Running head: EFFECTIVENESS OF UNIVERSAL PREVENTION

The Effectiveness of Two Universal Preventive Interventions in Reducing Children's Externalizing Behavior: A Cluster Randomized Controlled Trial

Tina Malti[1], Denis Ribeaud[2], and Manuel P. Eisner[3]

[1] University of Toronto

[2] Swiss Federal Institute of Technology Zurich (ETHZ)

[3] University of Cambridge

Malti, T., Ribeaud, D., & Eisner, M.P. (2011). The effects of two universal preventive interventionsto reduce children's externalizing behavior: A cluster-randomized controlled trial. *Journal of Clinical Child and Adolescent Psychology, 40(5),* 677-692.

This article can be retrieved using the following link: http://www.tandfonline.com/doi/abs/10.1080/15374416.2011.597084

The authors would like to express their sincere thanks to the children, parents, and teachers for participating in the study. Moreover, the authors are grateful to all the interviewers and undergraduate students for their help in data collection and coding. The authors would like to acknowledge financial support for the Zurich Project on the Social Development for Children by, amongst others, the Swiss National Science Foundation, the Jacobs Foundation, the Swiss Federal Office of Public Health, the Canton of Zurich Ministry of Education, and the Julius Baer Foundation. Correspondence concerning this article should be addressed to Tina Malti, Department of Psychology, University of Toronto, 3359 Mississauga Road North, Mississauga, ON L5L1C6, Canada. Electronic mail may be sent to tina.malti@utoronto.ca

Abstract

This paper reports the effectiveness of two universal prevention programs in reducing externalizing behavior in elementary school children. A sample of 1,675 first graders in 56 Swiss elementary schools was randomly assigned to a school-based social competence intervention, a parental training intervention, both, or control. Externalizing psychopathology and social competence ratings were provided by the children, primary caregivers, and teachers at the beginning and end of the 2-year program, with a follow-up 2 years later. Intention-to-treat analyses revealed that long-term effects on teacher- and parent-rated externalizing behavior were greater for the social competence intervention than for the control. However, for most outcomes, no statistically significant positive effects were observed.

Keywords: universal prevention, externalizing behavior, cluster randomized controlled trial, childhood

The Effectiveness of Two Universal Preventive Interventions in Reducing Children's Externalizing Behavior: A Cluster Randomized Controlled Trial

Aggression, violence, and bullying can seriously impact children's mental health and place them at risk for problems in adjustment in adolescence (Farrington, 2005). These types of externalizing behavior also interfere with children's ability to develop resilience (Masten & Wright, 2009). Based on an ecological understanding of human development (Bronfenbrenner, 1979), the present study used a combined prevention approach that focuses both on the school and family contexts. A cluster randomized controlled trial (RCT) was undertaken to evaluate the effects of two widely adopted universal preventive interventions aimed at reducing children's externalizing behavior. We evaluated the effects of these two interventions with long follow-up periods in a large longitudinal child sample and independent of program developers. The design of the study included all the components needed to optimize the quality of evidence-based violence prevention research.

Conceptually, we combined risk-focused prevention with a lifetime perspective on externalizing behavior (Eisner, Ribeaud, & Malti, in press). The key to risk-focused prevention is identification of the primary risk factors for the offending behavior and the implementation of preventive measures aimed at counteracting it. We also sought to maximize the protective factors that decrease the probability of problem behavior over a lifetime. From this lifetime perspective, context (e.g., family, school) contributes to shaping both continuity and changes in a child's symptoms and strengths (Bronfenbrenner, 1979). To be effective, prevention programs must identify the risk factors of the externalizing behavior, manipulate these risk factors in the desired direction through targeted activities, and observe whether the intended effects occur. This regimen implies that externalizing behavior can be reduced by effectively targeting family and school risk factors that influence problem behavior (Eisner et al., in press; see Metropolitan Area Child Study Research Group, 2002; Hahn et al., 2007).

We used universal prevention programs (i.e., interventions for general populations) to test how they might protect general populations of children. There are relatively few routinely delivered prevention programs because most child mental health services are targeted interventions (i.e., for children who are at risk for or who already have developed externalizing behavior problems). We chose a universal prevention strategy because there is a need to test interventions that address these issues by offering child development services aimed at an entire population (Spoth, Guyll, & Shin, 2009). A developmental risk-and-resilience perspective presupposes that all children can benefit from reducing risk and strengthening resilience factors such as social competence. Such development can lead to positive proximal outcomes that mitigate against externalizing symptoms (see Spoth et al., 2009). This universal approach also reduces the stigma of participation in prevention and intervention activities by clustering all the children within a school, regardless of their risk profile, rather than targeting a group consisting only of high-risk children (Malti & Noam, 2008).

Previous universal, school-based programs have often focused exclusively on promoting social competence. Research on two such programs, Incredible Years Teacher and Child Training, has yielded promising findings (Webster-Stratton, Reid, & Stoolmiller, 2008). Likewise, universal family interventions have proven to have positive effects on parenting and outcomes for children (for a review, see Farrington & Welsh, 2003).

Most previous prevention studies that focused on a combination of classroom- and family-based interventions were directed at children with elevated levels of aggression (e.g., Bierman, 1996). The findings from these studies have been mixed. For example, the Early Risers program uncovered few long-term changes in several externalizing symptoms (Bernat, August, Hektner, & Bloomquist, 2007). On the other hand, some universal combined prevention studies have yielded promising results that are stronger than those from studies of school-based or family-based interventions implemented in isolation. For example, the German EFFEKT

prevention program, which combines child and parental interventions, also yielded long-term improvement in children's externalizing behavior (Lösel, Beelmann, Stemmler, & Jaursch, 2006).

What is missing so far are effectiveness trials that test the single and combined effects of universal interventions in school and family contexts. Specifically, few studies have compared the relative and combined effects of universal programs designed to prevent externalizing behavior. An exception is an RCT conducted in England by Scott et al. (2009), who found that targeting multiple risk factors was an efficient way to reduce children's externalizing symptoms. Building on this study, we report results from an independent evaluation of the relative and combined effects of widely adopted, universal programs designed to prevent externalizing behavior in children. Successful independent replication of the relative and combined effects of existing prevention programs is essential for establishing effectiveness outside the controlled environment of developer-led trials, as model programs implemented with significant input from a program's developer cannot be reliably generalized to routine implementation (Eisner, 2009).

The Present Study

We implemented the PATHS (Promoting Alternative Thinking Strategies) program in the school context and the Triple-P (Positive Parenting Program) program in the family context. These programs were chosen because they are conceptually rooted in a developmental and risk-and-resiliency framework, the application of which has been shown to be effective in preventing aggression. For example, meta-analyses of school-based violence intervention programs (Wilson & Lipsey, 2007) suggest that effective programs such as PATHS tend to focus on measures such as enhancement of social competence (see Webster-Stratton et al., 2008). Meta-analyses of family-based violence intervention programs indicate that parental management training utilizing cognitive-behavioral techniques in developmental frameworks such as Triple-P are moderately effective in decreasing externalizing symptoms (Farrington & Welsh, 2003).

Our choice of programs was also based on an initial feasibility study conducted in collaboration with the school authorities of the city of Zurich, Switzerland (Eisner et al., in press). This study provided a resource and needs assessment of the target site that included the following components: (a) an evaluation of possible interventions suitable for the site, (b) an examination of site readiness that included an evaluation of the financial and organizational resources available for implementing the intervention, (c) a review of social and demographic characteristics of the site with a view to understanding the barriers and resources in the respective communities, and (d) an overview of the various stakeholders with whom partnerships needed to be established and maintained during the field experiment. This feasibility study confirmed that the chosen interventions would help meet the mental health needs of the city and thus provide a good evidential base.

Intervention in the School Context. PATHS is a research-based prevention program aimed at reducing externalizing behavior problems and enhancing social competence in primary school children (Greenberg & Kusché, 2002). It relies on an integrative model of children's risk-and-resiliency development. The underlying assumption is that the promotion of various aspects of social competence development reduces a set of well-known risk factors for aggression (Greenberg & Kusché, 2002; Greenberg et al., 2003). These risk factors include poor social-cognitive skills (Crick & Dodge, 1996), poor emotional skills (Arsenio, Gold, & Adams, 2006), and poor inhibition control (Riggs, Greenberg, Kusché, & Pentz, 2006). To reduce these risks, the PATHS lessons promote social-cognitive development, positive social behavior, and understanding of one's emotions.

Thoroughly evaluated, PATHS is 1 of only 11 programs recommended as effective by Blueprints of Violence Prevention at the University of Colorado (Greenberg & Kusché, 2002). Several rigorous trials of PATHS have been conducted. For example, the Conduct Problems Prevention Research Group (2002) screened over 9,000 kindergarten children at four sites in

three cohorts; 891 children were identified as high risk and then randomly assigned to intervention and control groups. Beginning in Grade 1, all children received PATHS. Teacher ratings of conduct problems gave modest effect-size evidence that the intervention was successful in preventing problem behavior at school. Riggs, Greenberg, Kusché and Pentz (2006) provided support for the effectiveness of the PATHS curriculum on inhibitory control in a sample of 318 second and third graders. In a recent large-scale RCT implemented in Grades 1 to 3, PATHS was found to have moderately positive overall effects on teacher-rated aggression and prosocial behavior (Bierman et al., 2010). However, in contrast to the present study, almost all past evaluations of PATHS were supervised by the developer of the program and were conducted in the US. This is important because, as noted above, trials independent of the program developers often show much less positive findings than those lacking such independence (Eisner, 2009). Very few independent trials have been conducted on PATHS.

