

Taking snapshots of preventive interventions

On the effectiveness of preventive interventions for youth
and how it relates to implementation and conflict of interest

Ferry X. Goossens

Cover by Robbert Oidtmann

Layout by Ferdinand van Nispen, *my-thesis.nl*

Printed by GVO drukkers & vormgevers B.V.

ISBN: 978-94-6332-111-2

© 2016 Ferry X. Goossens

All rights reserved. No part of this publication may be produced, stored in a retrieval system, or transmitted in any form of by any means, mechanically, by photocopy, by recording, or otherwise, without the permission from the author. The copyright of the articles that have been published, has been transferred to the respective journals.

Taking snapshots of preventive interventions

On the effectiveness of preventive interventions for youth
and how it relates to implementation and conflict of interest

"Taking snapshots of preventive interventions"

Over de effectiviteit van preventieve jeugdinterventies
en de relatie met implementatie en belangenverstrengeling
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college van promoties
in het openbaar te verdedigen op
vrijdag 10 februari 2017 des ochtends te 10.30 uur

door

Ferry Xaverius Goossens

geboren op 26 september 1978

te Utrecht

Promotor:

Prof. dr. B. Orobio de Castro

Copromotoren:

Dr. R. Monshouwer

Dr. S.A. Onrust

Beoordelingscommissie:

Prof. dr. M. Deković

Prof. dr. R.C.M.E. Engels

Prof. dr. C. Finkenauer

Prof. dr. A.E. Kunst

Prof. dr. T.A. van Yperen

"Het leven is een gedoetje."

- René Gude

Contents

| | | |
|------------------|--|------------|
| Chapter 1 | General introduction | 11 |
| Part 1 | Preventive interventions | 25 |
| Chapter 2 | Implementation of PATHS through Dutch municipal health services: a quasi-experiment International Journal of Conflict and violence (2012), 6, 234-248 | 27 |
| Chapter 3 | Effectiveness of a brief school-based intervention on depression, anxiety, hyperactivity, and delinquency: a cluster randomized controlled trial European Child & Adolescent Psychiatry (2016), 25, 639-648 | 53 |
| Chapter 4 | Effectiveness of an empowerment program for adolescent second generation migrants: a cluster randomized controlled trial Children and Youth Services Review (2016), 64, 128-135 | 79 |
| Part 2 | Conflict of interest | 109 |
| Chapter 5 | Dual roles in research: independence of research within a Dutch database of effective youth interventions De Psycholoog (2016), 6, 40-53 | 111 |
| Chapter 6 | Conflict of interest in school-based substance use prevention programs: a meta-regression analysis | 133 |
| Chapter 7 | General discussion | 157 |
| | Summary | 177 |
| | Samenvatting | 183 |
| | Dankwoord | 189 |
| | Over de auteur | 193 |

1

General introduction

General introduction

If we want to know whether preventive interventions work, we have to study their effects. These studies are basically snapshots, snapshots that we use as indicators for what the intervention will achieve when there is no camera at hand. In this dissertation we have taken a snapshot of three different interventions. In addition we have tried to gain more in-depth knowledge concerning the art of taking photos. Is it a matter of choosing the right angle for the best picture, retouching the photo, smiling when the camera is around, or making snapshots of daily life as it unfolds?

Evaluations of preventive interventions

The past decades many Western countries, including the Netherlands, have invested in the evaluation of preventive interventions in the field of public mental health. Examples of this are programs that aim to stimulate the social and emotional development of children facing social problems with their classmates (e.g. Vliek, Overbeek & Orobio de Castro, 2015), to target bullying in elementary schools (e.g. Kärnä et al., 2011), or to prevent depression in adults (e.g. Warmerdam et al., 2008). In these studies the main question generally is whether the intervention is effective in achieving its goals. This is an important question considering that if we want to enhance public mental health it is important to know which interventions are effective in reaching their goals and which are not. In this dissertation we have done so for three preventive interventions, which respectively aim to improve the social and emotional development of children in elementary school (PATHS), reduce alcohol use and mental health problems in students in secondary education (Prevention), and empower adolescent second generation migrants (POWER). The main aim of these studies was to test the effectiveness of these preventive interventions. An underlying objective was to gather information that could be helpful to others in making the decision whether, or how, the intervention should be implemented on a broad scale in the Netherlands.

Efficacy, effectiveness and dissemination trials

In general, the effects of preventive interventions are firstly studied in efficacy trials. Efficacy is the extent to which an intervention (technology,

treatment, procedure, service, or program) does more good than harm when delivered under optimal conditions (Flay et al., 2005a). Efficacy studies test interventions under optimal conditions, such as having ample resources, carefully supervised intervention personnel, and a high quality of program implementation. It is clear that this situation differs from day to day practice in implementing interventions, and therefore should be considered as a first check regarding whether the intervention is effective. Efficacy trials thereby distinguish from effectiveness studies in which program effects are tested under more daily circumstances, such as when the intervention is implemented as part of the daily routine of a professional, and without extra support from the researchers. This type of study generally gives a better idea of the results that can be achieved with the intervention in daily practice than efficacy studies. Normal variation in the quantity and the quality of the implementation of the intervention, caused by variation in professionals delivering the intervention, is now part of the study after all. Effectiveness trials subsequently also differ from dissemination trials, in which an intervention is tested when implemented on a large scale, for example in a whole region or nationwide. This type of research can be seen as the proof of success in the search for preventive interventions that indeed have a significant impact on the public mental health (Flay et al., 2005a).

This ordering simultaneously illustrates how interventions generally are developed. Starting in the ‘laboratory setting’ with an idea about how the behavior of the target group can be changed, and ending as a fully developed intervention with glossy materials that is implemented on a broad scale in daily life practice. All three types of research are useful in detecting interventions that will ultimately have an impact on the public health. As both the interventions PATHS and Preventure have found to be efficacious in earlier studies, and as POWER is an intervention that due to its nature can hardly be tested under ideal conditions, we performed three effectiveness trials. Ultimately, we wanted to know whether these interventions were effective before considering disseminating them on a broader scale.

Implementation

An essential difference between efficacy trials and effectiveness trials is the implementation process (Flay et al., 2005a; Flay et al., 2005b). In efficacy trials the aim is to basically implement the intervention as thoroughly as possible, and thereby test whether the intervention can change the participants' behavior, or the determinants of their behavior. An intervention can therefore also be seen as a theory about behavioral change. A theory that is packed into an intervention that, for example, consists of manuals, workbooks and exercises. However, if during an efficacy trial a theory of behavioral change is tested, then what is tested during an effectiveness trial? The answer is that the same theory of behavioral change is tested, yet, it is now tested in a more dynamic context. A context that could influence the quantity and quality of the delivery of the implementation of the intervention, and thereby also the study outcomes. To have a clear idea why an efficacious intervention is, or is not, also effective during an effectiveness trial, it is thus important to not only measure the outcome variables, but also measure the implementation process. Answers to questions such as "How was the intervention implemented?", and "Was the implementation according to plan?", are needed to understand the findings of effectiveness studies. In both the PATHS and POWER study we therefore also measured implementation variables in terms of the quantity and the quality of the intervention that was delivered. Subsequently, we have tested whether the intervention effects were influenced by these variables. In the Preventure study we did not measure the actual implementation, but some implementation factors were discussed that could have possibly influenced the intervention outcomes. These three studies are described in chapters two to four.

Conflict of interest

There is another aspect known to be of influence on intervention trial outcomes: the researchers themselves. It is known that intervention trials performed by the intervention developer, a colleague of him or her, or by researchers simultaneously working for the licensee company, in general show larger effect sizes than trials performed by independent researchers (Beelman & Lösel, 2006; Farrington & Welsh, 2003; Gorman &

Conde, 2007; Perlis et al., 2005; Petrosino & Soydan, 2005; Reyno & McGrath, 2006; St Pierre, Osgood, Mincemoyer, Kaltreider, & Kauh, 2006; Valentine, et al., 2011; Wilson, Lipsey & Derzon, 2003). There are two possible explanations for this phenomenon, known as the cynical view and the high fidelity view. This view asserts that researchers with a conflict of interest consciously manipulate the research results, and/or (un)consciously make choices during the study that increase the chance of positive intervention effects. The view holds that they - more than independent researchers - are a little too eager to find intervention effects. The second view asserts that the licensees and program developers are more aware of the theoretical foundation of the intervention. As a result, they are more knowledgeable than others on how to implement the intervention in the most faithful manner. This hypothesis assumes that nothing is substantially wrong with the study results. The only problem is that the findings lack external validity, i.e., they cannot be generalized to routine applications of the same program as the intervention developer cannot always be present during regular implementation (Eisner, 2009; Petrosino & Soydan, 2005). Nonetheless, the advantage of this mechanism being a mediating variable is that there are still opportunities to try to reach the same high level of implementation in daily routine as was achieved during the trial. When the cynical view is true, the study outcomes overestimate the effects of the intervention in daily life anyhow.

Triggered by the fact that the effectiveness of both the PATHS and Preventure interventions were largely based on studies performed by the intervention developers and licensees, and that the POWER intervention was licensed by the Trimbos Institute, I wanted to take a closer look at this topic in this dissertation. It appeared, however, impossible to statistically test the exact influence of this type of conflict of interest for the three interventions we studied, as the number of effect studies per intervention was limited. Therefore, we decided to study this phenomenon within two large datasets containing intervention studies. These studies are described in chapters five and six. We describe to what extent interventions that have been shown to be effective are studied by independent researchers, using both a Dutch database of evidence based public health interventions (chapter five), as well as an international

dataset of school-based preventive interventions targeting alcohol, tobacco and drugs use (chapter six). Subsequently we checked within the latter of these two datasets whether there is a relation between the effect size found during the trial, and being both the researcher and the intervention developer, and/or being both the researcher and working for the licensee of the intervention. Lastly we tested whether this relation could specifically be explained by the cynical or the high fidelity view.

Aims of this manuscript

The first aim of this dissertation was to test the effectiveness of three preventive mental health interventions for youth (PATHS, Preventure and POWER) for both the entire target group, as well as for subgroups in terms of gender (PATHS and POWER), age (PATHS), ethnicity (POWER), levels of problem behavior (POWER), and personality characteristics (Preventure). Moreover, we aimed to know whether the intervention effects of PATHS and POWER were moderated by the level of implementation. Our second aim was to describe to what extent being a developer-as-researcher, being his or her colleague, or being a licensee-as-researcher is common in intervention trials for youth. We wanted to know to what extent such a conflict of interest is associated with intervention trial outcomes, and whether this effect could especially be explained from the cynical view or the high fidelity view.

Outline of this manuscript

Chapter 2: PATHS

In chapter two we describe the effectiveness study of the PATHS intervention, a comprehensive school-based prevention program that aims to enhance social and emotional competence, and reduce behavioral and mental problems of children from Kindergarten to sixth grade. The intervention displayed positive intervention effects in studies (mainly) performed in the United States, and thus we wanted to test whether the intervention would also be effective in the Netherlands. To be more precise, whether PATHS would be effective in regular education while using an implementation strategy that would make it potentially possible to implement the intervention nationwide. The

effectiveness of PATHS was assessed using a quasi-experimental design with an intervention condition ($n = 674$) and a waiting list comparison condition ($n = 620$), each containing nine regular elementary schools. The intervention effects were measured during the first two years of the PATHS implementation. The research questions that are answered in chapter two are:

- 1) Is there an intervention effect of PATHS in terms of a significant reduction in externalizing and internalizing problem behaviors, and a significant improvement in social and emotional skills, as compared to the control group?
- 2) Is there a larger intervention effect in classes with higher implementation quality, and in classes with higher implementation quantity?

Chapter 3: Preventure

The effectiveness of Preventure is studied in chapter three. Preventure is a school-based intervention for adolescents that uses a personality-targeted approach and has displayed positive intervention effects in studies performed outside the Netherlands. We wanted to investigate whether the intervention would be useful for the Dutch context and tested the effectiveness of Preventure. We conducted a randomized controlled trial among 13 to 15-year-old adolescents in secondary schools. In this chapter we describe the secondary outcomes, which include a range of mental health outcomes at 2, 6, and 12 months post intervention. The effectiveness of the intervention on the primary, alcohol related, outcomes are published in Lammers et al. (2015). Fifteen schools were randomly assigned to the intervention condition (including 343 adolescents), and the control condition (including 356 adolescents). The intervention condition consisted of two 90 minute group sessions, carried out at the participants' schools by a qualified counsellor and a co-facilitator. The research questions that are answered in chapter three are:

- 3) Is there an intervention effect of Preventure in terms of a significant reduction of depression, anxiety, conduct

problems and hyperactivity rates, and delinquent risk taking behavior in the total intervention group as compared to the control group?

- 4) Are these intervention effects moderated by personality profiles?

