale construction and psychometric considerations. In H. Tinsley & S. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 65-94). San Diego: Academic Press.
- Diekstra, R., & Gravesteyn, C. (2008). Effectiveness of school-based social and emotional education programmes worldwide *Social and emotional education: an international analysis* (pp. 255-312). Santander, Spain: Fundacion Marcelino Botin.
- Duncan, T., & Duncan, S. (2004). An introduction to latent growth modeling. *Behavior Therapy*, 35, 33-363.
- Durlak, J. (1998). Common risk and protective factors in successful prevention

- programs. *American Journal of Orthopsychiatry*, 68(4), 512-520.
- Durlak, J., & Weissberg, R. (2005). Meta-analysis of 655 school, family and community PYD interventions. Retrieved 27 April 2011, from <http://www.casel.org>
- Durlak, J., Weissberg, R., Dymnicki, A., Taylor, R., & Schellinger, K. (2011). The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405-432.
- Durlak, J., & Wells, A. (1997). Primary prevention mental health programs for children and adolescents: A meta-analytic review. *American Journal of Community Psychology*, 25, 115-152.
- Edwards, A. (1983). *Techniques of attitude scale construction*. New York: Irvington.
- Enders, C. (2001). The performance of the full information likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement*, 61(5).
- ENSEC (2007). European Network for Social and Emotional Competence (ENSEC). Retrieved 7 March 2011, from <http://www.enseceurope.org>
- EU (2005). Final Green Paper. Improving the mental health of the population. Towards a strategy on mental health for the European Union. Brussels: Commission of the European Communities.
- Faggiano, F., Vigna-Taglianti, F., Burkhart, G., Bohrn, K., Cuomo, L., Gregori, D., Galanti, M. (2010). The effectiveness of a school-based substance abuse prevention program: 18-month follow-up of the EU-Dap cluster randomized controlled trial. *Drug and Alcohol Dependence*, 108(1-2), 56-64.
- Field, A. (2005). *Discovering statistics using SPSS (2nd edition)*. London, Thousand Oaks, New Delhi.
- Gadd, A. (2003). *Uppfattningar om värdet av att undervisa i ett socio-emotionellt träningsprogram*: Linköpings universitet, Institutionen för Beteendevetenskap, Psykologprogrammet.
- Geher, G., & Renstrom, K. (2004). Measurement issues in the emotional intelligence research. In G. Geher (Ed.), *Measuring emotional intelligence: common ground and controversy* (pp. 3-19). New York: Nova Sciences Publishers.
- Greenberg, M. (1996). *The PATHS Project: Preventive intervention for children. Final report to NIMH*. Seattle: Department of Psychology, University of Seattle.
- Greenberg, M. (2004). Current and future challenges in school-based prevention: the researcher perspective. *Prevention Science*, 5.
- Greenberg, M. (2010). The effects of a multiyear universal social-emotional learning program: the role of student and school characteristics. *Journal of Consulting and Clinical Psychology*, 78(2), 156-158.
- Greenberg, M., Domitrovich, C., & Bumbarger, B. (2001). The prevention of mental disorders in school-aged children: current state of the field. *Prevention and Treatment*, 4(Article 1).
- Gresham, S., & Elliott, S. (1990). *Social skills rating system manual*. Circle Pines: American Guidance Service.
- Gulbrandsson, K. (2008). From news to everyday use: the difficult art of implementation. Östersund: Swedish National Institute of Public Health.
- Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London, New York: Routledge.
- Hawkins, J., Catalano, R., Morrison, D., O'Donnell, J., Abbott, R., & Day, L. (1992). The Seattle Social Development project. In J. McCord & R. Tremblay (Eds.), *The prevention of antisocial behavior in children* (pp. 139-161). New York: Guilford Publications.

- Hibell, B., Anderson, B., Bjarnason, T., Kokkevi, A., Morgan, M., & Narusk, A. (1997). The 1995 ESPAD Report. Alcohol and other drug use among students in 26 European countries. Stockholm: The Swedish Council for Information on Alcohol and Other Drugs.
- Jaccard, J., Turrisi, R., & Choi, K. W. (1990). *Interaction effects in multiple regression*. Newbury Park, CA: Sage Publications.
- Lee, C.-Y., August, G., Realmuto, G., Horowitz, J., Bloomquist, M., & Klimes-Dougan, B. (2008). Fidelity at a distance: assessing implementation fidelity of the Early Risers prevention program in a going-to-scale intervention trial. *Prevention Science*, 9, 215-229.
- Lindberg, L., Larsson, N., & Bremberg, S. (1999). Ungdomars psykiska hälsa. Utvärdering av ett mätinstrument. *Rapport från samhällsmedicin*. Stockholm: Centrum för Barn- och Ungdomshälsa.
- MacCann, C., Matthews, G., Zeidner, M., & Roberts, R. (2004). The assessment of emotional intelligence: on frameworks, fissures, and the future. In G. Geher (Ed.), *Measuring emotional intelligence: common ground and controversy* (pp. 21-52). New York: Nova Science Publishers.
- Mangrulkar, L., Whitman, C., & Posner, M. (2001). Life skills approach to child and adolescent healthy human development. Washington DC: Adolescent Health and Development Unit, Division of Health Promotion and Protection, Pan American Health Organization.
- Marlowe, D. (2004). Drug court efficacy vs. effectiveness. Join Together to advance effective alcohol and drug policy, prevention, and treatment. Retrieved 20 April 2011, from <http://www.jointogether.org>
- Mayer, J., Caruso, D., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence*, 27, 267-298.
- Mazza, J., Fleming, C., Abbott, R., Haggerty, K., & Catalano, R. (2010). Identifying trajectories of adolescents' depressive phenomena: an examination of early risk factors. *Journal of Youth and Adolescence*, 39(6), 579-593.
- McKown, C., Gumbiner, L., Russo, N., & Lipton, M. (2009). Social-emotional learning skill, self-regulation, and social competence in typically developing and clinic-referred children. *Journal of Clinical Child & Adolescent Psychology*, 38(6), 858-871.
- Moffitt, T. (1993). Adolescence-limited and life-course-persistent antisocial behavior. A developmental taxonomy. *Psychological Review*, 100, 674-701.
- Moreira, P., Crusellas, L., Sá, I., Gomes, P., & Matias, C. (2010). Evaluation of a manual-based programme for the promotion of social and emotional skills in elementary school children: results from a 4-year study in Portugal. *Health Promotion International*, 25, 309-317.
- Murray, C., & Lopez, A. (1997). *The global burden of disease*. Cambridge: Harvard University Press.
- Muthén, B., & Muthén, L. (1998-2010). The MPlus Manual. Retrieved 20 April 2011, from <http://www.statmodel.com>
- Ouvinen-Birgerstam, P. (1985). *Jag tycker jag är*. Stockholm: Psykologiförlaget.
- Patel, V., Flisher, A., Nikapota, A., & Malhotra, S. (2008). Promoting child and adolescent mental health in low and middle income countries. 49(3), 313-334.
- Payne, A., & Eckert, R. (2010). The relative importance of provider, program, school, and community predictors of the implementation quality of school-based prevention programs. *Prevention Science*, 11(2), 126-141.
- Pearlin, L., Lieberman, M., Menaghan, E., & Mullan, J. (1981). The stress process. *Journal of Health and Social Behavior*, 22, 337-356.