Intervention in the Family Context. Triple-P is a multilevel parental and family training program aimed at strengthening parenting skills and reducing problem behavior in children (Sanders, 1999). It incorporates five levels of intervention, ranging from universal parenting information strategies (level 1) to a specialized, intensive intervention program for families facing multiple sources of distress (level 5). For the present study, we used the standard Triple-P (i.e., level 4). At this level, Triple-P is a group-based parent training program of four weekly 2-3 hour sessions with video elements and a parent workbook. The units address themes such as positive parenting, support for desired kinds of behavior, and avoidance of conflict escalation. To support active learning, the program makes use of video clips, group discussion, role playing, and homework for parents. After the course is completed, the program providers offer up to four weekly 20-min phone sessions during which the parents are invited to individually discuss issues arising out of their application of Triple-P principles and techniques in their family. The parents also receive a Triple-P handbook with practical advice about good parenting (Eisner & Meidert,

2011; Eisner, Nagin, Ribeaud, & Malti, 2010). In short, the interventions are intended to promote positive, effective parenting as a shield against child aggression and externalizing behavior as distal factors.

Triple-P is among the most thoroughly evaluated parental training programs in the world. A meta-analysis by Nowak and Heinrichs (2008) identified 55 studies assessing the effectiveness of Triple-P on a variety of outcome measures, including parenting and child problem behavior. Positive overall effects were reported for both categories. Prior research has also compared the enhanced, most intensive Triple-P intervention (level 5) and the standard Triple-P intervention (level 4). For example, Sanders, Markie-Dadds, Tully, and Bor (2000) conducted an RCT comparing standard and self-directed Triple-P (level 4), enhanced Triple-P (level 5), and wait-list control groups in a sample of young children (N = 305). The children in the enhanced Triple-P and standard Triple-P conditions showed significant improvement in observed disruptive behavior. Similarly, Sanders and McFarland (2000) documented that both standard and enhanced Triple-P reduced observed and parent-reported disruptive behavior in a sample of 47 children. (For an overview of RCTs on Triple-P, see Sanders, Turner, & Markie-Dadds, 2002). In contrast to the present study, most previous RCTs of standard Triple-P were supervised by the developer of the program, and they used small samples and/or samples of children with elevated levels of disruptive behavior. To date, standard Triple-P has not been evaluated systematically as a universal prevention program. Given that it was developed as a population-based preventive intervention offering a diverse set of options for families from different social backgrounds and representing different degrees of problem behavior, it is to be expected that overall population effects on children's social behavior should be found (see Eisner et al., 2010). Indeed, we chose the standard version of Triple-P for our intervention because previous studies have implemented this version as a universal parenting intervention (Prinz, Sanders, Shapiro, Whitaker, & Lutzker, 2009). Finally, our study is the first effectiveness trial of Triple-P in the sense that the program

was implemented under conditions that approximated how it is currently marketed in Switzerland.

Study Hypotheses. Our study was designed to compare PATHS, Triple-P, and PATHS + Triple-P with a control group in a cluster randomized longitudinal trial with a 2-year postintervention follow-up. We predicted that, compared to children in the control condition, children in all the treatment conditions would manifest greater short- and long-term reductions in externalizing behavior and greater increases in social competence. We assessed three different subdimensions of externalizing behavior (i.e., aggressive behavior, nonaggressive externalizing behavior, and impulsivity) because previous research indicates that PATHS has particularly strong effects on aggression prevention and impulsivity/inhibitory control (Riggs et al., 2006). Based on previous research with selected interventions (Reid, Webster-Stratton, & Hammond, 2007), we further hypothesized that these effects would be stronger in the PATHS + Triple-P condition than in the Triple-P condition, the PATHS condition, and the control condition. As PATHS is a classroom intervention that focuses on decreasing externalizing risks, we assumed that it would impact children's externalizing behavior as rated by their teachers. We also hypothesized that because of its focus on the reduction of externalization risks, the Triple-P program should have a particularly strong impact on children's externalizing behavior as rated by their parents. Based on previous studies documenting contextual effects on aggression prevention (Barkley, 2002), we hypothesized differential effects of PATHS and Triple-P at school and at home. Finally, we hypothesized that the PATHS + Triple-P intervention would affect children's externalizing behavior both in school and at home.

Method

Participants

The target population for the study was children entering the first year of elementary school in the city of Zurich, Switzerland. The data were taken from the Zurich Project on the

Social Development of Children (Z-Proso), an ongoing prospective longitudinal study of a cohort of children who entered elementary school in the city of Zurich in 2004 (for a detailed overview, see Eisner et al, in press). We used a cluster randomized sampling approach with school as the unit of randomization (Figure 1; for a detailed overview, see Eisner et al., in press). Because the targeted sample size and number of units (i.e., 50–60 schools) was small, direct randomization would have entailed a high risk of imbalance among the randomized groups. Random allocation can be expected to result in equivalence of the treatment groups only if the number of allocated units is relatively large. Therefore, a randomized block design was utilized, for which 14 blocks of 4 schools were created such that the schools within each block were similar in size and came from the same school district (i.e., the social background of the catchment area was comparable). The schools within each block were then randomly allocated to the treatment conditions (see Boruch, 1997). The randomization was computer generated.

All 90 public elementary schools in the city of Zurich were blocked by school size, and a stratified sample of 56 schools was then drawn. All the selected schools participated. The first three of four data collection waves took place at annual intervals between 2004/5 and 2006/7; wave 4 was conducted 2 years later in 2008/9 (Figure 1). The data-collection times corresponding to these waves are labeled T1 to T4 below. Each sweep collected data from the primary caregiver, the child, and the teacher.

The final sample consisted of 1,675 first graders (48% girls) from 56 elementary schools. At T1, when the children were 7 years old, 91% of the students were in regular classes, whereas the other 9% were in special-needs classes. At T1, the response rates were 81% for the child interviews (N = 1,361), 74% for the parent interviews (N = 1,240), and 81% for the teacher assessments (N = 1,350). At T2, when the children were 8 years old, the response rates were 95% for the parent interviews, 97% for the child interviews, and 96% for the teacher assessments; at T3, when the children were 9 years old, the response rates were 95% for the parent interviews,

96% for the child interviews, and 94% for the teacher assessments; at T4, when the children were 11 years old, the response rates were 86% for the parent interviews, 83% for the child interviews, and 92% for the teacher assessments.

The mean age of the children at the time of the child interviews (T1) was 7.45 years (*SD* = 0.39); 78% lived with their biological parents, 20% with their biological mother only, and 2% with their biological father only or with foster parents. As for the socioeconomic background of the primary caregiver, 25% had little or no secondary education, 30% had vocational training, 29% had attended vocational school or had a baccalaureate degree or advanced vocational diploma, and 16% had a university degree. Socioeconomic status (SES) was based on coding the caregiver's current profession; the codes were then transformed into an International Socio-Economic Index of occupational status (ISEI) score (Ganzeboom, Degraaf, Treiman, & Deleeuw, 1992). The final SES score was based on the highest ISEI score of the two caregivers. The average ISEI score of the households was 44.56 (*SD* = 17.82). A dummy variable was created and coded 0 if at least one parent was of Swiss nationality (55%) and coded 1 if both parents were of non-Swiss nationality (45%). The latter represented more than 80 countries of origin.

Interventions

The schools were randomly allocated to one of the four treatment conditions (PATHS, Triple-P, PATHS + Triple-P, control). School was chosen as the randomization unit to minimize potential crossover effects and because PATHS works best when school is the intervention unit (Greenberg & Kusché, 2002). Triple-P was implemented between waves 1 (T1) and 2 (T2), the latter being the end of the first year of elementary school. PATHS was implemented between waves 2 (T2) and 3 (T3), the latter being the second year of elementary school. In many multisystem prevention programs, school-based and family-based interventions are implemented simultaneously. Although this strategy may have advantages in terms of maximizing the dosage,

it has the disadvantage of limiting the availability of the resources needed to achieve optimal implementation quality. We therefore decided to implement Triple-P in Year 1 (2004/2005) and PATHS in Year 2 (2005/2006).

PATHS. The version used in the present study was that used in the Fast Track Project during the second school year (Bierman, 1996). This one-year program includes 46 primary lessons and several secondary ones. The content, methods, and materials were culturally adapted to the Swiss school system, and the materials were intensively tested in a pilot study (Eisner, Jünger, & Greenberg, 2006). PATHS lessons address problem-solving skills, social relationships, self-regulation, rule understanding, emotion understanding, and positive self-esteem. The PATHS classes consumed about 67 min per week during the 1-year program, an average of 2.4 sessions per week.

The teachers who implemented PATHS received a 2-day training course prior to the start of the experimental sessions. The five trainers, who are called coaches, were trained and supervised by an experienced Dutch expert who also manages the PATHS teacher education institute in the Netherlands. To increase implementation quality, the coaches were trained to visit the classes and provide feedback to the teachers. They visited each class four to six times during the implementation period, after which they discussed the lesson with the teacher. A refresher seminar was held midterm, and regular PATHS newsletters helped to create a sense of cohesion among the participating teachers. The city of Zurich had made the PATHS curriculum compulsory for teachers in the intervention group, and all classrooms in the 28 intervention schools were using it. The procedures that were used to monitor implementation closely followed suggestions by Greenberg and Kusche (2002). They included teacher and child questionnaires in addition to observations by the coach. These assessments included summaries of the content of all components of the intervention, as well as ratings by the teachers and coaches of how well the training and interventions were being implemented.