Chapter 4: POWER

Chapter four describes the effectiveness of an empowerment program for adolescent second generation migrants: POWER. The objective of POWER is to reduce the risk of marginalization and problem behavior by means of the development and expansion of the feeling of personal control over one's life and environment (i.e. mastery), problem-solving abilities, and competencies. The intervention consists of three elements: (1) a culturally sensitive empowerment group course for youngsters, (2) a course for their parents, and (3) a community approach, i.e. involving relevant local organizations during the project. To test its effectiveness we conducted a cluster randomized controlled trial including a pretest and posttest. The sample consisted of 248 Moroccan, Turkish, Surinamese, and Antillean/Aruban second generation migrants between 12 and 18 years of age from 16 lower-class neighborhoods in nine cities in the Netherlands. Participant groups were randomly assigned, and 132 youngsters were placed in 16 groups of the intervention condition and 116 youngsters were placed in 16 groups of the control condition. The research questions that are answered in chapter four are:

- 5) Is there an intervention effect of POWER in terms of a significant reduction of the participants' problem behavior and level of social marginalization, and does the intervention improve the participants' sense of mastery and coping skills as compared to the control group?
- 6) Are the intervention effects moderated by gender, the level of problem behavior at baseline, ethnicity, or the implementation process?

Chapter 5: Conflict of interest in a Dutch database of effective youth interventions

In chapter five we describe to what extent evidence based preventive interventions have been studied by independent researchers, in terms of that they were not also being the developer of the intervention, being his or her colleague, or employed by the licensee. We performed this study using 86 papers regarding 26 interventions that had the predicate of evidence based in the Dutch database of effective youth interventions. For every intervention, it was investigated who the developer(s) were, for which organization he and/or she worked at the time of the development of the intervention, and which organizations had the license or patent for the intervention, to determine if the authors of the research articles were indeed independent. In the results we present which percentage of the studies was performed by researchers with a conflict of interest in total, as well as per level of indications for evidence, and per intervention. We further describe which percentage of the research articles correctly contained a statement disclosing possible conflict of interests. The research questions that are answered in chapter five are:

- 7) In terms of percentages, to what extent are evidence based preventive interventions for youth studied by independent researchers in terms of that they were not also being the developer of the intervention, being his or her colleague, or employed by the licensee?
- 8) What percentage of research articles contain a correct statement disclosing possible conflict of interests?

Chapter 6: Conflict of interest in school-based substance use prevention programs

In chapter six we describe to what extent studies regarding school based interventions, in which substance use is one of the outcome variables, are studied by independent researchers. We used an international dataset of 241 studies regarding 288 interventions. In addition to what was examined in chapter five, in this study we also tested to what extent such a conflict of interest influences intervention trial outcomes

in terms of having a positive relation with the effect sizes, using meta-regression analyses. Furthermore we performed mediation analyses to test whether these differences in effect sizes could be explained from the cynical view or the high fidelity view. The research questions that are answered in chapter six are:

- 9) In terms of percentages, to what extent are school-based substance use prevention programs studied by researchers with a conflict of interest in terms of also being the developer of the intervention, and/or working for the licensee?
- 10) Is having a conflict of interest associated with the effect size of the study?
- 11) Is this effect mediated by the cynical view or the high fidelity view?

Chapter 7: General discussion

Chapter seven discusses the main findings of all the studies, the methodological issues, and provides some suggestions for future research. We advocate for more independent effectiveness and dissemination trials for preventive interventions, in which the implementation is measured just as carefully as the intervention outcomes. We also discuss what the implications of the findings are for the decision to, or how to, implement the present interventions on a broader scale in the Netherlands.

References

- Beelman, A. & Lösel, F. (2006). Child social skills training in developmental crime prevention: Effects on antisocial behavior and social competence. *PsycotHEMA*, 18, 603-610.
- Farrington & D.P., Welsh, B.C. (2003). Family-based Prevention of Offending: A Meta-analysis. *The Australian and New Zealand Journal of Criminology*, 36, 127-151.
- Eisner, M. (2009). No Effects in Independent Prevention Trials: Can We Reject The Cynical View? *Journal of Experimental Criminology*, 5, 163-83. doi:10.1007/s11292-009-9071-y.
- Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., Gottfredson, D., Kellam, S.G., ... Ji, P. (2005a). Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination.
- Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., Gottfredson, D., Kellam, S.G., ... Ji, P. (2005b). Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination. *Prevention Science*, 6, 151-175.
- Gorman D.M. & Conde E. (2007). Conflict of interest in the evaluation and dissemination of 'model' school-based drug and violence prevention programs. *Evaluation and Program Planning*, 30, 422-429.
- Kärnä, A., Voeten, M., Little, T.D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011). A large-scale evaluation of the KiVa antibullying program. *Child Development*, 82, 311-330.
- Lammers, J., Goossens, F., Conrod, P., Engels, R., Wiers, R.W., Kleijnjan, M. (2015). Effectiveness of a selective intervention program targeting personality risk factors for alcohol misuse among young adolescents: Results of a cluster randomized controlled trial. *Addiction*, 110, 1101-1109.
- Perlis, R.H., Perlis, C.S., Wu, Y., Hwang, C., Joseph, M. & Nierenberg, A.A. (2005). Industry Sponsorship and Financial Conflict of Interest in the Reporting of Clinical Trials in Psychiatry. *The American Journal of Psychiatry*, 162, 1957-1960.
- 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.
- Reyno, S.M., & McGrath, P.J. (2006). Predictors of parent training efficacy for child externalizing behavior problems - a meta-analytic review. *Journal of Child Psychology and Psychiatry*, 47(1), 99-111.
- St. Pierre, T.L., Osgood, D.W., Mincemoyer, C.C., Kaltreider, D.L., & Rauh, T.J. (2005). Results of an Independent Evaluation of Project ALERT Delivered in Schools by Cooperative Extension. *Prevention Science*, 6, 305-317.
- Valentine, J.C., Biglan, A., Boruch, R.F., González Castro, F., Collins, L.M., Flay, B.R., ... & Schinke, S.P. (2011). Replication in Prevention Science. *Prevention Science*, 12, 103-117.
- Vliek, L., Overbeek, G., & Orobio de Castro, B. (2015). A randomized trial of parent-child Topper Training: Effects on psychosocial problems, self-esteem and victimization. In L. Vliek, Effects of Kanjertraining (Topper Training) on emotional problems, behavioural problems and classroom climate (Proefschrift) (pp. 89-119). Universiteit Utrecht, Utrecht
- Warmerdam, L., Straten, van A., Twisk, J., Riper, H., & Cuijpers, P. (2008). Internet-Based Treatment for Adults with Depressive Symptoms: Randomized controlled Trial. *Journal of Medical Internet Research*, 10(4):e44.
- Wilson, S.J., Lipsey, M.W., & Derzon, J.H. (2003). The Effects of School-Based Intervention Programs on Aggressive Behavior: A Meta-Analysis. *Journal of Consulting and Clinical Psychology*, 71, 136-149.

Part 1

Preventive interventions

2

Implementation of PATHS through Dutch municipal health services: a quasi experiment

Ferry Goossens,
Evelien Gooren
Bram Orobio de Castro
Rees van Overveld
Goof Buijs
Karin Monshouwer
Simone Onrust
Theo Paulussen

Published as:

Goossens, F.X., Gooren, E.M.J.C., Orobio de Castro, B., Overveld, C.W. van, Buijs, G.J., Monshouwer, K., Onrust, S.A., & Paulussen T.G.W.M. (2012). Implementation of PATHS Through Dutch Municipal Health Services: A Quasi-Experiment. *International Journal of Conflict and Violence*, 6, 234-248.

Authors' contributions:

FG carried out the study, analyzed the data, and wrote the manuscript. EG coordinated the acquisition of data, cleaned the data, and revised the manuscript. BODC revised the manuscript. KVO trained and supported the municipal health services and revised the manuscript. GB contributed to the design of the study, developed the implementation materials, and revised the manuscript. RM and SO revised the manuscript. TP designed the study and revised the manuscript. All authors have given final approval of the version to be published.

Conflict of interest:

FG, EG, BODC, GB, RM, SO and TP declare they have no conflict of interest. KVO is employed at the Dutch licensee of PATHS.

Abstract

Only a limited number of effectiveness studies have been performed to study the benefits of efficacious behavior problems prevention programs for children when implemented through national health service systems. This study uses a quasi-experimental design to test the effectiveness of the school-based PATHS prevention program (Providing Alternative THinking Strategies) when implemented through Dutch municipal health services by health promotion professionals. A sample of 1,294 children was followed for two years: 674 children attending nine schools providing PATHS and 620 children in nine comparison schools. We hypothesized finding an intervention effect of PATHS in terms of a significant reduction in teacher- and student-rated externalizing and internalizing problem behaviors, and a significant improvement in teacher-, student-, and peer-rated social skills and emotional skills. In fact, the results show low levels of program implementation and no intervention effects on problem behavior or social and emotional skills, suggesting that it is hard to reproduce positive intervention effects where an efficacious social-emotional prevention program is implemented through a national health service. More research is needed on the specific conditions required to implement efficacious programs effectively.

Introduction

It has been shown that school-based prevention programs aimed at social and emotional learning can be efficacious (Durlak et al., 2011). Usually these programs are evaluated in efficacy trials characterized by optimal conditions, such as well-trained and carefully supervised intervention personnel, and ample resources, and by involvement of the program developers in the implementation process and research (Flay et al., 2005). The better the program implementation and the greater the program fidelity, the stronger the effects are (Beelmann & Raabe, 2009; Eisner, 2009; Salmivalli, Raukiainen & Voeten, 2005; Wilson & Lipsey, 2007). However, efficacy trials have high internal validity but weaker external validity, for example actual utilization under normal community conditions (Weisz et al., 1995). Very few effectiveness studies have examined the benefits of efficacious prevention programs on a large scale (Dodge, 2009). Given that to date few efficacious prevention programs have been successfully implemented on a large scale (Elliot & Mihalic, 2004), and their population impact on the prevention of problem behavior is very modest (Dodge, 2009), more research on this topic is needed.

The PATHS Program

PATHS is a comprehensive school-based prevention program that aims to enhance social and emotional competence and reduce behavioral and mental problems of children from kindergarten to sixth grade. PATHS is based on five conceptual models that form a theoretical framework focusing on a broad range of protective and risk factors for internalizing and externalizing problem behavior. The curriculum provides one to three sessions each week focusing on self-control, emotional understanding, positive self-esteem, relationships, and interpersonal problem-solving skills. Lessons include didactic instruction, role playing, class discussion, modeling by teachers and peers, social and self-reinforcement, and worksheets. The lessons are grouped in three major units: readiness and self-control, feelings and relationships, and interpersonal cognitive problem-solving (Greenberg & Rusché, 1993).

PATHS in the Netherlands

For a couple of decades now, the efficacious school-based prevention program PATHS (Providing Alternative THinking Strategies) (Domitrovich, 1999; Kusché & Greenberg, 1994) has been implemented in the Netherlands by a small group of dedicated professionals employed at the Dutch national licensee, which has directly trained staff at three hundred schools over the last twenty-five years. However, with approximately seven thousand elementary schools in the Netherlands this is no more than 5% of the potential, and its impact on society is therefore probably still limited. The present study examines the effectiveness of PATHS when implemented through health promotion professionals from Dutch municipal health services (MHSS). In this implementation strategy, the national licensee trained MHS health promotion professionals who in turn trained teachers. As it is the daily job of such professionals to implement school-based prevention programs on topics like bullying, alcohol, obesity, sexually transmitted diseases, and social emotional competence, it was assumed that they could also be a useful link in implementing PATHS.

PATHS Efficacy and Effectiveness

PATHS has been studied in a number of trials within a variety of populations, including children in regular education, children at risk of behavior problems, children in special education, and children with hearing impairments. As the current study concerns the implementation of PATHS for children in regular education, we concentrate on studies focusing on that target group, as well as studies that examine the implementation process.

The results of the first PATHS study in regular education indicated that the intervention was effective in improving grade 2 and 3 children's range of vocabulary and fluency in discussing emotional experiences, their efficacy beliefs regarding the management of emotions, and their developmental understanding of some aspects of emotions (Greenberg et al., 1995). A second, larger study of 329 second and third graders showed that the intervention promoted inhibitory control, verbal fluency, and diminished internalizing and externalizing problem behaviors (Riggs et

al., 2006). A third study, with 246 pre-school children, showed that children exposed to PATHS intervention had higher emotion knowledge skills and were more socially competent than peers (Domitrovich, Cortes & Greenberg, 2007). In all these studies, the program developers were involved in the research, and the level of support was high (i.e. teachers received monthly or even weekly consultation from the project staff to enhance the quality of implementation).