- Peetsma, T., Hascher, T., van der Veen, I., & Roede, E. (2005). Relationship between adolescents' self-evaluations, time perspectives, motivation for school and their achievement in different countries and at different ages. *European Journal of Psychology of Education, 20*(3), 209-225.
- Peleg-Oren, N., Saint-Jean, G., Cardenas, G., Tammara, H., & Pierre, C. (2009). Drinking alcohol before age 13 and negative outcomes in late adolescence. *Alcoholism: Clinical and Experimental Research, 33*(11), 1966-1972.
- Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human Development, 15*, 1-12.
- Rohrbach, L., Dent, C., Skara, S., Sun, P., & Sussman, S. (2007). Fidelity of implementation in Project Towards No Drug Abuse (TND): a comparison of classroom teachers and program specialists. *Prevention Science, 8*, 125-132.
- Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55.
- Rutter, M. (2007). Psychopathological development across adolescence. *Journal of Youth Adolescence, 36*, 101-110.
- Sampson, R., & Laub, J. (2003). Life-course desisters? Trajectories of crime among delinquent boys followed to age 70. *Criminology, 41*, 555-593.
- Shochet, I., Dadds, M., Holland, D., Whitefield, K., Harnett, P., & Osgarby, S. (2001). The efficacy of a universal school-based program to prevent adolescent depression. *Journal of Clinical Child Psychology, 30*(3), 303-315.
- Spivack, G., & Shure, M. (1994). *Social adjustment of young children: a cognitive approach to solving real-life problems*. San Francisco: Jossey Bass.
- Stattin, H., & Kerr, M. (2009). Challenges in intervention research on adolescent development. *Journal of Adolescence, 32*(6), 1437-1442.
- UNESCO (2006). Life skills. Retrieved 12 April 2011, from [www.unicef.org/lifeskills/index.html](http://www.unicef.org/lifeskills/index.html)
- UNICEF (1989). Convention on the Rights of the Child. Retrieved 12 December 2010, from <http://www.unicef.org>
- Weissberg, R., Caplan, M., & Bennetto, L. (1988). *The Yale-New Haven social problem-solving (SPS) program for young adolescents*. New Haven, CT: Yale University.
- Welsh, B., Sullivan, C., & Olds, D. (2010). When early crime prevention goes to scale: a new look at the evidence. *Prevention Science, 11*, 115-125.
- Vermunt, J. K., & Magidson, J. (2005). *Latent GOLD 4.0 user's guide*. Belmont, MA: Statistical Innovations Inc.
- Vermunt, J. K., & Van Dijk, L. A. (2001). A non-parametric random coefficient approach: the latent class regression model. *Multilevel Modeling Newsletter, 13*, 6-13.
- WHO (1997). Life skills education for children and adolescents in school: an introduction and guidelines to facilitate the development and implementation of life skills programmes. Geneva: World Health Organization.
- WHO (1999). Partners in Life Skills Education: Conclusions from a United Nations Inter-Agency Meeting WHO/MNH/MHP/99.2. Retrieved 12 April 2010, from [www.who.int/mental\\_health/media/en/30.pdf](http://www.who.int/mental_health/media/en/30.pdf)
- WHO (2003). Skills for Health: Skills-based education including life skills. An important component of a child-friendly/health-promoting school. Geneva: World Health Organization.
- von Marées, N., & Petermann, F. (2010). Effectiveness of the "Verhaltenstraining in der Grundschule" for promoting social competence and reducing behavior problems [in German]. *Prax Kinderpsychol Kinderpsychiatr., 59*(3), 224-241.

- Vygotsky, L. (1978). *Mind in society*. Cambridge MA: Harvard University Press.
- Özdemir, M. (2010). *Adolescent self-efficacy beliefs in multiple contexts. An analysis of individual, peer, family and neighborhood factors*. Saarbrücken, Germany: VDM Verlag Dr MüllerGmbH & Co. KG.