The checklists completed by the coaches indicated that, on average, 27 of the 30 obligatory lessons, 30 of the recommended vignettes, and 25 small group activities were completed in the classes. The coaches also gave high ratings to the implementation quality of the 308 PATHS classes that they observed during their classroom visits. The quality of classroom leadership, child motivation, and teaching of PATHS concepts received marks of 88%, 82%, 74% respectively. Overall, the teachers liked the training: 85% rated it good or very good and 88% evaluated the curriculum positively. Sixty-one percent of them rated the coaching as supportive. The child questionnaires were distributed 2 to 4 months into the program. Of the children enrolled at the time, 86% reported that they liked PATHS much or very much, and 85% reported that they were familiar with the key concepts in PATHS (i.e., emotion cards, child of the day; supplemental materials for descriptive statistics across sites are available upon request).

Triple-P. The standard Triple-P (i.e., level 4) as used in the present study is a parental training course consisting of four units, each lasting 2–2.5 hr and presented in a group format. The routines address issues such as positive parenting and how to avoid the escalation of conflicts. Overall, 1,235 parents (74% of the target sample) agreed to participate in the study at wave 1. The target sample (treatment conditions) consisted of 819 families from the 28 schools selected for Triple-P, leaving 856 for the control condition. For 257 of the children in the experimental conditions, at least one parent enrolled for the program (31%), and in 76 cases (9%) both parents signed up. Of the 257 parents who did enroll, 220 of them (86%; 27% of the target population) attended at least one session. The mean number of sessions attended by parents was 3.07 (SD = 1.42). Parents of 153 children (19% of the target sample) completed all four course units. Overall, 70% of those paents present at the first session fully completed the program. This corresponds to 18.6% of the initial target sample.

Parents who enrolled in the program were more likely than nonenrollees to report oppositional defiant behavior, aggression, a conflictual family climate, and high SES (Eisner et

al., in press). We also asked the parents for their motives to enroll in the course. Interestingly, most parents reported that they came because of their interest in parenting programs (71%). In contrast, 18% of the parents reported that they enrolled because of the presence of at least one child problem behavior or parent education problem. An examination of parental engagement revealed that the parents who completed the program differed from the target population (Eisner et al., 2010). For example, they were more likely to come from breadwinner families, to be Swiss, and to have high SES (for a more detailed description, see Eisner & Meidert, 2011). Additional steps were undertaken to motivate non-German speaking families with an immigrant background to participate. First, the Triple-P information package was translated into the ten languages most widely spoken by the immigrant minorities in Zurich. Furthermore, Triple-P International agreed to translate the complete program into Albanian, Portuguese, and Turkish. In Zurich, these three languages are spoken by significant immigrant minorities who, on average, are at a considerable social disadvantage. showing low levels of education and/or employment. Triple-P could thus be offered in five different languages in conjunction with the original English version. Also, bilingual Triple-P providers contacted all the Turkish-, Albanian-, and Portuguese-speaking parents in the target sample individually to explain the goals of the program to the parents and motivate them to participate.

Experienced Triple-P providers with a background in psychology, education, or guidance counseling who were licensed by Triple-P Switzerland were selected from a pool of applicants. All the German-speaking providers had significant experience in presenting the courses. New providers for the Albanian, Turkish, and Portuguese programs were recruited by the implementation team and trained by Triple-P Switzerland. In addition to obtaining a Triple-P license, the trainers attended two training courses to prepare themselves for the project. In these courses, they had the opportunity to discuss difficult situations with an expert of the family education center of the city of Zurich or an experienced school psychologist. When necessary,

bilingual Triple-P providers offered translations to parents with immigrant backgrounds. To assure uniform implementation quality across groups, the providers were invited to two meetings to discuss key situations and receive coaching from highly experienced providers, and a parent-teacher conference was held to inform and motivate the parents to participate. An implementation team composed of local school authorities managed recruitment and the organization of the Triple-P courses.

The target group included all parents of first grade children in the 28 schools allocated to the Triple-P condition. After the start of the school year, the schools sent the parents information about the project and the parenting program, as well as an enrollment form. The parents were informed that the school authorities supported the program and encouraged participation. Also, as a complement to the mailed information package, the Triple-P providers introduced the program during the first parent-teacher meetings of Grade 1. The courses were offered in every school district, and travel distances were generally less than 1 mile. To reduce barriers created by difficult work schedules, the program was offered in the mornings, afternoons, and evenings, and the parents could choose which weekday they preferred. To ensure close proximity to where the parents lived, the courses were held in nearby schools, community centers, church centers, and youth centers. They began in May 2005, about 6 months after the median date of the baseline parent interviews. They were completed in early July 2005, about 2 months before the start of the postassessment. The course providers estimated that 93% of the intended course material was presented. The costs of the course were fully covered by the city of Zurich, and parents even received funding for a babysitter when necessary.

Overall, 41 Triple-P courses were conducted by the providers associated with Triple-Switzerland: 33 in German, 3 in Turkish, 2 each in Portuguese and Albanian, and 1 in English. The number of participants per course across the four sessions varied between 5 and 12 (at the first session, $M = 6.73$, $SD = 2.44$). Parental satisfaction with the program and the presentation of

the course materials were assessed by questionnaires and provider checklists. Data collected from the Triple-P trainers revealed that treatment fidelity was high. The Triple- P trainers evaluated how many aspects of the programs were regularly conveyed to participants on a scale from 0–100%. For the German-speaking courses, this rate was 93%; for the foreign-language courses, it was 90%. Overall participant satisfaction with the program was high (M = 4.33, SD = 0.89), as were the ratings of provider competence (M = 4.65, SD = 0.73), both on 5-point scales.

Measures

Externalizing behavior. The teachers, parents, and children evaluated the externalizing behavior of the children at all four measurement times (T1–T4) using Tremblay et al.'s (1991) Social Behavior Questionnaire (SBQ). This instrument has been used in a variety of longitudinal studies, and it has been shown to be sensitive to behavior changes in multiple intervention studies (e.g., Lacourse et al., 2002; Lösel et al., 2006). An expert-team approach was adopted for the translations (for a detailed description, see Eisner & Parmar, 2007, p. 14). Although the SBQ has not been previously used in Switzerland, it has been translated into German and the German version has been used in Germany (Lösel et al., 2006). The German and English versions were used as the basis for the expert translations.

Items on the teacher and parent interview were assessed on a 5-point Likert scale. The children were shown drawings of specific behaviors of a child and asked whether (s)he sometimes does what is shown in the pictures. A yes/no format was utilized so that the items could be easily understood by children at age 7. The assessments were based on the Dominic interactive measure, which has been shown to have moderate to excellent reliability and validity for young children (Linares Scott, Short, Singer, Russ, & Minnes, 2006). Three main subdimensions of externalizing behavior were measured: Aggressive Behavior (AB; e.g., "is cruel, bullies, or is mean to others"; 11-12 items), Impulsivity/ADHD (ADHD; e.g., "The child is impulsive, acts without thinking"; 8-9 items), and Non-aggressive Conduct Disorders (NACD;

e.g., "The child tells lies and cheats"; 4-5 items). The full version of the SBQ was administered to the teachers, parents, and children at T1 and T3. Because of the time limits imposed on the interviews, a shortened version of the SBQ, which excludes the ADHD subscale, was administered at T2 and T4 (see Table 1).

For the AB subscale, the reliabilities (Cronbach's α) across the four waves were .93 for all the teachers' reports; they ranged from .79 to .81 (M = .80) for parents' reports and from .72 to .76 (M = .73) for children's self-reports. For the ADHD subscale, α ranged from .90 to .91 (M = .91) for teachers' reports, from .71 to .79 (M = .75) for parents' reports, and from .62 to .78 (M = .71) for children's self-reports. For the NACD subscale, the reliabilities ranged from .69 to .78 (M = .74) for teachers' reports; the reliabilities for parent-reported and child-reported NACD were too low (M = .57; M = .33, respectively) to justify including these scales in the analyses, so they were dropped.

Social competence. The children's social competence was operationally defined as (a) prosocial behavior and (b) social-cognitive skills. Teachers, parents, and children evaluated the prosocial behavior of the children at all four times using the Prosocial Behavior subscale of the SBQ (Tremblay et al., 1991). This subscale contains seven to 10 items, depending on the group filling it out. A sample item is: "The child will try to help someone who has been hurt." Across the four waves, α ranged from .91 to .93 (M = .92) for teachers' ratings, from .77 to .83 (M = .80) for parents' ratings, and from .59 to .79 (M = .66) for children's self-ratings.

The children's social-cognitive skills were measured at T1 and T3 by having them respond to four hypothetical vignettes: playing on a swing, participating in a game, laughing at someone, and stealing a ball. The four scenarios, which were adapted from previous research (Crick & Dodge, 1996), consist of a three-frame sequence of gender-matched cartoons. For the first story, the child is read the following text:

Pretend that this is you and that this is another child. The other child has been on the swing

for a long time and doesn't seem to want to share the swing with you. You would really like to play on the swing.