These intensive procedures may have contributed significantly to the implementation quality and the effects found. This assumption is supported by the results of an effectiveness trial involving 350 first graders in six inner-city public schools in a high-risk urban community in the United States (Ram, Greenberg & Walls, 2003). In this study, significant intervention effects were found only in schools where both implementation quality and support from the principal were high. Another study showed that it was not the number of PATHS sessions received, but the quality of these sessions that positively influenced the outcomes (CPPRG, 1999).

Interestingly, PATHS with less intense support was also studied. A Dutch study concerning the effectiveness of PATHS for boys with severe aggressive behavior problems showed positive effects on proactive and reactive aggression (Louwe et al., 2007a), in a context where training and support were provided directly by the national licensee. However, a negative trend was found for children in special needs schools, where implementation quality was substantially lower than in other schools (Louwe et al., 2007b). A recent study in Zurich showed intervention effects on teacher- and parent-rated aggressive behavior, and teacher-rated ADHD, but no significant positive effects for nine other externalizing and internalizing outcomes. The teachers who implemented PATHS received two days' training, and local coaches were trained to visit the classes and provide feedback to the teachers (Malti, Ribeaud & Eisner, 2011). Finally, the intervention was studied in ten US public elementary schools (SCDRC, 2010). Of the twenty child-level outcomes, none was significant. The authors suggest that the lack of positive findings was probably caused by the control condition not being a non-treatment condition,

but a standard practice condition including schools that use other social and character development activities. In sum, PATHS has been shown to be efficacious and potentially effective, but effectiveness depends to a large extent on the implementation conditions.

Aims and Hypotheses

The present study explored the effectiveness of PATHS when implemented through a regular health service system, i.e. Dutch municipal health services. We hypothesized finding an intervention effect of PATHS in terms of a significant reduction in teacher- and student-rated externalizing and internalizing problem behaviors, and a significant improvement in teacher-, student-, and peer-rated social and emotional skills. We further hypothesized finding a larger effect in classes with higher implementation quality, and in classes with higher implementation quantity.

Method

The effectiveness of PATHS was assessed using a quasi-experimental design with an intervention ($n = 674$) and a waiting list comparison ($n = 620$) condition, each containing nine regular elementary schools. The intervention effects were measured during the first two years of PATHS implementation.

Implementation Strategy

First, the national project team recruited MHSs to participate in the study. All thirty Dutch MHSs were approached by e-mail, and three were willing and able to participate in the study. At each participating MHS, three health promotion professionals were recruited. These nine professionals were facilitated in their task of training and supporting teachers by means of 1) a two-day PATHS course run by the Dutch PATHS licensee, 2) a manual, a model recruitment letter, and a PATHS brochure for recruiting schools, 3) a pre-implementation teacher training protocol, and 4) follow-up support by the Dutch PATHS licensee.

Second, the health promotion professionals sent all principals of elementary schools in their region the PATHS model recruitment letter and the PATHS brochure. If principals and teachers expressed the intention to participate in the study, an implementation agreement was signed between the school and the MHS.

Third, schools were allocated randomly to either the intervention condition or the waiting list comparison condition, with four exceptions. In order to maximize ecological validity, we refrained from randomizing four of the schools. All the schools were very willing to implement PATHS, but two preferred to start two years later for organizational reasons and two were willing to participate only if they could start directly. These requests were complied with.

Fourth, under supervision of the national licensee, the health promotion professionals provided a two-day pre-implementation training course for the teachers in the intervention condition. Directly after this course, all teachers in the intervention condition started giving their PATHS lessons.

Fifth, in the course of each school year the teachers received a half-day booster session. Additionally, the PATHS coordinators received a half-day support session. Each school designated a staff member as PATHS coordinator with the task of supporting the implementation in their school by delivering supervision and feedback to the teachers.

Sixth, each school in the experimental condition organized, in co-operation with the MHS, an information meeting for parents. Parents also received frequent written information about the content of the program.

The Health Promotion Context

The PATHS strategy had previously been applied successfully to the implementation of a school-based sex education program by MHSS in the Netherlands (Wiefferink et al., 2005). The strategy had a positive impact on teachers' extent of use, as well as their curriculum-related beliefs. Moreover, implementing school-based prevention programs

through MHS health promotion professionals is a common strategy for the implementation of school-based prevention programs in the Netherlands. So although this strategy clearly differs from the more intensive procedures advised by the developers – and as generally used in PATHS trials – it is a good example of how prevention programs are implemented in the Dutch context of school-based health promotion and thus complies with our aim of exploring the effectiveness of PATHS when implemented through a regular health service system.

The Intervention

The version of PATHS used in this study (SVO, 2005) consisted of 161 lessons of 20 to 30 minutes, spread over the eight years of elementary school. It was an update of a Dutch translation of the US curriculum for regular schools that had already been in use for several years in the Netherlands, supplemented by a translation of the pre-school PATHS program (Domitrovich et al., 1999). During the study, all children in the intervention condition received the PATHS program for two years, consisting of approximately forty PATHS lessons. As children in the higher grades did not start the PATHS lessons from kindergarten, they received extra lessons to inform them about the story line and basic principles of the PATHS program. Nevertheless, as PATHS is a cyclical program, all major units are discussed each school year. The study proposal was approved by the medical ethical committee, which stated that passive informed consent was adequate.

Data Collection

Teacher assessments and student assessments were conducted in both conditions: at the start (T0) and the end (T1) of the first year, and at the start (T2) and the end (T3) of the second year. Each teacher received his/her questionnaires by post and was asked to fill out ten to thirty questionnaires each wave. Child questionnaires were completed in face-to-face interviews with the three youngest cohorts that were followed in this study (kindergarten, grade 1 and grade 3) and by means of a self-report questionnaire for the oldest cohort (grade 5). The student assessments lasted 20 to 30 minutes per child per wave and were conducted by approximately seventy-five trained graduate psychology students who

did not know the intervention condition of the school. Questionnaires concerning implementation quantity were sent to all teachers in the intervention condition four times a year. Implementation quality was measured twice, at the end of the first year and at the end of the second year. Children received a gift worth +/- €0.75 at each measurement point. To incentivize school participation, training and materials were provided free of charge (+/- €4,000).

Measures

Problem Behavior

The Problem Behavior at School Interview (PBSI) (ErasmusMC, 2000) is a forty-two-item questionnaire enquiring about externalizing problems: attention deficit and hyperactivity (ADH), oppositional defiant disorder (ODD), conduct problems, and relational aggression, and internalizing problems: anxiety and depression. Teachers rated the child's behavior on a five-point Likert scale ranging from 0 (never applicable) to 4 (often applicable). Cronbach's α in this study varied between .78 and .92. The convergent validity of the PBSI was found to be good as indicated by the correlations between the PBSI and the Teacher's Report Form (Achenbach, 1991), which were .75 for externalizing behavior and .55 for internalizing behavior (Witvliet et al., 2010). This measure was used for all cohorts.

The Social Experience Questionnaire - Teacher Report (SEQ-T) (Crick and Grotpeter 1996) was used to measure relational victimization, physical victimization, and prosocial behavior on a 5-point Likert scale ranging from 0 (never applicable) to 4 (often applicable). The relational victimization and physical victimization scales included three items and the prosocial behavior scale four items. Cronbach's α was .87 (relational victimization), .85 (physical victimization), and .75 (prosocial behavior). This measure was used for all four cohorts.

The twenty-four-item Dimensions of Depression Profile for Children (DDPC) (four-point Likert format) was used to measure the level of depression in the oldest cohort (Harter & Nowakowski, 1987). It contained four subscales:

depressed mood ($\alpha = .69$), self-blame ($\alpha = .59$), low energy/interest ($\alpha = .75$), low global self-worth ($\alpha = .77$), and one total score ($\alpha = .85$).

Social and Emotional Skills

Peer social preference was obtained using peer nominations of like most and like least, as described by Coie, Dodge, and Coppotelli (1982). Children (cohorts 1 and 4) were asked to nominate an unlimited number of classmates that they liked most and that they liked least. Each child could therefore be nominated by each classmate as "liked most," "liked least," or gain no score. For each child, the liked most as well as the liked least nominations were summed and divided by the number of children in the class minus one (self-nomination was not allowed). The standardized liked least score was then subtracted from the standardized liked most score to generate a social preference score. Social preference is generally regarded as a reliable and valid measure of sociometric status (Cillessen & Mayeux, 2004; Rubin et al., 2006).

The teacher-based Preschool and Kindergarten Behavior Scale (PKBS) (Merrel, 1996) used for cohorts 1 and 2 is designed to assess social skills and problem behaviors in children aged three to six years. In this study, the social skills scale, which has three subscales, was used. The social cooperation scale included twelve items reflecting behaviors and characteristics deemed important in following instructions from adults, cooperating and compromising with peers. The social interaction scale included 11 items reflecting behaviors and characteristics deemed important in gaining and maintaining the acceptance and friendship of others. The social independence scale included eleven items reflecting behaviors and characteristics deemed important in achieving social independence within the domain of the peer group. For all items, responses were based on a four-point scale. The internal consistency of all three subscales was high (respectively $\alpha = .89$, $\alpha = .87$, $\alpha = .86$).

The Head Start Competence Scale (HSCS) (Domitrovich, Cortes & Greenberg, 2001) used for cohorts 1 and 2 is a twelve-item measure of children's social and emotional skills reflecting interpersonal relationships and emotion regulation. Teachers were asked to indicate on a four-point scale how

well each item on the scale described the child. Internal consistency was high ($\alpha = .95$).

Children's emotional awareness scores were obtained for cohort 1 and 4 using the Levels of Emotional Awareness Scale for children (LEAS-C), which assesses the complexity of children's emotional awareness (Bajgar et al., 2005). It contains twelve interpersonal scenarios featuring the child and another person. After each scenario had been presented, the children were asked to describe their own feelings and those of the other person. For each scenario, the self and other response was rated on a scale ranging from 0 to 4. The higher of the scores for the self and the other was taken as the total score for each scenario. In cases where both the self and the other score were 4, a total score of 5 was awarded. The total scores were summed across the scenarios. Cronbach's α ranged from .89 to .92 over the assessments.

The child-based Difficulties in Emotion Regulation Scale (DERS) (Gratz & Roemer, 2004) was used to measure deficits of emotion regulation. It contains thirty-six questions (five-point Likert scale) in six scales: non-acceptance of emotional responses ($\alpha = .73$), difficulties engaging in goal-directed behavior ($\alpha = .82$), impulse control difficulties ($\alpha = .80$), lack of emotional awareness ($\alpha = .78$), limited access to emotion regulation strategies ($\alpha = .74$), and lack of emotional clarity ($\alpha = .61$). This measure was used for the oldest cohort only.

To measure the children's affective sharing of others' emotions, a short ten-item version of Bryant's Empathy Index was used (Bryant, 1982; De Wied et al., 2007). This child-based measure was used for the youngest and oldest cohort only. Cronbach's α was .68.

Implementation Quality

Implementation quality was operationalized as "conceptual use of the program" i.e. to what extent do teachers act according to the PATHS basic principles. Teachers received a list of ten questions describing daily classroom situations. For each situation, teachers could choose one of four answers that described how they would react in this specific situation. The answers varied from most desired reaction according to

the PATHS basic principles (score = 4) to least desired reaction (score = 1). All scores were averaged, resulting in a mean score for the first year and a mean score for the second year. For each class, we calculated one mean score (range 1 - 4) from these two scores.

Implementation Quantity

To measure implementation quantity (i.e. completeness), teachers completed a monthly log describing all the required lessons and elements thereof, recording whether they completed each specific element of each lesson. For both the first year and the second year, intervention completeness was assessed as the completed proportion of all prescribed activities for that year. We summed these two proportions to obtain one total score (range 0 - 2).

Covariates

The Peabody Picture Vocabulary Test - Third Edition (PPVT-III) (Dunn & Dunn, 1997) was used to measure verbal ability. This assessment is a well-known and widely used measure of children's receptive vocabulary. A Word Comprehension Quotient (WCQ) score, using age-appropriate norms, was calculated from the raw total number of correct answers. The internal consistency of the PPVT-III standard scores ranged from .92 to .98. Child verbal ability was included as a covariate because of its potential to affect children's performance in the testing situation.

Statistical Analyses

The data were analyzed using Stata version 11.1 (StataCorp, 2009) over all 1,294 students in accordance with the intention-to-treat principle. Missing data were handled through Full Information Maximum Likelihood estimation (FIML). ANOVAs and Chi-square analyses were conducted to check whether there was a balanced distribution of student characteristics and outcome variable values across the two conditions at baseline ($\rho < .05$).

The data in this study is hierarchically structured, i.e. measurement waves (T0, T1, T2, and T3) are nested within students, students are nested within classes, and classes are nested within schools. In such a hierarchical structure, standard statistical formulas will underestimate the sampling

variance, and consequently lead to significance tests with an inflated alpha level (type 1 error rate). Multilevel models are specifically geared toward the statistical analyses of nested or clustered data (Hox, 2010). In the present study we use multilevel mixed-effects linear regression analyses (Twisk, 2006) to test, for each of the outcome variables, whether changes over time varied as a function of the intervention.

For each outcome variable we analyzed the whole development of each outcome variable over time in one analysis by calculating the three change scores ($T_0 - T_1$, $T_1 - T_2$, and $T_2 - T_3$) for each variable. However, as there were baseline differences between the conditions we had to correct the change scores and used the analysis of covariance combination approach as described by Twisk and De Vente (2008) for that purpose. In this analytical approach, the change between baseline and the first reassessment is corrected for the baseline value by subtracting the individual baseline value from the first individual change score. For computational reasons only, a correction was also made for the remaining two change scores. These three adjusted change scores per variable were included in a longitudinal multilevel mixed-effects linear regression model. The resulting β coefficient represents the intervention effect over the whole period, i.e. from T_0 to T_3 .

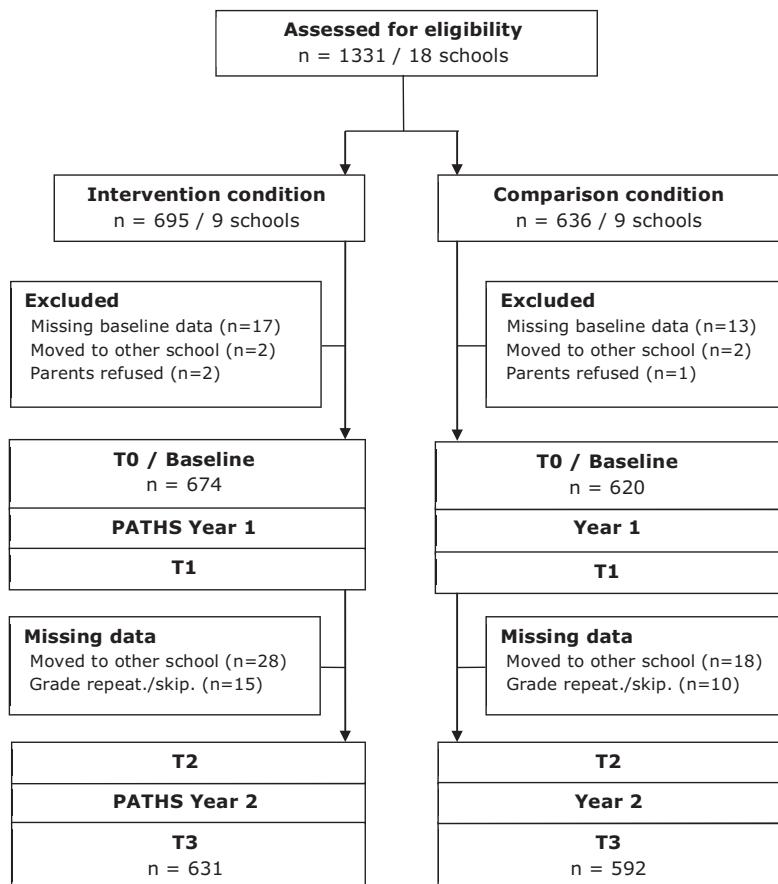
In order to assess short-term effects, the analyses were also performed for the $T_0 - T_1$ period alone (i.e. the models included only the $T_0 - T_1$ adjusted change score). In addition, we tested for possible moderating effects of group, gender, program completeness, and conceptual use by including interaction terms between these variables and condition in the models (each interaction term was tested in separate models). Because of multiple testing (27 outcomes), the level of statistical significance was set at $p < .01$ in all tests.

The Sample

In total, 1,331 children (five to 11 years old) from kindergarten and elementary school grades 1, 3, and 5 were eligible for inclusion (Figure 1). The 18 participating schools were located in rural areas and provincial towns in the western (Noord-Holland) and eastern part (Gelderland

and Overijssel) of the Netherlands. The baseline data for 30 children were missing or incomplete. Four children were excluded because they moved to another school soon after the baseline assessment, and three children's parents refused to allow their children to participate in the study. Therefore the baseline sample included 1,294 children. Of these, 65 changed school during the study. In accordance with the intention-to-treat principle, we sought to collect data from these children by sending a questionnaire to their home address, asking the parents to fill out the forms. We were able to collect data from nineteen of them. We were also able to collect reassessment data for ten of the thirty-five children who repeated or skipped a grade. Participation in our study was high (97%), and the attrition rate low (5% at last assessment).

Figure 1. Participation flowchart



Age, gender, ethnicity, and verbal ability did not differ between the intervention and the comparison group (Table 1). Significant baseline differences ($\rho < .05$) were present with respect to the level of ADH ($F(1,1292) = 10.443, \rho < .001$), ODD ($F(1,1292) = 20.896, \rho < .001$), conduct problems ($F(1,1292) = 30.338, \rho < .001$), relational aggression ($F(1,1292) = 60.891, \rho < .001$), anxiety ($F(1,1293) = 11.400, \rho < .001$), depression ($F(1,1292) = 10.161, \rho < .001$), relational victimization ($F(1,1292) = 45.082, \rho < .001$), physical victimization ($F(1,1292) = 45.594, \rho < .001$), prosocial behavior ($F(1,1292) = 5.098, \rho < .05$), low energy ($F(1,1292) = 4.924, \rho < .05$), social interaction ($F(1,1292) = 4.722, \rho < .05$), social independence ($F(1,1292) = 4.725, \rho < .05$), and social and emotional skills ($F(1,1292) = 10.691, \rho < .01$). In general, the levels of problem behavior were higher in the intervention condition, and the levels of social and emotional skills were lower (Table 2). These differences were statistically corrected for in the analyses.

Table 1. Socio-demographic variables at baseline by cohort

| | | Intervention condition | Comparison condition |
|----------------------------|-------|---------------------------|---------------------------|
| Cohort 1 (Kindergarten) | n | 158 | 166 |
| | Male | 56.3 % | 52.4 % |
| | Age | 5.4 years ($SD = .36$) | 5.5 years ($SD = .31$) |
| | Dutch | 98.6 % | 92.4 % |
| Cohort 2 (Grade 1) | n | 159 | 151 |
| | Male | 52.2 % | 46.4 % |
| | Age | 6.5 years ($SD = .38$) | 6.5 years ($SD = .42$) |
| | Dutch | 96.7 % | 95.1 % |
| Cohort 3 (Grade 3) | n | 173 | 152 |
| | Male | 49.1 % | 56.6 % |
| | Age | 8.6 years ($SD = .53$) | 8.5 years ($SD = .50$) |
| | Dutch | 92.1 % | 93.2 % |
| Cohort 4 (Grade 5) | n | 184 | 151 |
| | Male | 44 % | 47 % |
| | Age | 10.5 years ($SD = .42$) | 10.6 years ($SD = .45$) |
| | Dutch | 94.4 % | 94.5 % |

Results

Table 2 shows the outcome variables over time for both conditions. A decline in the PBSI, SEQ-T (except prosocial behavior scale), DDPC, and DERS scores represents a decrease in these problem behaviors/skills. An increase in the PEER, PKBS, HSCS, LEAS, EMPATHY, and prosocial behavior scale (SEQ-T) scores represents an improvement in these skills.

Table 2. Mean scores (M) and standard deviations (SD) at baseline and reassessments for all four cohorts combined

| Measure | Scale | Intervention condition | | | | | | Comparison condition | | | | | |
|---------|------------------------------|------------------------|---------|-------|-------|-------|-------|----------------------|---------|-------|-------|-------|----|
| | | T0 | | T1 | | T2 | | T3 | | T0 | | T1 | |
| | | M | SD | M | SD | M | SD | M | SD | M | SD | M | SD |
| PBSI | ADH* | 1.26 | (.81) | 1.00 | 1.07 | 1.11 | .98 | 1.11 | (.84) | .97 | 1.10 | 1.07 | |
| | ODD* | 1.11 | (.75) | .96 | .98 | .98 | .91 | .91 | (.79) | .88 | .88 | .89 | |
| | Conduct problems* | .62 | (.55) | .51 | .54 | .56 | .45 | .45 | (.52) | .45 | .41 | .42 | |
| | Relational aggression* | 1.12 | (.73) | .94 | 1.05 | 1.04 | .80 | .80 | (.76) | .82 | .89 | .86 | |
| | Anxiety* | 1.29 | (.71) | 1.10 | 1.20 | 1.15 | 1.16 | 1.16 | (.72) | 1.09 | 1.13 | 1.01 | |
| | Depression* | .92 | (.64) | .86 | .95 | .92 | .80 | .80 | (.69) | .81 | .84 | .82 | |
| | Relational victimization* | .87 | (.69) | .72 | .75 | .79 | .62 | .62 | (.63) | .61 | .62 | .62 | |
| | Physical victimization* | .57 | (.55) | . | .45 | .50 | .37 | .37 | (.50) | . | .31 | .29 | |
| | Prosocial behavior* | 2.68 | (.66) | . | 2.78 | 2.78 | 2.76 | 2.76 | (.66) | . | 2.76 | 2.83 | |
| | Depressed mood | 10.63 | (2.99) | 10.37 | 10.06 | 9.98 | 10.35 | 10.35 | (3.00) | 9.93 | 9.82 | 9.88 | |
| DDPC | Self blame | 15.31 | (2.74) | 14.43 | 14.17 | 14.44 | 15.01 | 15.01 | (3.09) | 14.63 | 14.28 | 14.27 | |
| | Low energy* | 11.08 | (3.46) | 10.68 | 10.45 | 10.53 | 10.25 | 10.25 | (3.25) | 10.40 | 9.78 | 10.19 | |
| | Low self-worth | 10.23 | (3.30) | 9.71 | 9.71 | 9.79 | 9.70 | 9.70 | (3.25) | 9.44 | 9.43 | 9.50 | |
| | Total depression | 47.25 | (9.14) | 45.24 | 44.39 | 44.73 | 45.31 | 45.31 | (8.89) | 44.39 | 43.31 | 43.90 | |
| | Peer nominations | .11 | (.27) | .15 | .18 | .15 | .13 | .13 | (.29) | .17 | .18 | .17 | |
| | Social cooperation | 31.65 | (4.35) | 32.59 | 32.35 | 32.24 | 32.25 | 32.25 | (4.00) | 32.69 | 32.02 | 32.32 | |
| | Social interaction* | 24.45 | (5.02) | 25.76 | 25.44 | 26.07 | 24.54 | 24.54 | (5.22) | 25.46 | 24.79 | 25.55 | |
| HSCS | Social independence* | 27.95 | (4.11) | 29.07 | 28.58 | 28.57 | 28.35 | 28.35 | (4.16) | 28.91 | 28.15 | 28.87 | |
| | Social and emotional skills* | 21.61 | (7.39) | 23.63 | 23.55 | 23.98 | 24.47 | 24.47 | (7.06) | 25.08 | 24.45 | 25.48 | |
| | Emotional awareness | 25.04 | (11.27) | 31.47 | 34.03 | 35.90 | 23.84 | 23.84 | (10.69) | 28.10 | 30.15 | 30.98 | |
| | Lack emotional awareness | 21.31 | (5.67) | 22.50 | 23.60 | 24.51 | 20.93 | 20.93 | (5.50) | 22.49 | 22.20 | 22.43 | |
| | Non-accept. of emot. resp. | 11.91 | (4.46) | 10.86 | 10.14 | 9.91 | 11.61 | 11.61 | (4.63) | 10.50 | 10.04 | 10.7 | |
| | Diff. in goal-directed beh. | 13.19 | (5.14) | 12.73 | 11.87 | 11.30 | 13.19 | 13.19 | (5.33) | 12.13 | 11.54 | 11.51 | |
| | Impulse control difficulties | 13.16 | (5.09) | 12.69 | 12.49 | 11.94 | 13.60 | 13.60 | (5.67) | 12.17 | 11.57 | 11.90 | |
| | Limited access strategies | 17.42 | (5.59) | 17.20 | 15.99 | 15.83 | 17.60 | 17.60 | (5.97) | 15.97 | 15.38 | 15.76 | |
| | Lack of emotional clarity | 15.52 | (4.32) | 14.54 | 14.64 | 14.60 | 15.35 | 15.35 | (4.16) | 14.19 | 14.06 | 14.44 | |
| | EMPATHY | 5.20 | (2.36) | 5.21 | 4.91 | 4.94 | 4.99 | 4.99 | (2.32) | 5.14 | 5.19 | 5.24 | |

* Significant difference between intervention condition and comparison condition at baseline ($p < .05$).¹ To limit the number of questions teachers had to answer each wave, this questionnaire was sent at two instead of three reassessments.