Afterwards, the child is asked the following question: "What could you say or do so that you could play on the swing?" This question measures children's social problem-solving strategies. The responses were audiotaped and later coded in the following categories: (a) aggressive strategies (e.g., "I'd just push him off the swing"), (b) socially competent strategies (e.g., "I'll ask to take turns"), and (c) other strategies (authority orientation, irrelevant). Category (c) is not considered further in this paper, because we were interested only in aggressive and problem-solving strategies. Two independent coders rated the total content of all the transcripts. The interrater agreement (Krippendorff's α) across the categories was .80 at T1 and .87 at T3. The raters discussed their disagreements with each other until a consensus was reached. Proportional mean scores for aggressive and competent problem-solving strategies were then created.

Procedure

The parents were asked to sign an informed consent form for their child's participation at the beginning of the first interview and 81% gave consent. A consent form was given again at wave 4. Parents who did not give consent tended to be overrepresented in four of the ethnic minority groups (i.e., Albanian, Portuguese, Serbo-Croatian, Turkish). The study procedures were approved by the data security commissioner of the city of Zurich. The Z-Proso study is registered with the International Standard Randomized Controlled Trial board (http://www.controlled-trials.com/ISRCTN84472990). To optimize the participation rates for the high proportion of parents with an immigrant background (57%), all contact letters and parent interviews were translated into the eight languages spoken by the most prominent immigrant minorities in Zurich (i.e., Albanian, English, Italian, Portuguese, Serbian/Bosnian/Croatian, Spanish, Tamil, and Turkish). Special care was taken to recruit native speakers or cross-culturally competent interviewers for these larger immigrant communities. The computer-

assisted face-to-face interviews of the parents lasted an average of 1 hr and were conducted at the parent's home. The interview partner was usually the mother, because this was the person most involved in the child's upbringing. The parental interview contained questions on parenting, family and school background, and the child's development. In the first three waves, computer-assisted personal child assessments lasting 45 min were conducted at the school. In the fourth wave, classroom-based paper-and-pencil surveys lasting approximately 90 min were utilized. The child's teacher completed a questionnaire on the child's social development and returned it by mail. The questionnaire also included items on the child's behavior and social development. The interviews were conducted by 44 interviewers intensively trained by the research team, especially in techniques for interviewing children.

Data Analysis

Hierarchical linear modeling (HLM Version 6.08) was used to assess the effects of the PATHS and Triple-P programs on child externalizing behavior and social competence over time. We recoded treatment assignment as two dummy variables to compare the PATHS and Triple-P conditions separately with the control condition. Thus, a standard approach to coding a 2 x 2 design (2 levels of Factor A crossed with 2 levels of Factor B) was used to analyze program effects. This 2 x 2 design allowed us to specify the different timings of the interventions as well as the inclusion of interactions involving the PATHS + Triple-P condition. The cross-product of the PATH + Triple-P interaction answers the question of whether adding PATHS improves the effects of Triple-P and whether adding Triple-P improves the effects of PATHS. The models incorporated three levels: data-collection wave (level 1), child (level 2), and school (level 3). These levels were employed in conjunction with a two-way interaction between time and intervention to measure the treatment effects. This approach enabled testing of whether schools in each of the treatment conditions varied in the degree to which the children's behavior changed from wave 1 to wave 4, while controlling for the nested structure of the data. Although we

considered parenting group as an additional level in our HLM models, the structure of our data did not allow for it because parenting groups were not necessarily held in schools but nearby community centers. In addition, the average group size was small, which would have resulted in insufficient power. The Zurich school system requires that children remain in the same class with the same teacher from Grade 1 to Grade 3, but they enter new classes in Grade 4 (i.e., middle school). The new teachers at T4 were blinded to the treatment conditions.

For the final analyses, multiple imputation was used to account for missing data as follows. For the teachers, only 6% of the datapoints were missing and they were randomly distributed in the database. Because Little's MCAR test was not significant, $\chi^2(135) = 152.75$, multiple imputation was not necessary for the teachers' data. For the children, 15% of the datapoints were missing, but these were not distributed randomly in the database. In this case, the MCAR test was significant, $\chi^2(168) = 212.32$, $p < .05$. For the parents, 21% of the datapoints were missing and they were not distributed randomly. The MCAR test was significant, $\chi^2(54) = 93.88$, $p < .01$. Therefore, multiple imputation was carried out for the parent and child data using the expectation maximization method to estimate the values for the missing datapoints.

Results

Initial Group Equivalence and Attrition

Table 1 presents descriptive statistics for all outcome variables by rater and condition across time. ANOVAs were conducted to determine the equivalence of the treatment and control groups across the outcome variables. The models took account of the nesting of students within schools, treating schools as a random variable. As can be seen from the descriptive statistics in Table 1, some measures of the children's behavioral problems were higher in the intervention than in the control conditions at baseline. However, our preliminary analyses revealed no statistically significant baseline differences on any of the teacher, parent, or child outcome measures across treatment conditions. Additional protection against chance bias was provided by

our stratified randomization procedure (Roberts & Torgerson, 1999). Because of this approach, the similarity of the intervention and control conditions at pretest was found to be satisfactory for the purpose of assessing outcomes. Attrition was low across the four waves of data collection and was comparable in the four treatment conditions. Specifically, children in the control condition completed an average of 3.68 waves, children in the PATHS condition an average of 3.76 waves, children in the Triple-P condition an average of 3.64 waves, and children in the PATHS + Triple-P condition an average of 3.72 waves (Figure 1).

Analyses of Program Impact on Outcome Variables

Given that longitudinal and intervention research has consistently shown the importance of child- and family-level factors in predicting the outcome variables of interest (e.g., Bierman et al., 2010; Raver et al., 2009), and that our preliminary analyses confirmed significant effects of the main independent variables on outcomes, we controlled for gender, special- education classes, nationality, and household SES (including its interaction with the intervention conditions) in all the multilevel analyses (see Raver et al., 2009; Roberts & Torgerson, 1999). As the interaction terms were not significant in any of the models, they are not displayed in the relevant parts of Tables 2 and 3. Furthermore, because previous studies indicate that program outcomes are affected by the initial level of problem behavior (e.g., Bierman et al., 2010), we tested the moderating role of initial levels of externalizing behavior. These moderator effects were tested by three-way interactions between intervention, initial level of behavior, and data-collection time. We also used preliminary models to test the immediate posttest effects of the interventions on externalizing behavior and social competence at T3. As none of these models were significant, we report only long-term effects in the following.

Preliminary, unconditional three-level models were first run to ascertain the proportion of the variance of each dependent variable that could be attributed to child and school level. Following the recommendation of Kim (2009), we computed intraclass correlation coefficients

(ICCs) for this purpose. ICCs in three-level models help to identify how much variance is explained at the school level relative to the total variance. The average ICC for school level was .25 across all teacher-reported outcome variables, .02 across all parent-rated outcome variables, and .04 across all child-rated outcome variables. Next, multilevel models were run to obtain the "intent-to-treat" estimates for the effects of the interventions on the children's externalizing behavior and social competence. These models used the SBQ posttest scores to assess these effects at both the child and school levels. The treatment effects on children's externalizing behavior are presented in Table 2. We also computed effect sizes for the significant Treatment x Time interactions by multiplying the β estimate of the interaction term by the number of time points and dividing by $\sqrt{((\text{var}(y\_1)+\text{var}(y\_n) - 2*\text{cov}(y\_1,y\_n))}$, where var($y\_n$) is the estimated variance at T1 and T4, and cov(y_1,y_n) is the estimated covariance between T1 and T4 (cf. Fonagy et al., 2009).

The results for the teacher ratings of externalizing symptoms suggest that, as expected, children in the PATHS condition showed a greater overall decline in the externalizing symptoms of aggressive behavior and Impulsivity/ADHD than their control group counterparts on the SBQ (both $p < .05$). The effect sizes were moderate: $d = 0.42$ for Aggressive Behavior and $d = 0.46$ for Impulsivity/ADHD scores. The effect of the PATHS intervention on impulsivity/ADHD was moderated by baseline impulsivity/ADHD, with the effect of the intervention being nonsignificant for children with low baseline scores on Impulsivity/ADHD. In other words, only children with high impulsivity/ADHD at baseline benefited from the PATHS intervention, as indicated by a reduction in teacher-rated ADHD at followup. In addition, increased teacher-rated aggressive behavior, nonaggressive externalizing behavior, and impulsivity/ADHD were predicted by high baseline problem behavior and the child being male. A non-Swiss background predicted increased teacher-rated aggression.

As for externalizing symptoms rated by the parents, children in the PATHS group showed a

greater decline on SBQ aggressive behavior than their control group counterparts ($p < .05$), but the effect size was small: $d = .26$. For the PATHS + Triple-P condition, the effect of the intervention on aggression was moderated by baseline aggression, with the effect of the intervention at follow-up being nonsignificant for children with low aggression scores. The corresponding relationships with impulsivity/ADHD as the dependent variable were nonsignificant. Increased parent-reported aggressive behavior and impulsivity/ADHD were predicted by baseline problem behavior, the child being male, the child being enrolled in a special-education class, and a Swiss background.

There were no statistically significant treatment effects for child-reported aggressive behavior. However, in the combined PATHS + Triple-P condition, the effect of the intervention on aggression was moderated by baseline aggression, with the intervention effect at follow-up being nonsignificant for children with high initial aggression scores. Additionally, child-reported aggressive behavior was predicted by the child being male and by having a Swiss background (see Table 2). There were also no statistically significant treatment effects for child-reported impulsivity/ADHD. However, in the Triple-P condition the effect of the intervention on impulsivity/ADHD was moderated by baseline impulsivity/ADHD, with the effect of the intervention at followup being nonsignificant for children with low initial Impulsivity/ADHD scores. Furthermore, increased child-reported impulsivity/ADHD was predicted by baseline impulsivity/ADHD, by the child being male, and the child being enrolled in a special-education class.

Next, we evaluated the effects of the interventions on the three measures of social competence: prosocial behavior, aggressive problem solving, and socially competent problem solving. Table 1 displays the means of these variables by rater and treatment condition across data-collection times. Table 3 presents the multilevel findings for social competence. Contrary to our hypotheses, there is no evidence of a clear benefit from program participation on prosocial

behavior. The findings for the child- and family-level control variables showed that increased prosocial behavior as reported by teachers, parents, and children was predicted by the child being female and baseline prosocial behavior. Additionally, increased teacher-reported prosocial behavior was predicted by a Swiss background and low SES. Increased child-reported prosocial behavior was also predicted by not being enrolled in a special-education class and by high SES. Parent-reported increased prosocial behavior was predicted by a non-Swiss background (Table 3).

The mean proportion of aggressive problem-solving strategies showed the typical developmental decrease from Grade 1 to Grade 3, as indicated by the significant time effect in the multilevel models. However, the children in the PATHS and PATHS + Triple-P conditions did not differ significantly from the control children in the mean proportion of aggressive problem-solving strategies they generated. Likewise, the mean proportion of socially competent problem-solving strategies increased from the first to the third grade, but again there were no significant differences across intervention conditions. The results for problem solving revealed that increased aggressive problem-solving skills were predicted by baseline aggressive problem-solving skills, by the child being male, and by the child being enrolled in a special-education class. On the other hand, increased socially competent problem-solving skills were predicted by baseline socially competent problem-solving skills, by the child being female, and by the child not being enrolled in a special-education class (Table 3).

Discussion

The purpose of this study was to evaluate the effects of two interventions on child externalizing symptomatology and social competence in two different ecological contexts (school and family). A large-scale RCT of a cohort of children attending public schools in Zurich, Switzerland was conducted with high-quality implementation and long follow-up periods. Because our study was an independent replication of previous studies lacking such

independence, the potential bias inherent in studies where the evaluation is conducted by the program developers was reduced.

As expected, according to teacher reports, the PATHS intervention was more effective than no intervention in reducing the children's long-term impulsivity/ADHD and aggressive behavior. In the 5th year, or maintenance phase, PATHS remained superior to no intervention in terms of teacher-reported externalizing behavior. In addition, according to parent reports, PATHS helped reduce aggressive behavior. The effect sizes were moderate for the primary long-term outcome variables. The teachers who rated students at T4 were not the same teachers who conducted the intervention and were blinded to condition assignment, a factor which adds to the generality of the findings. These findings are consistent with the view that universal school-based social competence interventions such as PATHS can reduce the long-term likelihood that children will engage in externalizing problem behavior as rated by teachers and parents. This is consistent with results from previous studies that tested samples in the United States (e.g., Riggs et al., 2006). In contrast, no differences between intervention and no intervention were found in the children's ratings of their own externalizing behavior. Even though all our measures were reasonably reliable, the teacher reports were not highly correlated with the parent and child reports of similar behavior. This finding underscores the importance of collecting ratings from multiple sources when assessing the effects of an intervention. Our study is not the first to show that parents, teachers, and children can provide inconsistent data regarding problem behavior in children (Kraemer et al., 2003). This inconsistency could also help explain why positive results have not fully generalized across settings.

In contrast to our hypotheses, the Triple-P intervention had no significant effect on children's overt externalizing behavior, and the PATHS + TRIPLE-P treatment did not have any stronger effects on externalizing behavior than PATHS alone. Our results thus provide no evidence that Triple-P by itself decreases externalizing problem behavior. These findings are

inconsistent with previous RCTs on Triple-P (Sanders et al., 2002). However, they are in line with the previous evidence base, which has provided equivocal data on the effectiveness of community-based preventative approaches to parental training. Durlak and Wells (1997), for example, conducted a meta-analysis of the effects of different types of primary prevention programs on behavioral problems in children and adolescents. For the 10 studies that used parental training as the primary intervention, the effect size was nonsignificant with $d = 0.16$. Findings from more recent studies on community-based parenting programs have yielded mixed results. Whereas Gross et al. (2009) reported positive effects on child problem behavior, others such as Hiscock et al. (2008) found no significant effects of competently implemented, parent-focused universal interventions on children's externalizing behavior. This inconsistency can likely be attributed to differences in methodology, samples, and program implementation. These differences might also explain the weaker than anticipated effects found in other recent evaluations of Triple-P. For example, studies that found significant program effects generally tested smaller, more selective samples (Nowak & Heinrichs, 2008). As Triple-P offers increasingly intensive support to families with the greatest need (level 5), the use of the standard (level 4) intervention might have limited the impact of Triple-P on some participants in our study. Previous studies, on the other hand, have not revealed strong differences between levels 4 and 5 on behavioral outcomes (Sanders et al., 2002). As our study was among the first to use standard Triple-P as a universal prevention strategy in a large-scale field trial, new RCTs on the effectiveness of Triple-P are necessary.

Despite the fact that some significant main effects for the PATHS intervention on externalizing behavior were found at follow-up, it should be emphasized that many of our analyses revealed no significant treatment effects at all. This might to some extent be related to the fact that small-sample studies tend to produce large effect sizes, whereas large-sample field trials often produce effect sizes close to zero (Eisner, 2009). In addition, large field trials

conducted by independent investigators have tended to yield less positive results than developer-led studies with tight control over all aspects of the study (Petrosino & Soydan, 2005).

Additionally, the findings indicate that the baseline measures of externalizing psychopathology predicted the corresponding externalizing outcomes. We also found evidence for the moderating role of baseline behavior on these outcomes. Three of the four significant effects were in the direction that children with high levels of baseline problem behavior benefited more from either or both interventions than children with low levels of baseline problem behavior. This finding is in line with previous RCTs (e.g., Bierman et al., 2010; Raver et al., 2009) and may indicate that children at risk for behavioral problems benefit most from these types of interventions.

Contrary to our expectations, there was no significant effect on child externalizing psychopathology after termination of treatment; this effect did not appear until follow-up. This delayed treatment effect might be related to the fact that all the teacher assessments at T4 were made by new teachers who were blinded as to treatment condition and who had PATHS-treated and untreated children in their classrooms. Perhaps it was easier for these T4 teachers to detect change than for the T3 teachers, whose classes had only treated children. Thus, the effects of the treatments on externalizing behavior depended to some extent on the composition of the classes.

Our second major finding is that the interventions did not clearly increase any of the tested dimensions of social competence (i.e., prosocial behavior and sociocognitive skills). This finding is surprising, as previous large-scale RCTs have shown positive effects of social competence programs on these variables (Biermann et al., 2010). Perhaps some teachers implement the program components aimed at increasing social competence more competently than others. Hence, analyzing the effects of differences in program implementation and the impact of these differences on social competence seems warranted as a next step. In our study, some of the teachers might have rated the children's behavior more realistically (i.e., came to less positive

conclusions) than teachers in the control group because their participation in PATHS led them to reflect on the children's prosocial behavior in everyday situations. In addition, the data show that teacher ratings of prosocial behavior changed in the expected direction from T1 to T3, whereas the change of ratings from T3 to T4 was in the opposite direction in all conditions (although the change was most pronounced in the PATHS condition). Perhaps children in the PATHS condition, who remained with the same teacher for 3 years, became closer to their teacher as a result of the intervention. When they changed to a new teacher at wave 4, their experience may have become more negative. Future longitudinal data from our study will help test this speculation.

Several limitations of the study should be noted. First, as with any longitudinal school-based study, there were missing data. In our case, however, the attrition was low and comparable across conditions. Second, we did not investigate proximal indicators of child externalizing and social competence, such as parenting. On the other hand, we did control for the child and family variables known to influence externalizing behavior and social competence. Future analyses with moderating variables will indicate if particular subgroups of children benefit from PATHS and/or Triple-P. Third, although we adapted the curricula to the cultural needs of our different subgroups, there may be social dynamics unique to our diverse sample that make it more difficult to implement a family-oriented curriculum such as Triple-P successfully. Fourth, our results may not generalize to schools where children do not have the same teacher and peers across the first school grades, since teachers and students in Swiss schools are likely to have continued to use the PATHS skills. Future research is needed to compare how different grouping structures in schools may relate to treatment effects.

Implications for Research, Policy, and Practice

Despite these limitations, our findings indicate that universal, school-based interventions such as PATHS can potentially ameliorate the long-term impact of externalizing

psychopathology. At the same time, many of our analyses revealed no significant treatment effects at all. However, it has been shown that large-sample, independent field trials generally produce weaker effects than small-sample, developer-led trials (Eisner, 2009). Inconsistent with the few existing previous trials, our findings also suggest that combining universal school- and family-based interventions has no additional effect. Furthermore, the family-based intervention did not have any treatment effects, and we have shown elsewhere that there were also no effects of the parenting program for highly adherent parents (Eisner et al., 2010). The discrepancies in the findings may result from differences in the methodological rigor of the studies, and additional research is needed to further disentangle these inconsistencies.