First, intervention effects were examined by longitudinal multilevel mixed-effects linear regression analyses (Table 3). We combined all four cohorts and tested whether changes over time varied as a function of the intervention. This first set of analyses resulted in a positive intervention effect found on the emotional awareness scale. Note that this is the only significant intervention effect out of twenty-seven tests.

Second, we checked whether there was a positive effect in the first year of implementation. We performed the same analyses as above, but limited them to the T0 - T1 adjusted change score. We found no intervention effects on any of the variables in the first year.

Third, we tested whether intervention effects were moderated by grade. We found three significant moderations on the prosocial scale. Effects were stronger for children in kindergarten in the intervention condition and weaker for children in grades 1 and 5 in the intervention condition. The analyses also showed that the effects for the LEAS were stronger for children in kindergarten. Otherwise, grade did not moderate outcomes.

Fourth, we checked whether there was a difference in effect for boys and girls by adding an interaction term (sex x condition). We did not find any gender-related difference.

Fifth, we tested whether intervention effects depended on the level of program completeness. Mean completeness was 50% in the first year ($SD = 23\%$) and 49% in the second year ($SD = 24\%$). We summed these two proportions to obtain one score (mean = .99, $SD = .42$, range .16 - 1.70), and tested whether this score was related to intervention effects. The analyses showed that program completeness did not moderate intervention effects.

Sixth, we tested whether intervention effects depended on implementation quality. The mean level of "conceptual use" was around 3.05 in the first year ($SD = .27$, range 2.30 - 3.75) and 3.07 in the second year ($SD = .37$, range 2.20 - 3.70). We calculated a mean of these two scores for each class and tested whether this score was related to intervention effects. Conceptual use did not moderate intervention effects.

Table 3. Intervention effect (β coefficient) over all four assessments for all four cohorts combined

| Measure | Scale | β | SE | Z | p | 95% CI | |
|---------|-----------------------------|---------------------------------------|-------|-------|---------|--------|-------|
| PBSI | ADH | -.022 | .029 | -0.74 | 0.458 | -.079 | |
| | ODD | -.012 | .028 | -0.42 | 0.675 | -.067 | |
| | Conduct problems | .008 | .020 | 0.41 | 0.681 | -.032 | |
| | Relational aggression | .005 | .030 | 0.18 | 0.857 | -.054 | |
| | Anxiety | .017 | .032 | 0.53 | 0.599 | -.045 | |
| | Depression | .005 | .031 | 0.17 | 0.868 | -.055 | |
| | Relational victimization | .025 | .039 | 0.64 | 0.522 | -.051 | |
| | Physical victimization | .102 | .071 | 1.45 | 0.148 | -.036 | |
| | Prosocial behavior | .074 | .075 | 0.99 | 0.323 | -.073 | |
| | Depressed mood | .016 | .168 | 0.09 | 0.925 | -.314 | |
| SEQ-T | Self blame | .114 | .173 | 0.66 | 0.509 | -.225 | |
| | Low energy | -.032 | .182 | -0.18 | 0.860 | -.388 | |
| | Low self-worth | .090 | .188 | 0.48 | 0.632 | -.279 | |
| | Total depression | .104 | .473 | 0.22 | 0.825 | -.823 | |
| | Peer nominations | .004 | .012 | 0.33 | 0.738 | -.019 | |
| | Social cooperation | -.065 | .174 | -0.37 | 0.710 | -.280 | |
| | Social interaction | .466 | .295 | 1.58 | 0.115 | -.113 | |
| | Social independence | .187 | .276 | 0.68 | 0.498 | -.353 | |
| | Social and emotional skills | .111 | .328 | 0.34 | 0.736 | -.532 | |
| | Emotional awareness | .186 | .393 | 3.02 | 0.003 * | .416 | |
| PEER | LERS | Lack emotional awareness | .694 | .305 | 2.27 | 0.023 | |
| | DERS | Non-acceptance of emotional responses | -.059 | .245 | -0.24 | 0.809 | .095 |
| | | Difficulty in goal-directed behavior | -.139 | .277 | -0.50 | 0.616 | -.539 |
| | | Impulse control difficulties | .135 | .297 | 0.46 | 0.648 | .421 |
| | | Limited access strategies | .123 | .332 | 0.37 | 0.712 | -.683 |
| | | Lack of emotional clarity | .141 | .298 | 0.47 | 0.635 | .404 |
| | EMPATHY | Empathy | -.200 | .114 | -1.76 | 0.079 | -.446 |
| | | | | | | | .717 |

* Significant difference between intervention condition and comparison condition ($p < .01$).

Note: A minus sign (-) before the β of the PBSI, SEQ-T (except prosocial behavior scale), DDPC, and DERS represents a decrease in these problem behaviors/skills in the intervention condition over time, relative to the control condition.

Discussion

The purpose of this study was to examine the effectiveness of the efficacious school-based prevention program PATHS when implemented through Dutch municipal health services. In contrast to other studies, the intervention was implemented by health promotion professionals from a regular health service system rather than dedicated PATHS trainers. As health promotion professionals implement school-based prevention programs on a regular basis, this linkage group could in the long run potentially support the national licensee in the dissemination of PATHS. However, virtually no intervention effects were found in this study.

It seems unlikely that the underlying theoretical model of the PATHS intervention can explain the lack of intervention effects. Previous research has demonstrated that PATHS is effective when delivered adequately. The implementation process therefore seems a more likely reason. In the present study, there appears to be a large gap between the intended intervention and the intervention participants actually received. The implementation strategy most probably affected the teacher support negatively. This resulted in low program completeness and probably affected some other implementation variables negatively as well. This finally resulted in poor intervention outcomes.

Although we cannot exactly identify the (implementation) variables that moderated the outcomes, we can conclude that the tested PATHS implementation strategy is not a recipe for effective prevention of problem behavior on a large scale.

Strengths and Limitations

This is the first study to test the effectiveness of PATHS when implemented through a regular health service, and sought to explore how problem behavior could be prevented on a large scale. Furthermore, this study had very high participation (97%) and low attrition (5%). We also had a fairly large sample that gave us enough power to detect small effects, even for the four grades/cohorts separately. It must further be recognized that the results are fairly comprehensive, and the measures

included constructs of social and emotional skills and problem behavior of both teachers and children, on four occasions.

However, the study also suffers from important limitations. First, we did not randomize all schools. This may have caused baseline differences between the two conditions. However, we corrected for these differences with a thorough statistical technique specially developed for this kind of problem (Twisk & De Vente, 2008). A second limitation is the limited duration of implementation. As PATHS is an eight-year curriculum, two years of implementation may have been insufficient to achieve the desired change. However, previous studies were able to report positive outcomes within a similar timespan. Third, instead of using independent observations, self-reporting was used to assess teachers' quality and quantity of program implementation. Self-reports may have led to an overestimation of the quantity and quality of implementation. Besides, our number of implementation measures was limited. Fourth, although we did not monitor this, the lack of effects could have been caused by the level of standard social and character development activities that were given in the control group, for example by other preventive programs. Nonetheless, the practical value of PATHS would have been reflected in additional effects on top of care as usual, and this is exactly what we studied.

Conclusions

It seems unfortunate that the combination of an efficacious prevention program and a health service system specifically designed to be a linking agent for the implementation of school-based health promotion interventions does not provide better results – especially as universal prevention programs can only have population impact on the prevention of problem behavior when they are implemented on a large scale.

One could question whether this implementation strategy was the best possible. Obviously, the implementation strategy could have been enhanced with greater levels of support for both health promotion professionals and teachers. However, our aim was not to study the effectiveness of an intervention using a theoretically optimal

implementation strategy, but rather to study its effectiveness when employing an implementation strategy that has been used successfully before, but, even more importantly, is commonly used and achievable within the Dutch public health sector. Along with efficacy trials, this kind of study helps us to identify interventions and implementation strategies that could be helpful (or not) in preventing problem behavior.

In addition, we believe that the present findings are probably not just relevant for PATHS or limited to prevention in the realm of social behavior. Implementation seems likely to play a key role in establishing societal impact. Our study underlines the importance of studying the process of transferring outcomes of efficacy studies to the more naturalistic settings for program implementation, and of monitoring program application in different countries and settings. More, well-designed large-scale field trials are urgently needed to provide policymakers with realistic estimates of the investments required to obtain intervention effects that are replicable at population level.

Acknowledgements

This study was funded by grant number 2430.0031 from the Dutch Health Care Research Organization (ZonMw). Furthermore, we are grateful to all who participated in this study (health promotion professionals, teachers, children, and their parents).

References

- Achenbach, Thomas M. 1991. *Manual for the Teacher's Report Form and 1991 Profiles*. Burlington: University of Vermont, Department of Psychiatry.
- Bajgar, Jane, Joseph Ciarrochi, Richard Lane, and Frank P. Deane. 2005. Development of the Levels of Emotional Awareness Scale for Children (LEAS-C). *British Journal of Developmental Psychology* 23:569-86. doi:10.1348/026151005X35417
- Beelmann, Andreas, and Tobias Raabe. 2009. The Effects of Preventing Antisocial Behavior and Crime in Childhood and Adolescence: Results and Implications of Research Reviews and Meta-Analyses. *European Journal of Developmental Science* 3:260-81. doi:10.3233/DEV-2009-3305
- Bryant, Brenda R. 1982. An Index of Empathy for Children and Adolescents. *Child Development* 53 (2): 413-25. doi:10.2307/1128984
- Cillessen, Antonius H. N., and Lara Mayeux. 2004. Sociometric Status and Peer Group Behavior: Previous Findings and Current Directions. In *Children's Peer Relations: From Development to Intervention*, ed. Janis B. Kupersmidt and Kenneth A. Dodge, 3-20.. Washington, DC: American Psychological Association.
- Coie, John D., Kenneth A. Dodge, and Heide Coppotelli. 1982. Dimensions and Types of Social Status: A Cross-Age Perspective. *Developmental Psychology* 18:557-70.
- CPPRG (Conduct Problems Prevention Research Group). 1999. Initial Impact of the Fast Track Prevention Trial for Conduct Problems: II. Classroom Effects. *Journal of Consulting and Clinical Psychology* 67:648-57. doi: 10.1037/0022-006X.67.5.648
- Crick, Nicki R., and Jennifer R. Grotjeter. 1996. Children's Treatment by Peers: Victims of Relational and Overt Aggression. *Development and Psychopathology* 8:367-80. doi:10.1017/S0954579400007148
- De Wied, Minet, Cora Maas, Stephanie Van Goozen, Marjolijn Vermande, Rutger Engels, Wim Meeus, Walter Matthys, and Paul Goudena. 2007. Bryant's Empathy Index: A Closer Examination of its Internal Structure. *European Journal of Psychological Assessment* 23:99-104. doi:10.1027/1015-5759.23.2.99
- Dodge, Kenneth A. 2009. Community Intervention and Public Policy in the Prevention of Antisocial Behaviour. *Journal of Child Psychology and Psychiatry* 50:194-200. doi:10.1111/j.1469-7610.2008.01985.x
- Domitrovich, Celene E., Rebecca C. Cortes, and Mark T. Greenberg. 2001. Head Start Competence Scale Technical Report. Unpublished manuscript, Pennsylvania State University.
- Domitrovich, Celene E., Rebecca C. Cortes, and Mark T. Greenberg. 2007. Improving Young Children's Social and Emotional Competence: A Randomized Trial of the Preschool "PATHS" Curriculum. *Journal of Primary Prevention* 28 (2): 67-91. doi:10.1007/s10935-007-0081-0
- Domitrovich, Celene E., Mark T. Greenberg, Rebecca C. Cortes, and Carol Kusché. 1999. *Manual for the Preschool PATHS Curriculum*. South Deerfield, MA: Channing-Bete.
- Dunn, Lloyd M., and Leota M. Dunn. 1997. *Peabody Picture Vocabulary Test - Revised*. Circle Pines, MN: American Guidance Service.
- Durlak, Joseph A., Roger P. Weissberg, Allison B. Dymnicki, Rebecca D. Taylor, and Kriston B. Schellinger. 2011. The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child Development* 82:405-32. doi:10.1111/j.1467-8624.2010.01564.x
- Eisner, Manuel. 2009. No Effects in Independent Prevention Trials: Can We Reject The Cynical View? *Journal of Experimental Criminology* 5:163-83. doi:10.1007/s11292-009-9071-y
- Elliott, Delbert S., and Sharon Mihalic. 2004. Issues in Disseminating and Replicating Effective Prevention Programs. *Prevention Science* 5:47-53. doi:10.1023/B:PREV.0000013981.28071.52
- ErasmusMC. 2000. *Problem Behaviour at School Interview*. Rotterdam: Department of Child and Adolescent Psychiatry.
- Flay, Brian R., Anthony Biglan, Robert F. Boruch, Felipe González Castro, Denise Gottfredson, Sheppard Kellam, Eve K. Mościcki, Steven Schinke, Jeffrey C. Valentine, and Peter Ji. 2005. Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination. *Prevention Science* 6:151-75. doi:10.1007/s11121-005-5553-y
- Gratz, Kim L., Lizabeth Roemer. 2004. Multidimensional Assessment of Emotion Regulation and Dysregulation: Development, Factor Structure and Initial Validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioural Assessment* 26:41-54. doi:10.1023/B:JOBA.0000007455.08539.94
- Greenberg, Mark T., and Carol A. Kusché. 1993. *Promoting Social and Emotional Development in Deaf Children: The PATHS Project*. Seattle, WA: University of Washington Press.