Among the broader implications of our findings is that results from studies in which program developers play a role in the program's implementation cannot always be generalized to other contexts. Independent replications provide useful information about the extent to which programs are effective under real-world conditions similar to those in routine dissemination. We therefore concur with others (St. Pierre et al., 2006) that a larger number of high-quality independent field trials is an essential step towards creating a better evidence base to guide interventions aimed at preventing externalizing problem behavior in children. As a next step in learning how theory-based practices can be effectively applied at the population level, researchers should include in their agenda an independent evaluation and replication of the efficacy of different dissemination strategies (Gottfredson et al., 2006), as well as greater integration between program theory and the independent evaluation of RCTs (Deaton, 2010). Independent replication is needed to minimize the attributability of findings to self-selection and expectancy effects at the level of the participating aggregate units and to increase their generalizability (Malti, 2011).

References

Arsenio, W., Gold, J., & Adams, E. (2006). Children's conceptions and displays of moral emotions. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 581-610). Mahwah: Lawrence Erlbaum.

Barkley R.A. (2002). Psychosocial treatments for attention-deficit/hyperactivity disorder in children. *Journal of Clinical Psychiatry,* 63 (suppl 12), 36-43.

Bernat, D.H., August, G.J., Hektner, J.M., & Bloomquist, M.L. (2007). The Early Risers preventive intervention: Testing for six-year outcomes and meditational processes. *Journal of Abnormal Child Psychology, 35,* 605-617.

Bierman, K. (1996). Integrating social-skills training interventions with parent training and family-focused support to prevent conduct disorder in high-risk poluations. The Fast Track Multisite Demonstration Project. The Conduct Problems Prevention Research Group. *Annals of the New York Academy of Sciences, 794*, 256-264.

Bierman, K.L., Coie, J.D., Dodge, K.A., Greenberg, M.T., Lochman, J.E., McMahon, R.J., Pinderhughes, E., & Conduct Problems Prevention Research Group (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), 156-168.

Boruch, R. (1997). *Randomized experiments for planning and evaluation: A practical guide (applied social research method series, No 44)*. Thousand Oaks: Sage.

Bronfenbrenner, U. (1979). *The ecology of human development.* Cambridge, MA: Harvard University Press.

Conduct Problems Prevention Research Group (2002). Evaluation of the first three years of the Fast Track Prevention Trial with children at high risk for adolescent conduct problems. *Journal of Abnormal Child Psychology, 30,* 19–35.

Crick, N. R., & Dodge, K. A. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child Development, 67*, 993–1002.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature, 48*, 424-455.

Durlak, J. A., & Wells, A. M. (1997). Primary prevention mental health programs for children and adolescents: A meta-analytic review. *American Journal of Community Psychology, 25*(2), 115-152.

Eisner, M. (2009). No effects in independent prevention trials: Can we reject the cynical view? *Journal of Experimental Criminology, 5*, 163–183.

Eisner, M., Jünger, R., & Greenberg, M. (2006). Gewaltpraevention durch die Foerderung emotionaler und sozialer Kompetenzen in der Schule: Das PATHS/ PFAD Curriculum [preventing violence by the promotion of emotional and social competences in schools: The PATHS curriculum]. *Praxis der Rechtspsychologie, 16* (1/2), 144-168.

Eisner, M., & Meidert, U. (2011). From enrollment to utilization – stages of parental engagement in a universal parent training program. *Journal of Primary Prevention, 32* (2), 83-93.

Eisner, M., Nagin, D. Ribeaud, D., & Malti, T. (2010). Effects of a universal parenting program for highly adherent parents: A propensity score matching approach. *Manuscript under revision.*

Eisner, M., & Parmar, A. (2007). Doing criminological research in ethnically and culturally diverse contexts. In K. Roy & A. Winkup (Eds.), *Doing research in crime and justice, 2nd edition* (pp. 171-199). Oxford: Oxford University Press.

Eisner, M., Ribeaud, D., & Malti, T. (in press-a). Large-scale criminological field experiments: The Zurich Project on the Development of Children. In S. Gadd, S. Karstedt, & S. Messner (Eds.), *Sage Handbook of Criminological Research Methods.* London: Sage.

Farrington, D. (2005). Childhood origins of antisocial behavior. *Clinical Psychology and*

Psychotherapy, 12, 177–190.

Farrington, D. P., & Welsh, B. C. (2003). Family-based prevention of offending: A meta-analysis. *Australian and New Zealand Journal of Criminology*, *36*, 127–151.

Fonagy, P., Twemlow, S. W., Vernberg, E. M., Mize Nelson, J., Dill, E .J., Little, T. D., et al. (2009). A cluster randomized controlled trial of child-focused psychiatric consultation and a school systems-focused intervention to reduce aggression. *Journal of Child Psychology and Psychiatry, 50*, 607–616.

Ganzeboom, H. B. G., Degraaf, P. M., Treiman, D. J., & Deleeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, *21*, 1–56.

Gottfredson, D., Kumpfer, K., Polizzi-Fox, D., Wilson, W., Puryear, V., Beatty, P., et al. (2006). The Strengthening Washington D.C. Families Project: A randomized effectiveness trial of family-based prevention. *Prevention Science, 7*, 57–74.

Greenberg, M. T., & Kusché, C. A. (2002). *Promoting alternative thinking strategies: Blueprint for violence prevention* (Book 10; 2nd ed). Boulder: University of Colorado.

Greenberg, M., Weissberg, R., O'Brien, M., Zins, J., Fredewricks, L., Resnik, H., et al. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional and academic learning. *American Psychologist, 58*, 466–474.

Gross, D., Garvey, C., Julion, W., Fogg, L., Tucker, S., & Mokros, H. (2009). Efficacy of the Chicago Parent Program with low-income African American and Latino parents of young children. *Prevention Science, 10*, 54–65.

Hahn, R., Fuqua-Whitley, D., Wethington, H., Lowy, J., Crosby, A., Fullilove, M., et al. (2007). Effectiveness of universal school-based programs to prevent violent and aggressive behavior: A systematic review. American Journal of Preventive Medicine, 33 (2, Supplement 1), S114-S129.

Hiscock, H., Bayer, J. K., Price, A., Ukoumunne, O. C., Rogers, S., & Wake, M. (2008). Universal parenting programme to prevent early childhood behavioural problems: Cluster randomised trial. *BMJ, 336,* 318-321.

Kim, J.-S. (2009). Multilevel analysis: An overview and some contemporary issues. In R.E. Millsap & A. Maydeu-Olivares (Eds.), t*he Sage handbook of quantitative methods in psychology* (pp. 337-361). Thousand Oaks: Sage.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce. W. T., & Kupfer, D .J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160,* 1566–1577.

Lacourse, E., Côté, S., Nagin, D. S., Vitaro, F., Brendgen, M., & Tremblay, R. E. (2002). A longitudinal-experimental approach to testing theories of antisocial behavior development. *Development and Psychopathology, 14,* 909–924.

Linares Scott, T. J., Short, E. J., Singer, L. T., Russ, S. W., & Minnes, S. (2006). Psychometric properties of the Dominic Interactive assessment. A computerized self-report for children. *Assessment, 13*(1), 16-26.

Lösel, F., Beelmann, A., Stemmler, M., & Jaursch, S. (2006). Prävention von Problemen des Sozialverhaltens im Vorschulalter: Evaluation des Eltern- und Kindertrainings EFFEKT [preventing behavior problems in preschool-aged children: evaluation of the parent- and child skills training EFFEKT]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 35*, 127–139.

Malti, T. (2011). Evaluating the impact of life skills programs for youth: A selective review. In Jacobs Foundation (Ed.), *Jacobs Foundation guidelines on monitoring and evaluating life skills for youth development (pp. 64-79).* Zurich, Switzerland: Jacobs Foundation.

Malti, T., & Noam, G. G. (Eds.) (2008). Where youth development meets mental health and

education: The RALLY approach. *New Directions for Youth Development, No. 120*. San Francisco: Jossey-Bass.

Masten, A. S., & Wright, M. O'D. (2009). Resilience over the lifespan: Developmental perspectives on resistance, recovery, and transformation. In J. W. Reich, A. J. Zautra, & J. S. Hall (Eds.), *Handbook of adult resilience* (pp. 213–237). New York: Guilford Press.

Metropolitan Area Child Study Research Group (2002). A cognitive-ecological approach to preventing aggression in urban settings: Initial outcomes for high-risk children. *Journal of Consulting and Clinical Psychology, 70* (1), 179-194.

Nowak, C., & Heinrichs, N. (2008). A comprehensive meta-analysis of Triple-P-positive parenting program using hierarchical linear modeling: Effectiveness and moderating variables. *Clinical Child and Family Psychology Review, 11*, 114–144.

Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology, 1*, 435–450.

Prinz, R.J., Sanders, M.R., Shapiro, C.J., Whitaker, D.J., & Lutzker, J.R. (2009). Population-based prevention of child maltreatment: The U.S. Triple P system population trial. *Prevention Science, 10*, 1-12.

Reid, M.J., Webster-Stratton, C., & Hammond, M. (2007). Enhancing a social classroom social competence and problem-solving curriculum by offering parent training to families of moderate- to high-risk elementary school children. *Journal of Clinical Child and Adolescent Psychology, 36* (4), 605-620.

Riggs, N. R., Greenberg, M. T., Kusche, C. A., & Pentz, M. A. (2006). The mediational role of neurocognition in the behavioral outcomes of a social-emotional prevention program in elementary school students: Effects of the PATHS Curriculum. *Prevention Science, 7*, 91-102.

Raver, C.C., Jones, S.M., Li-Grining, C., Zhai, F., Metzger, M.W., & Solomon, B. (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology, 77* (2). 302-316.

Roberts, C., & Torgerson, D. J. (1999). Understanding controlled trials: Baseline imbalance in randomized controlled trials. *British Medical Journal, 319*, 185.

Sanders, M. R. (1999). The Triple-P-positive parenting program: Towards an empirically validated multilevel parenting and family support strategy for the prevention of behavior and emotional problems in children. *Clinical Child and Family Psychology Review, 2*, 71–90.

Sanders, M. R., Markie-Dadds, C., Tully, L. A., & Bor, W. (2000). The Triple P—Positive Parenting Program: A comparison of enhanced, standard, and self-directed behavioral family intervention for parents of children with early onset conduct problems. *Journal of Consulting and Clinical Psychology, 68*, 624–640.

Sanders, M. R., & McFarland, M. L. (2000). The treatment of depressed mothers with disruptive children: A controlled evaluation of cognitive behavioral family intervention. *Behavior Therapy, 31*, 89–112.

Sanders, M.R., Turner, K.M.T., & Markie-Dadds (2002). The development and dissemination of the Triple P-Positive Parenting Program: A multi-level, evidence-based system of parenting and family support. *Prevention Science*, 3, 173-198.

Scott, S., Sylva, K., Doolan, M., Price, J., Jacobs, B., Cook, C., et al. (2009). Randomized controlled trial of parent groups for child antisocial behaviour targeting multiple risk factors: The SPOKES project. *Journal of Child Psychology and Psychiatry, 51*, 48–57.

Spoth, R., Guyll, M., & Shin, C. (2009). Universal intervention as a protective shield against exposure to substance use: Long-term outcomes and public health significance. *American Journal of Public Health, 99*, 2026-2033.

St Pierre, T. L., Osgood, D. W., Mincemoyer, C. C., Kaltreider, D. L., & Kauh, T. J. (2006). Results of an independent evaluation of project alert delivered in schools by cooperative extension. *Prevention Science, 6*(4), 305-317.

Tremblay, R. E., Loeber, R., Gagnon, C., Charlebois, P., Larivee, S., & LeBlanc, M. (1991). Disruptive boys with stable and unstable high fighting behavior patterns during junior elementary school. *Journal of Abnormal Child Psychology, 19*, 285–300.

Webster-Stratton, C., Reid, M. J., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: Evaluation of the Incredible Years Teacher and Child Training Programs in high-risk schools. *Journal of Child Psychology and Psychiatry, 49,* 471–488.

Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine, 33* (2, Supplement 1), S130–S143.

Table 1
Means (SDs) of Outcome Variables

	Treatment Condition															
Outcome Variable	Control (*n* = 356)				PATHS only (*n* = 360)				Triple-P only (*n* = 339)				PATHS X Triple-P (*n* = 306)			
Teacher Reports	T1	T2	T3	T4	T1	T2 [a]	T3	T4	T1	T2	T3	T4	T1	T2 [a]	T3	T4
Aggressive behavior	0.51 (0.68)	0.41 (0.55)	0.45 (0.60)	0.53 (0.68)	0.56 (0.63)	0.56 (0.61)	0.62 (0.69)	0.48 (0.68)	0.61 (0.70)	0.58 (0.69)	0.62 (0.64)	0.57 (0.69)	0.69 (0.72)	0.67 (0.68)	0.63 (0.61)	0.58 (0.71)
Nonaggressive ext. behavior	0.30 (0.46)	0.22 (0.38)	0.23 (0.43)	0.26 (0.47)	0.29 (0.44)	0.30 (0.45)	0.32 (0.50)	0.23 (0.44)	0.31 (0.52)	0.30 (0.49)	0.39 (0.57)	0.32 (0.49)	0.42 (0.52)	0.42 (0.59)	0.40 (0.58)	0.28 (0.46)
Impulsivity/ ADHD	1.11 (0.97)	0.85 (0.94)	0.87 (0.94)	1.14 (1.04)	1.27 (1.02)	1.15 (1.00)	1.13 (1.01)	1.00 (0.89)	1.23 (0.98)	1.06 (0.98)	1.08 (0.94)	1.16 (1.06)	1.41 (0.97)	1.37 (0.93)	1.22 (0.86)	1.14 (0.96)
Prosocial behavior	2.16 (0.82)	2.14 (0.88)	2.42 (0.92)	2.26 (0.79)	2.34 (0.79)	2.43 (0.76)	2.54 (0.77)	2.24 (0.80)	1.99 (0.84)	2.17 (0.78)	2.23 (0.74)	2.12 (0.80)	2.20 (0.80)	2.34 (0.82)	2.36 (0.85)	2.17 (0.77)
Parent Reports																
Aggressive behavior	0.63 (0.44)	0.66 (0.44)	0.66 (0.43)	0.53 (0.38)	0.63 (0.46)	0.71 (0.47)	0.68 (0.46)	0.51 (0.39)	0.58 (0.40)	0.66 (0.43)	0.65 (0.42)	0.52 (0.35)	0.55 (0.38)	0.62 (0.41)	0.61 (0.41)	0.48 (0.32)
Impulsivity/ ADHD	1.23 (0.65)	b	1.31 (0.71)	1.31 (0.72)	1.21 (0.64)	b	1.32 (0.70)	1.22 (0.67)	1.19 (0.65)	b	1.30 (0.65)	1.33 (0.70)	1.21 (0.64)	b	1.29 (0.63)	1.24 (0.67)
Prosocial behavior	2.56 (0.51)	2.67 (0.52)	2.67 (0.55)	2.67 (0.52)	2.60 (0.53)	2.67 (0.55)	2.65 (0.53)	2.70 (0.59)	2.59 (0.56)	2.71 (0.52)	2.70 (0.51)	2.71 (0.57)	2.56 (0.56)	2.70 (0.53)	2.65 (0.53)	2.77 (0.55)
Child Reports																
Aggressive behavior	0.17 (0.17)	0.13 (0.16)	0.12 (0.15)	0.22 (0.20)	0.20 (0.19)	0.17 (0.18)	0.15 (0.16)	0.21 (0.20)	0.16 (0.17)	0.13 (0.15)	0.12 (0.15)	0.20 (0.18)	0.15 (0.16)	0.13 (0.16)	0.12 (0.17)	0.19 (0.20)
Impulsivity/ ADHD	0.17 (0.18)	b	0.17 (0.19)	b	0.19 (0.20)	b	0.18 (0.20)	b	0.16 (0.19)	b	0.15 (0.18)	b	0.16 (0.18)	b	0.14 (0.19)	b
Prosocial behavior	0.81 (0.19)	.88 (.16)	0.92 (0.12)	0.89 (0.14)	0.82 (0.18)	0.88 (0.16)	0.90 (0.16)	0.90 (0.15)	0.83 (0.16)	0.89 (0.13)	0.92 (0.13)	0.91 (0.12)	0.81 (0.17)	0.89 (0.13)	0.90 (0.14)	0.89 (0.15)
Socially competent problem solving	0.70 (0.29)	b	0.77 (0.26)	b	0.72 (0.27)	b	0.77 (0.26)	b	0.72 (0.26)	b	0.74 (0.27)	b	0.73 (0.25)	b	0.79 (.24)	b
Aggressive problem solving	0.15 (0.21)	b	0.10 (0.17)	b	0.16 (0.21)	b	0.10 (0.17)	b	0.13 (0.19)	b	0.10 (0.17)	b	0.14 (0.20)	b	0.08 (0.16)	b

[a] T2 is a second baseline score, because of the time-lagged implementation of the PATHS only and the PATHS+ Triple-P interventions.

[b] No measure at assessment point.