- Greenberg, Mark T., Carol A. Kusché, Elizabeth T. Cook, and Julie P. Quamm. 1995. Promoting Emotional Competence in School-Aged Children: The Effects of the PATHS Curriculum. *Development and Psychopathology* 7:117-136. doi:10.1017/S0954579400006374
- Harter, Susan, and M. Nowakowski. 1987. *Manual for the Dimensions of Depression Profile for Children and Adolescents*. Denver, CO: University of Denver, Department of Psychology.
- Hox, Joop J. 2010. Multilevel Analysis. Techniques and Applications, 2d ed. New York: Routledge.
- Kam, Chi-Ming, Mark T. Greenberg, and Carla T. Walls. 2003. Examining the Role of Implementation Quality in School-Based Prevention using the PATHS Curriculum. *Prevention Science* 4:55-63. doi:10.1023/A:1021786811186
- Rusché, Carol A., and Mark T. Greenberg. 1994. *The PATHS Curriculum*. South Deerfield, MA: Channing-Bete.
- Louwe, Jos J., Cornelis W. van Overveld, Welmoet Merk, Bram Orobio de Castro, and Willem Hoops. 2007a. De invloed van het Programma Alternatieve Denkstrategieën op reactieve en proactieve agressie bij jongens in het primair onderwijs: Effecten na één jaar. *Pedagogische Studiën* 84:277-92.
- Louwe, Jos J., Cornelis W. van Overveld, Welmoet Merk, Bram Orobio de Castro, and Willem Hoops. 2007b. Het PAD van de school en de leraar. In *Jaarboek Lectoraten Faculteit Educatie* 2007. Utrecht: Hogeschool Utrecht.
- Malti, Tina, Denis Ribeaud, and Manuel Eisner. 2011. The Effectiveness of Two Universal Preventive Interventions in Reducing Children's Externalizing Behaviour: A Cluster Randomized Controlled Trial. *Journal of Clinical Child and Adolescent Psychology* 40 (5): 677-92. doi:10.1080/15374416.2011.597084
- Merrell, Kenneth W. 1996. Socio-Emotional Assessment in Early Childhood: The Preschool and Kindergarten Behaviour Scales. *Journal of Early Intervention* 20:132-45. doi:10.1177/105381519602000205
- Riggs, Nathaniel R., Mark T. Greenberg, Carol A. Kusché, and Mary Ann Pentz. 2006. The Mediational Role of Neurocognition in the Behavioural Outcomes of a Social-Emotional Prevention Program in Elementary School Students: Effects of the PATHS Curriculum. *Prevention Science* 7:91-102. doi:10.1007/s11121-005-0022-1
- Rubin, Kenneth H., Julie C. Wojslawowicz, Linda Rose-Krasnor, Cathryn Booth-LaForce, and Kim B Burgess. 2006. The Best Friendships of Shy/Withdrawn Children: Prevalence, Stability, and Relationship Quality. *Journal of Abnormal Child Psychology* 34 (2): 143-57. doi:10.1007/s10802-005-9017-4
- Salmivalli, Christina, Ari Kaukainen, and Marinus Voeten. 2005. Anti-Bullying Intervention: Implementation and Outcome. *British Journal of Educational Psychology* 75:465-87. doi:10.1348/000709905X26011
- SCDRC (Social and Character Development Research Consortium). 2010. *Efficacy of School Wide Programs to Promote Social and Character Development and Reduce Problem Behaviour in Elementary School Children (NCER 2011-2001)*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.
- StataCorp. 2009. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.
- SVO (Seminarium voor Orthopedagogiek). 2005. *PAD, Programma Alternatieve Denkstrategieën, Lessen voor sociale en emotionele vaardigheden in de basisschool*. Utrecht: Trimbos Institute.
- Twisk, Jos W. R. 2006. *Applied Multilevel Analysis*. Cambridge: Cambridge University Press.
- Twisk, Jos W. R., and Wieke De Vente. 2008. The Analysis of Randomised Controlled Trial Data with More than One Follow-Up Measurement. A Comparison Between Different Approaches. *European Journal of Epidemiology* 23:655-60. doi:10.1007/s10654-008-9279-6
- Weisz, John R., Geri R. Donenberg, Susan S. Han, and Danika Kauneckis. 1995. Child and Adolescent Psychotherapy Outcomes in Experiments versus Clinics: Why the Disparity? *Journal of Abnormal Child Psychology* 23:83-106. doi:10.1007/BF01447046
- Wiefferink, Karin H., Jos Poelman, M. Linthorst, Ine Vanwesenbeeck, J. C. M. van Wijngaarden, and Theo G. W. Paulussen. 2005. Outcomes of a Systematically Designed Strategy for the Implementation of Sex Education in Dutch Secondary Schools. *Health Education Research* 20 (3): 323-33. doi:10.1093/her/cyg120
- Wilson, Sandra J., and Mark W. Lipsey. 2007. School-Based Interventions for Aggressive and Disruptive Behavior. Update of a Meta-Analysis. *American Journal of Preventive Medicine* 33 (2S): S130-43. doi:10.1016/j.amepre.2007.04.011
- Witvliet, Miranda, Pol A. C. van Lier, Pim Cuijpers, and Hans M. Koot. 2010. Change and Stability in Childhood Clique Membership, Isolation from Cliques, and Associated Child Characteristics. *Journal of Clinical Child and Adolescent Psychology* 39:12-24. doi:10.1080/15374410903401161

3

Effectiveness of a brief school based intervention on depression, anxiety, hyperactivity, and delinquency: a cluster randomized controlled trial

Ferry Goossens
Jeroen Lammers
Simone Onrust
Patricia Conrod
Bram Orobio de Castro
Karin Monshouwer

Published as:

Goossens, F.X., Lammers, J., Onrust, S.A., Conrod, P.J., Orobio de Castro, B., & Monshouwer, K. (2016). Effectiveness of a Brief School Based Intervention on Depression, Anxiety, Hyperactivity and Delinquency: A Cluster Randomized Controlled Trial. *European Child & Adolescent Psychiatry*, 25, 639-648.

Authors' contributions:

FG carried out the study (together with JL), analyzed the data (together with JL and SO), and wrote the manuscript. JL, SO, BODC, and KM revised the manuscript. PC contributed to the design of the study, supervised the implementation, developed the original intervention materials, and revised the manuscript.

Conflict of interest:

FG, JL, SO, BODC, and KM declare that they have no conflict of interest. PC is the developer and licensee of Preventure.

Abstract

Problematic substance use and mental health problems often co-occur in adolescents. Effective schoolbased interventions that are brief and target multiple problems are promising in the field of health promotion. Preventure is a brief, school-based, selective preventive intervention, tailored to four personality profiles. Preventure has already proved effective on alcohol outcomes. Previous trials also reveal effects on several mental health outcomes, yet the evidence for these outcomes is limited. This study presents the results of the Dutch Preventure Trial, on a range of mental health outcomes. In a cluster RCT, including 699 high risk students (mean age 14 years), the intervention effects on mental health problems at 2, 6, and 12 months post intervention were tested in the total high risk population and in four specific personality groups. No significant intervention effects were found on 22 from the 24 tests. A positive intervention effect on anxiety was found in the anxiety sensitivity personality group at 12-month follow-up, and a negative intervention effect on depression was found at 12-month follow-up in the negative thinking group. In post hoc growth curve analyses these effects were not found. This study found no convincing evidence for the effectiveness of Preventure in the Netherlands on mental health problems. This finding is not in line with the results of an earlier effectiveness study in the UK. This highlights the need for more research into the knowledge transfer model of interventions, to ensure that interventions are effective in a variety of circumstances.

Introduction

Problematic substance use and mental health problems often co-occur in adolescents (De Graaf, Bijl, Smit, Vollebergh & Spijker, 2002; Kessler, 2004). Studies have shown that an important part of this co-occurrence is explained by an overlap in risk factors (Kendler, Prescott, Myers & Neale, 2003). Consequently, if these risk factors are causally related to both types of problems, effective preventive interventions targeting these risk factors are likely to reduce both substance use and mental health problems. Indeed, the school-based preventive interventions Preventure (counsellor based) and Adventure (teacher based), which were originally designed to prevent problematic alcohol use, have been shown to reduce not only risky drinking behaviors (Conrod, Stewart, Comeau & Maclean, 2006; Conrod, Castellanos & Mackie, 2008; Conrod, Castellanos & Mackie, 2011; O'Leary-Barrett, Mackie, Castellanos-Ryan, Al-Khudhairy & Conrod, 2010) and illicit drug use (Conrod, Castellanos-Ryan & Strang, 2010), but also a variety of mental health problems, including depression, anxiety, and a number of conduct disorders (Castellanos & Conrod, 2006; O'Leary-Barrett et al., 2013).

Preventure and Adventure use a personality-targeted approach, based on the theory that personality is an important construct for understanding and preventing adolescents' alcohol use and abuse (Comeau, Stewart & Loba, 2001; Cooper, Frone, Russel & Mudar, 1995). Both interventions, discern four personality groups which receive interventions tailored to their personalities. These groups are high levels of negative thinking (NT), anxiety sensitivity (AS), sensation seeking (SS), and low levels of impulse control (IMP) (Conrod, Stewart, Comeau & Maclean, 2006). These personality profiles have been identified as risk factors not only for substance abuse (Comeau, Stewart & Loba, 2001; Cooper, Frone, Russel & Mudar, 1995; Blackwell, Conrod & Hansen, 2002; Conrod & Woicik, 2002; Malmborg et al., 2010; Pulkkinen & Pitkänen, 1994; Shall, Kemeny & Matlzman, 1992; Stewart, Peterson & Pihl, 1995), but also for mental health problems like depression (Joiner, 2000), anxiety (Maller & Reiss, 1992), hyperactivity (Moeller, Barratt, Dougherty, Schmitz & Swann, 2001), and reckless behavior (Arnett, 1994).

The effects of Preventure and Adventure on risky drinking behaviors have been studied extensively, mostly in Canadian and English samples of adolescents. Studies show a variety of 'condition' and 'condition x personality risk group' effects on risky drinking behaviors up to 24 months post intervention (Conrod, Stewart, Comeau & Maclean, 2006; Conrod, Castellanos & Mackie, 2008; Conrod, Castellanos & Mackie, 2011; O'Leary-Barrett, Mackie, Castellanos-Ryan, Al-Khudhairy & Conrod, 2010). The effectiveness of Preventure on the primary alcohol related outcomes is currently being tested in the Netherlands, using the same sample as used for this paper (Lammers et al., 2011). Binge drinking rates did not significantly differ between the intervention and control group at 12 months follow-up. Intention-to-treat analyses revealed no significant intervention effects on alcohol use and problem drinking at 12 months follow-up. However, post-hoc latent-growth analyses revealed significant effects on the development of binge drinking, and binge drinking frequency (Lammers et al., 2015).

To date, the effectiveness of Preventure on mental health problems has been studied in one trial in London (Castellanos & Conrod, 2006). This study among 423 students found a moderate intervention effect in the NT group on depression scores at 6-month follow-up and a similar effect in the AS group on panic attacks and truancy. A small intervention effect was found for shoplifting in the total high risk sample, as well as a moderate effect on this outcome in the IMP group. Yet, the results are limited to short-term mental health outcomes (i.e. at 6 months follow-up). Analysis of the secondary mental health outcomes of the Adventure program at 24 months follow-up showed that the intervention had generic effects in reducing depression and anxiety symptoms across all high risk youth, and personality-specific effects in reducing conduct problems in impulsive youth (O'Leary-Barrett et al., 2013).

Aims

The aim of the present study is to test the effectiveness of Preventure in the Netherlands on a range of mental health outcomes at 2, 6, and 12 months post intervention. Based on previous finding (Castellanos & Conrod, 2006; O'Leary-Barrett et al., 2013) and the theoretical underpinnings

of Prevention, we expect a reduction of depression, anxiety, conduct problems and hyperactivity rates, and delinquent risk taking behavior in the total intervention group as compared to the control group. Furthermore, we expect these effects to be moderated by personality, i.e. reductions in depression rates are strongest in the NT group, anxiety rates in the AS group, and hyperactivity rates, conduct problems and delinquent risk-taking behavior in both the IMP and SS group. Results of the primary alcohol related outcomes are published yet (Lammers et al., 2015) and the remaining (secondary) outcomes (drinking motives, smoking behavior, and marihuana consumption) will be published in future, separate journal articles.