Table 2

Parameter Estimates (Standard Errors) of Long-term Treatment Effects on Children's Externalizing Behavior

	Teacher Reports			Parent Reports		Child Reports	
Parameter	AB	NACD	ADHD	AB	ADHD	AB	ADHD[a]
Intercept	0.80 (0.18)***	0.31 (0.10)**	1.70 (0.21)***	0.73 (0.08)***	1.46 (0.12)***	0.20 (0.03)***	0.16 (0.04)***
Time	0.03 (0.03)	0.01 (0.02)	0.03 (0.04)	0.02 (0.01)*	0.01 (0.01)	0.01 (0.00)*	0.01 (0.01)
PATHS	0.12 (0.12)	0.03 (0.06)	0.30 (0.15)*	0.06 (0.04)	-0.04 (0.07)	0.06 (0.02)*	0.03 (0.02)
Triple-P	0.04 (0.12)	0.01 (0.07)	0.13 (0.13)	-0.01 (0.04)	-0.06 (0.06)	0.01 (0.02)	0.01 (0.03)
PATHS X Triple-P	-0.01 (0.19)	0.06 (0.11)	-0.01 (0.21)	-0.08 (0.06)	0.04 (0.10)	-0.08 (0.03)*	-0.02 (0.04)
Time × PATHS	-0.08 (0.04)*	-0.01 (0.02)	-0.11 (0.06)*	-0.03 (0.02)*	0.00 (0.02)	-0.01 (0.01)	-0.01 (0.01)
Time × Triple-P	-0.06 (0.04)	-0.01 (0.03)	-0.10 (0.08)	0.01 (0.02)	0.03 (0.02)	-0.01 (0.01)	-0.01 (0.01)
Time × (P X T)	0.05 (0.06)	-0.02 (0.04)	0.10 (0.09)	0.03 (0.03)	-0.02 (0.03)	0.02 (0.01)	0.00 (0.01)
Baseline behavior	1.09 (0.03)***	1.09 (0.02)***	1.06 (0.02)***	1.12 (0.02)***	1.10 (0.02)***	1.05 (0.03)***	1.35 (0.02)***
Time x PATHS x BB	0.02 (0.03)	0.02 (0.03)	-0.04 (0.02)**	0.02 (0.01)	-0.02 (0.02)	-0.02 (0.02)	-0.05 (0.03)
Time x Triple-P x BB	0.03 (0.03)	0.01 (0.04)	-0.02 (0.03)	0.00 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.06* (0.03)
Time x (P X T) x BB	-0.05 (0.04)	-0.03 (0.05)	0.03 (0.03)	-0.05 (0.02)*	-0.01 (0.04)	0.05* (0.03)	0.07 (0.04)
Control variables							
Girl	-0.21 (0.03)***	-0.08 (0.02)***	-0.40 (0.04)***	-0.10 (0.02)***	-0.20 (0.04)***	-0.05 (0.01)***	-0.02 (0.01)**
Special-class education	0.05 (0.08)	0.06 (0.06)	0.17 (0.11)	.010 (0.03)*	0.26 (0.06)***	0.02 (0.01)	0.03 (0.02)*
Non-Swiss nationality	0.07 (0.03)*	0.04 (0.02)	0.06 (0.07)	-.014 (0.02)***	-0.13 (0.04)**	-0.02 (0.01)*	-0.01 (0.01)
Socioeconomic status	-0.01 (0.01)	-0.01 (0.00)	-0.01 (0.00)	0.01 (0.01)***	0.00 (0.00)	0.00 (0.00)	0.01 (0.00)

AB = Aggressive behavior. NACD = Nonaggressive externalizing behavior. ADHD = Attention deficit hyperactivity disorder. P = Paths. T = Triple-P. P X T = PATHS X Triple-P. BB = Baseline behavior.

[a]Child ratings of impulsivity/attention-deficit-disorder were collected at T3.

*$p < .05$. **$p < .01$. ***$p < .001$

Table 3
Parameter Estimates (Standard Errors) for Long-term Treatment Effects on Children's Social Competence

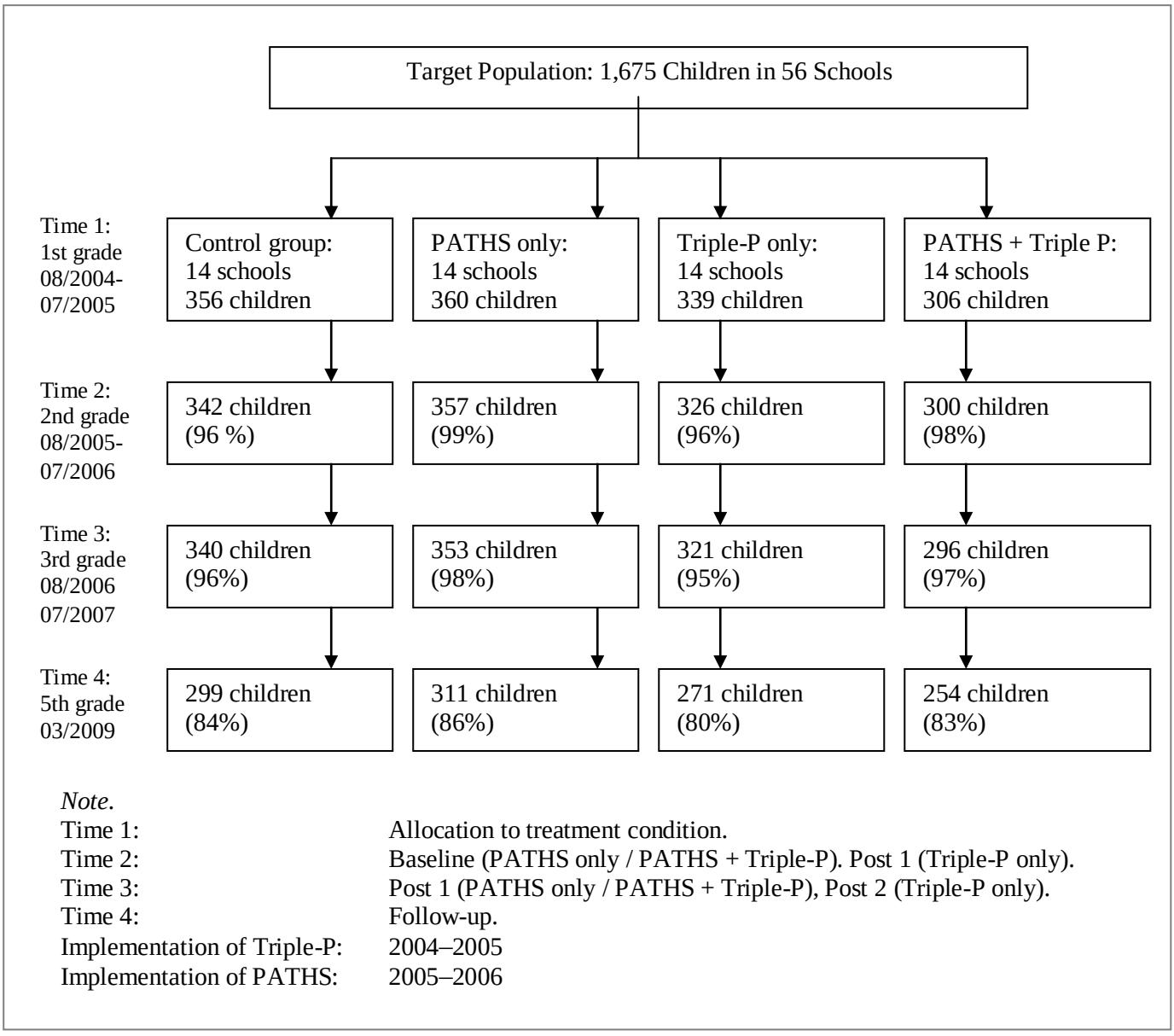
	Prosocial Behavior			Problem Solving	
Parameter	Teacher Reports	Parent Reports	Child Reports	Aggressive[a]	Socially Competent[a]
Intercept	1.87 (0.28)***	2.12 (0.09)***	0.69 (0.03)	0.20 (0.04)***	0.61 (0.05)***
Time	-0.02 (0.07)	0.06 (0.03)*	0.05 (0.01)	-0.03 (0.01)**	0.05 (0.02)**
PATHS	0.30 (0.24)	0.04 (0.06)	0.01 (0.02)	0.03 (0.03)	0.03 (0.05)
Triple-P	-0.05 (0.21)	0.03 (0.05)	0.03 (0.02)	-0.05 (0.03)	0.07 (0.05)
PATHS X Triple-P	-0.07 (0.32)	-0.09 (0.07)	-0.02 (0.03)	0.03 (0.04)	-0.07 (0.06)
Time × PATHS	-0.08 (0.08)	-0.01 (0.03)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.02)
Time × Triple-P	-0.01 (0.08)	0.03 (0.04)	-0.01 (0.01)	0.02 (0.01)	-0.03 (0.02)
Time × (PATHS X Triple-P)	0.05 (0.10)	0.01 (0.05)	0.01 (0.01)	-0.02 (0.01)	0.04 (0.02)
Control variables					
Girl	0.49 (0.04)***	0.20 (0.03)***	0.05 (0.01)***	-0.05 (0.01)***	0.09 (0.01)***
Special-class education	-0.10 (0.07)	-0.03 (0.05)	-0.03 (0.01)**	0.04 (0.02)*	-0.08 (0.02)**
Non-Swiss nationality	-0.09 (0.04)*	0.11 (0.03)**	0.01 (0.01)	-0.01 (0.01)	0.01 (0.01)
Socioeconomic status	-0.01 (0.02)*	0.00 (0.00)	0.01 (0.00)**	0.00 (0.00)	0.01 (0.00)

[a]Aggressive and socially competent problem solving skills were measured at T3.

*$p < .05$. **$p < .01$. ***$p < .001$.

Figure Caption

Figure 1. Flow diagram of study participation and treatment status, Time 1 to Time 4



Note.
Time 1: Allocation to treatment condition.
Time 2: Baseline (PATHS only / PATHS + Triple-P). Post 1 (Triple-P only).
Time 3: Post 1 (PATHS only / PATHS + Triple-P), Post 2 (Triple-P only).
Time 4: Follow-up.
Implementation of Triple-P: 2004–2005
Implementation of PATHS: 2005–2006