Method

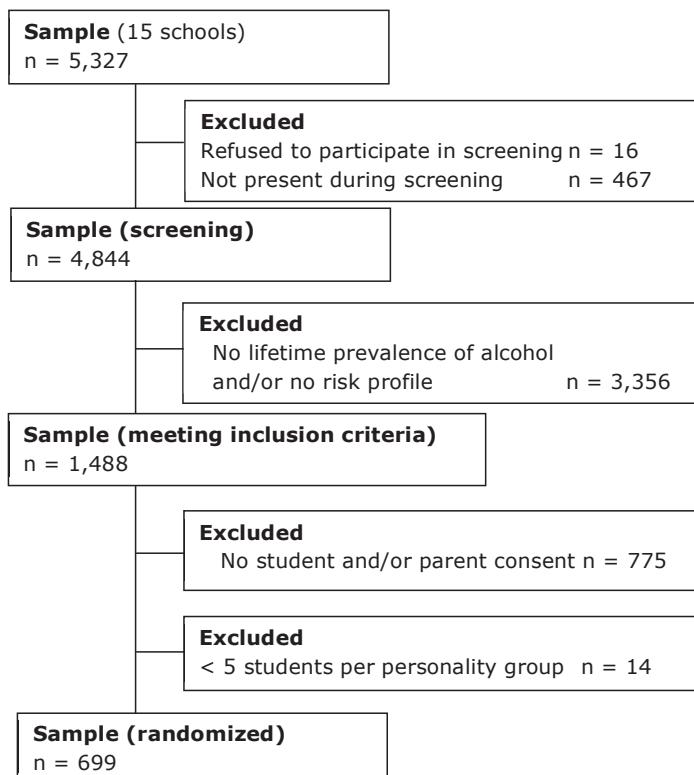
This study represents an analysis of the mental health secondary outcomes of a cluster randomized controlled trial with an intervention and a control condition, testing the intervention effects on mental health symptoms at 2, 6, and 12 months after the delivery of the intervention. The full study design is described in a study protocol (Lammers et al., 2011) and the trial is registered in the Netherlands Trial Register (NTR1920).

Participants

A screening survey among all students attending grades 8 and 9 at each of the 15 participating schools was conducted in order to assess whether students were eligible to enter the trial. The inclusion criteria were defined as follows: (1) lifetime use of at least one glass of alcohol, (2) scoring at least one standard deviation above the sample mean on one of the four personality risk scales (AS, SS, NT, or IMP) of the Substance Use Risk Profile Scale (SURPS) (Woicik, Stewart, Phil & Conrod, 2009), and (3) attending a school where at least five students per personality risk group were eligible and willing to be included in the intervention condition (in order to create workshops with sufficient interaction between students).

From the 4,844 adolescents who completed the screening questionnaire, 1,488 (31%) met the first two inclusion criteria. If a student scored high on more than one subscale, he or she was assigned to the personality group in which he or she showed the largest statistical deviation with respect to the z-scores. Of those 1,488 respondents, 713 (48%) were willing to participate (active consent of both parent and student). However, 14 students were excluded because at their (intervention or control) school the number per personality risk group was too small (i.e. fewer than five students per workshop). This procedure resulted in 699 students (mean age at baseline 14 years) (see Figure 1).

Figure 1. Participant flowchart



| Intervention condition (7 schools) | | | | |
|---|-----|-----|-----|-----|
| n = 343 (100%) | | | | |
| | NT | AS | IMP | SS |
| n | 99 | 66 | 80 | 98 |
| male % | 34% | 21% | 55% | 71% |
| Present at T1 (2 months) | | | | |
| n = 283 (83%) | | | | |
| | NT | AS | IMP | SS |
| n | 78 | 58 | 63 | 84 |
| Present at T2 (6 months) | | | | |
| n = 263 (78%) | | | | |
| | NT | AS | IMP | SS |
| n | 70 | 57 | 58 | 78 |
| Present at T3 (12 months) | | | | |
| n = 246 (72%) | | | | |
| | NT | AS | IMP | SS |
| n | 66 | 51 | 52 | 77 |

| Control condition (8 schools) | | | | |
|--------------------------------------|-----|-----|-----|-----|
| n = 356 (100%) | | | | |
| | NT | AS | IMP | SS |
| n | 87 | 56 | 96 | 117 |
| male | 32% | 32% | 65% | 78% |
| Present at T1 (2 months) | | | | |
| n = 297 (84%) | | | | |
| | NT | AS | IMP | SS |
| n | 68 | 53 | 80 | 96 |
| Present at T2 (6 months) | | | | |
| n = 289 (81%) | | | | |
| | NT | AS | IMP | SS |
| n | 72 | 45 | 80 | 92 |
| Present at T3 (12 months) | | | | |
| n = 284 (80%) | | | | |
| | NT | AS | IMP | SS |
| n | 69 | 49 | 78 | 88 |

Consent

Parents were informed of the study (screening and intervention) through a letter sent by the school to the home address asking them to contact the researchers by phone or e-mail if they did not wish their child to participate in the study (passive informed consent). Parents were told that the intervention was a coping-skill training designed to reduce adolescent risk taking, with alcohol abuse as an example. To assure participants' confidentiality, parents were not explicitly informed about any of the selection variables of the study. On the day of the screening, students were given information on the screening, the ethical issues (confidentiality and the voluntary nature of participation), and the intervention. Parents and students provided active informed consent to participate in the intervention part of the study.

The study was approved by the Medical Ethical Commission for Mental Health (METIGG), i.e. the study was evaluated by the METIGG, which considered the study did not fall within the WMO Act (Medical Research Involving Human Subject Act). However, for the consent procedure, we adhered to the guidelines and recommendations of the METIGG.

Randomization

After the screening and consent procedure was completed, the schools were randomized to one of the two conditions (experimental or control). This occurred at school and not class or individual level to avoid contamination between conditions. A randomization scheme was used, stratified by level of education provided by the school (only lower secondary education; only higher secondary education; both lower and higher secondary education) and, within these three strata, by school size (50% largest schools; 50% smallest schools).

Intervention

Intervention condition

The intervention involved two group sessions, carried out at the participants' schools. The group sessions were tailored to one of the four personality profiles, so there were four different groups of two sessions each. Both group sessions lasted 90 minutes and were spread across two weeks. The intervention used student manuals. The original student manuals, developed in Canada, were translated and adapted to the Dutch cultural and school context. The theoretical basis of Prevention and the content of the intervention are described in the study protocol paper (Lammers et al., 2011).

Control condition

Students assigned to the control group received no further intervention.

Treatment integrity

The intervention was provided by qualified counsellors and co-facilitators. The three counsellors and two co-facilitators attended a 2-day training session led by Dr P.J. Conrod and Dr N. Castellanos from King's College, London, who developed the original intervention. Furthermore, all counsellors had practiced the two group sessions at a pilot school with supervision and feedback. These supervised interventions were run with students from a pilot school, not recruited for the Prevention trial. Also, each counsellor's first two group sessions were observed by a supervisor who had participated in the Prevention training session. All

the counsellors were provided with feedback during four peer reviewing meetings under the guidance of the same supervisor.

At the first workshop, 80% of participants were present, and at the second workshop 71%. In total, 71% of the students attended both workshops. Students who did not attend both workshops were more likely to have recently been binge drinking (59% vs. 45%) ($\chi^2(1) = 5.12, p < .024$) and were more likely to skip one or more of the follow-up measurements ($\chi^2(1) = 25.87, p < .0001$) than students who attended both workshops.

Data collection

The data were collected by means of an online questionnaire asking for demographics, personality risk profile, drinking behaviors, and outcomes. The data collection took place during a regular lesson (approximately 50 minutes), and questionnaires were administered by a trained research assistant from the Trimbos Institute. The baseline/screening survey was collected 2 to 6 months before the intervention started. Students assigned to the experimental or control condition received the follow-up questionnaires at 2, 6, and 12 months after completion of the intervention period.

The response on the 2- (83%), 6- (79%), and 12-month (76%) follow-up measurement was high. Those who only completed the screening questionnaire (7% of all 699 respondents) were more likely to attend a lower-level educational school than those who completed at least one of the three follow-up questionnaires (53% vs. 34%, $\chi^2(1) = 8.20, p < .004$).

Measures

Outcomes

Depression

The level of depression was measured with the Dutch version of the widely used 20-item Centre for Epidemiological Studies Depression Scale (CES-D) (Bouma, Ranchor, Sanderman & Van Sonderen, 1995). The CES-D scale ranges from 0 to 60, with higher scores indicating a higher level of

symptoms. Example items are: "During the last week, I felt hopeful about the future," "During the last week, I thought my life had been a failure." The measure showed a high internal consistency ($\alpha = .80$).

Anxiety

The Childhood Anxiety Sensitivity Index (CASI) (Silverman, Fleisig, Rabian & Peterson, 1991; Van Widenfelt, Siebelink, Goedhart & Treffers, 2002) is a self-report questionnaire to assess children and adolescents' fear of anxiety symptoms. Silverman et al. (1991) constructed the CASI as a developmentally sensitive version of the adult Anxiety Sensitivity Index (Reiss, Peterson, Gursky & McNally, 1986) for younger test-takers. Respondents are asked to rate their degree of agreement with each of 18 items (e.g. "It scares me when I feel nervous") on a 3-point Likert scale ranging from 0 (*none*) to 2 (*a lot*). The CASI showed acceptable two-week test-retest reliability and good criterion-related validity in clinical and nonclinical samples of children and teens (Walsh, Stewart, McLaughlin & Comeau, 2004; Weems, Hammond-Laurence, Silverman & Ginsburg, 1998). The internal consistency in the present sample was high ($\alpha = .87$).

Hyperactivity and conduct problems

The Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997) is a behavioral screening questionnaire measuring strengths and difficult behavior in adolescents. The SDQ yields five scale scores (emotional symptoms, conduct problems, hyperactivity, peer problems, and prosocial behavior), each consisting of five items and a scale score from 0 to 10. The SDQ has been shown to be highly correlated with other well-established measurement instruments, including the Child Behavior Checklist (Goodman & Scott, 1999). In this study the conduct scale and the hyperactivity scale were used. The internal consistency of the hyperactivity scale was acceptable ($\alpha = .68$). As the instrument elicits information on the previous six months, hyperactivity was only assessed at 6 and 12 months post intervention. The internal consistency of the conduct problems scale was too low ($\alpha = .50$) and therefore it was decided not to use the data of this scale for further analyses.

Delinquent behavior

To assess delinquent behavior, we used a 14-item self-report measure (Baerveldt, 1992). For each of the different types of delinquent behavior (e.g. shoplifting, stealing a bicycle or a moped, vandalizing a bus/train, carrying a knife or weapon), respondents were asked to indicate whether they had committed these offenses in the previous six months. A positive answer was scored a 1. By summing up the item scores, a total score was calculated, ranging from 0 to 14. As the instrument elicits information on the previous six months, delinquent behavior was only assessed at 6 and 12 months post intervention.

Screening

Lifetime prevalence of alcohol use

The lifetime prevalence of alcohol use was measured with information on the number of times respondents had used alcohol in their lives (Hibell et al, 2009; O'Malley, Bachmann & Johnston, 1983). Those answering once or more were classified as alcohol users.

Personality risk profile

The SURPS (Woicik, Stewart, Phil & Conrod, 2009) consists of 23 non-overlapping items that assist in discriminating personality dimensions independent of substance use behavior. Items are scored on a 4-point scale, 1 = *strongly disagree*, 2 = *disagree*, 3 = *agree*, 4 = *strongly agree*. The personality dimension Negative Thinking (seven items) refers to hopelessness, which might lead to depressive symptoms. A sample item on the Negative Thinking subscale is "I feel that I'm a failure." The Anxiety Sensitivity dimension (five items) measures fear of bodily sensations. The Sensation Seeking subscale (six items) measures the tendency to seek out thrilling experiences, and the tendency to act without thinking is measured by the Impulsivity scale (five items). In the present study, for each of the four personality risk scales, respondents scoring at least one standard deviation above the sample mean were coded as 1 on the corresponding risk scale, all others were coded as 0. Studies in both adolescent and adult samples in several countries, including the Netherlands, have shown that this scale has good internal reliability,

good convergent and discriminant validity, and adequate test-retest reliability (Conrod & Woicik, 2002; Malmberg et al, 2010; Woicik, Stewart, Phil & Conrod, 2009; Krank, Stewart, Wall, Woicik & Conrod, 2011; Wiers, Ames, Hofmann, Krank & Stacey, 2010). The instrument has been successfully applied in the Netherlands (Malmberg et al, 2010). In the present sample, the internal consistency of the subscales appeared to be adequate (NT: $\alpha = .84$, AS: $\alpha = .72$, IMP: $\alpha = .69$, SS: $\alpha = .66$).

Demographics

A questionnaire assessing demographic information asked participants to report on their age, sex, and ethnic background. The child's ethnicity was determined by the mother's country of birth. If the mother was born in the Netherlands, the child's ethnicity was determined by the father's country of birth. The educational level was classified as low = (1) including the three lowest levels of lower secondary professional education or high = (0) including the highest level of lower secondary professional education or higher general secondary education or pre-university education.

Recent binge drinking

Recent binge drinking was defined as drinking at least five alcoholic drinks at one occasion in the last month (Hibell et al, 2009; O'Malley, Bachmann & Johnston, 1983). Those answering once or more were classified as recent binge drinkers.

Power

The sample size was based on the primary outcomes of this Prevention study; a reduction in the percentage of students engaging in binge drinking. The first step in that power calculation resulted in 183 students in each condition to test the hypothesis in a 2-sided test at alpha = 0.05 and a power of (1-beta) = 0.80. Because of the loss of power due to randomization of schools and the increase in error because of applying a multiple imputation procedure to fill in missing values, 256 respondents per condition (intervention and control) needed to be included at baseline (see Lammers et al. (2011) for the full calculation). Simple post-hoc analyses showed that with 183 students an effect size of 0.33 can be

determined for the social emotional outcomes in a 2-sided test at alpha = 0.05 and a power of (1-beta) = 0.80. This effect size represents a small to moderate effect size. The power was also expected to be sufficient to assess potential moderating effects of personality groups, as interaction effects were calculated on the total sample.

Statistical Analyses

The effects of Preventure in the intervention condition were tested using multivariate regression analyses in Mplus version 6.11 with the MLR estimator (Muthén & Muthén, 1998-2010). The analyses of interest were 'condition' and 'condition x personality risk group interaction' effects. The analyses result in standardized Beta coefficients, testing whether the 2, 6 and 12 month follow-up scores on outcomes varied as a function of the intervention within the full sample, and whether the follow-up scores on a specific outcome within a specific personality group varied as a function of the intervention (for example a reduction in depression score in the NT group).

As age, sex, educational level, and binge drinking were uneven distributed across the conditions, and as previous studies have showed that these variables are correlated to the development of our outcomes measures (Monshouwer et al., 2012; Odgers et al., 2008; Moffit & Caspi, 2001), all analyses were conducted with these variables as covariates to control for any possible bias stemming from the imbalance. However, following De Boer et al (2015) we will also present the unadjusted analyses, i.e. with only the baseline measure of the outcome variable as a covariate in the model (De Boer, Waterlander, Ruijper, Steenhuis & Twisk, 2015).

By means of a latent-growth curve approach, *post-hoc* analyses were conducted to examine the effect of the Dutch version of Preventure on the linear increase on the outcomes variables. As three follow-up measures is seen as a minimum to perform this type of analysis, we can only show the results for depression and anxiety. A latent-growth model approaches the analysis of repeated measures from the perspective of an individual growth curve for each subject; each growth curve has a certain initial level (intercept) and a certain rate of change over time

(slope) (Duncan, Duncan, Strycker, Li & Alpert, 1999). In this study the outcome slope was regressed on the Prevention intervention condition variable controlling for the outcome measure intercept. Again both adjusted and unadjusted analyses were conducted. The fit of the models was assessed by the following fit indexes: χ^2 , comparative fit index (CFI), Tucker-Lewis index (TLI) and root mean square error of approximation (RMSEA). Due to the sensitivity of the χ^2 goodness-of-fit test to sample sizes, the fit indices CFI, TLI and RMSEA were used. Except for the values of RMSEA (which would be satisfactory if smaller than 0.08), goodness-of-fit values greater than 0.90 are considered an acceptable fit (Bentler & Bonett, 1980).

Missing observations at follow-up were imputed using multiple regression imputation as implemented in Mplus 6.11. To correct for non-independence because of clustering of the data, the TYPE = COMPLEX procedure in Mplus was used. Following Schulz and Grimes (2005) the level of statistical significance was set at $p < .05$, and no Bonferroni correction was applied (Schulz & Grimes, 2005). In order to assess the influence of workshop retention, we ran all analyses on a dataset including only students who attended both workshops.

Results

Baseline

Table 1 shows the distribution of gender, age, ethnicity, educational level and recent binge drinking in both the intervention and control condition. Tables 2 and 3 show the means and standard deviations of the outcome variables for the total sample and for each of the four personality groups, separately. There were no significant baseline differences between the two conditions on any of the outcome variables.

Table 1. Socio-demographics of intervention and control condition

| | Intervention condition | Control condition |
|------------------------|------------------------|-------------------|
| Male | 47% | 57% |
| Age | 13.9 years (SD = .98) | 14.1 (SD = .77) |
| Dutch | 87% | 87% |
| Low level of education | 43% | 26% |
| Binge drinking | 49% | 37% |

Table 2. Baseline and follow-up levels of outcomes for the total sample

| | Intervention condition | Control condition | |
|--|------------------------|-------------------|--------------|
| | | Mean (SD) | Mean (SD) |
| Depression (range 0-60) | Baseline | 16.55 (7.84) | 16.44 (7.90) |
| | 2 months | 14.76 (6.70) | 14.91 (6.68) |
| | 6 months | 14.79 (7.55) | 14.46 (6.66) |
| | 12 months | 15.98 (8.71) | 14.94 (6.98) |
| Anxiety (range 0-36) | Baseline | 7.83 (6.29) | 7.85 (5.81) |
| | 2 months | 6.88 (6.19) | 6.98 (5.41) |
| | 6 months | 6.50 (6.07) | 6.62 (5.67) |
| | 12 months | 6.54 (6.40) | 6.84 (5.96) |
| Hyperactivity ^a (range 0-10) | Baseline | 5.56 (2.32) | 5.31 (2.21) |
| | 6 months | 5.36 (2.29) | 5.12 (2.34) |
| | 12 months | 5.30 (2.44) | 5.15 (2.36) |
| Delinquency ^a (range 0-14) | Baseline | 2.18 (2.28) | 1.85 (2.07) |
| | 6 months | 2.42 (2.86) | 2.45 (2.93) |
| | 12 months | 2.39 (3.13) | 2.28 (2.87) |

a Hyperactivity and delinquent behaviour was only assessed at 6 and 12 months post intervention because the measurement instrument elicits six-monthly information.

Table 3. Baseline and follow-up levels of outcomes for the four risk profiles

| | | Intervention condition | Control condition |
|---|-----------|------------------------|-------------------|
| | | Mean (SD) | Mean (SD) |
| Depression (range 0-60) (NT only) | Baseline | 22.16 (9.70) | 22.04 (8.64) |
| | 2 months | 17.25 (7.08) | 17.61 (7.09) |
| | 6 months | 16.23 (7.48) | 16.24 (7.42) |
| | 12 months | 18.44 (9.01) | 15.48 (6.94) |
| Anxiety (range 0-36) (AS only) | Baseline | 12.02 (6.32) | 12.82 (6.48) |
| | 2 months | 9.66 (7.24) | 10.38 (6.02) |
| | 6 months | 8.70 (6.82) | 10.54 (5.55) |
| | 12 months | 7.90 (6.57) | 10.89 (7.24) |
| Hyperactivity (range 0-10) (IMP only) | Baseline | 6.54 (2.17) | 6.31 (2.12) |
| | 6 months | 5.97 (2.13) | 5.86 (2.13) |
| | 12 months | 5.81 (2.18) | 5.78 (2.14) |
| Hyperactivity (range 0-10) (SS only) | Baseline | 5.36 (2.12) | 4.93 (2.17) |
| | 6 months | 5.51 (2.20) | 4.82 (2.37) |
| | 12 months | 5.29 (2.45) | 4.98 (2.36) |
| Delinquency (range 0-14) (IMP only) | Baseline | 2.54 (2.37) | 2.21 (2.21) |
| | 6 months | 2.96 (3.22) | 2.61 (2.65) |
| | 12 months | 2.95 (3.46) | 2.74 (3.36) |
| Delinquency (range 0-14) (SS only) | Baseline | 2.41 (2.28) | 2.12 (2.00) |
| | 6 months | 2.92 (2.76) | 2.96 (3.12) |
| | 12 months | 2.55 (3.04) | 2.68 (2.78) |

Intervention effects for the total high risk sample

We first tested main condition effects in the overall high risk sample. No significant differences were found between the intervention and control condition on any of the outcome variables (see Table 4).

Table 4. Effect of Prevention on depression, anxiety, hyperactivity and delinquency within the total sample

| | | β^1 (95% CI) | P | β^2 (95% CI) | P |
|---------------|-----------|----------------------|------|----------------------|------|
| Depression | 2 months | -0.021 (-0.09, 0.05) | 0.54 | -0.014 (-0.09, 0.06) | 0.70 |
| | 6 months | 0.016 (-0.06, 0.09) | 0.69 | 0.021 (-0.09, 0.13) | 0.71 |
| | 12 months | 0.048 (-0.04, 0.13) | 0.27 | 0.064 (-0.05, 0.18) | 0.26 |
| Anxiety | 2 months | -0.008 (-0.09, 0.07) | 0.84 | -0.007 (-0.10, 0.09) | 0.89 |
| | 6 months | -0.011 (-0.09, 0.07) | 0.79 | -0.014 (-0.10, 0.07) | 0.76 |
| | 12 months | -0.024 (-0.11, 0.07) | 0.60 | -0.024 (-0.12, 0.07) | 0.63 |
| Hyperactivity | 6 months | 0.015 (-0.06, 0.09) | 0.69 | 0.020 (-0.05, 0.09) | 0.59 |
| | 12 months | -0.006 (-0.08, 0.06) | 0.86 | 0.005 (-0.05, 0.09) | 0.87 |
| Delinquency | 6 months | -0.040 (-0.15, 0.07) | 0.47 | -0.039 (-0.15, 0.07) | 0.48 |
| | 12 months | -0.015 (-0.11, 0.08) | 0.74 | -0.017 (-0.11, 0.07) | 0.71 |

¹ Adjusted for sex, age, education and binge drinking.² Unadjusted for sex, age, education and binge drinking. β = Standardized logistic regression coefficient.

Intervention effects per personality group

At second we tested condition effects in specific personality groups. Table 5 shows the Beta coefficients of fourteen tests, testing intervention effectiveness on specific outcomes within specific personality groups. For depression in the NT group, a significant effect was found at 12-month follow-up. The descriptive results (Table 3) suggest that, although depressive feelings were reduced in both the intervention and the control group, the effect was stronger in the control group. Tables 3 and 5 also show that, as hypothesized, a stronger reduction of anxiety sensitivity scores was found in the AS intervention group than in the control condition at 12-month follow-up.

Table 5. Effect of Preventure on depression, anxiety, hyperactivity and delinquency within specific personality groups

| | | β^1 (95% CI) | P | β^2 (95% CI) | P |
|-----------------------------|-----------|-----------------------|-------|------------------------|-------|
| Depression (NT only) | 2 months | -0.014 (-0.14 ; 0.11) | 0.83 | -0.014 (-0.14 ; 0.11) | 0.82 |
| | 6 months | -0.031 (-0.14 ; 0.08) | 0.58 | -0.024 (-0.13 ; 0.08) | 0.66 |
| | 12 months | 0.116 (0.01 ; 0.22) | 0.04* | 0.115 (0.01 ; 0.22) | 0.04* |
| Anxiety (AS only) | 2 months | -0.012 (-0.13 ; 0.11) | 0.84 | -0.012 (-0.14 ; 0.11) | 0.85 |
| | 6 months | -0.076 (-0.19 ; 0.04) | 0.20 | -0.077 (-0.19 ; 0.04) | 0.20 |
| | 12 months | -0.128 (-0.25 ; 0.00) | 0.05* | -0.133 (-0.26 ; -0.01) | 0.04* |
| Hyperactivity (IMP only) | 6 months | -0.021 (-0.12 ; 0.08) | 0.70 | -0.022 (-0.12 ; 0.08) | 0.66 |
| | 12 months | -0.020 (-0.14 ; 0.10) | 0.75 | -0.022 (-0.15 ; 0.11) | 0.73 |
| Hyperactivity (SS only) | 6 months | 0.073 (-0.05 ; 0.19) | 0.24 | -0.071 (-0.05 ; 0.19) | 0.25 |
| | 12 months | 0.011 (-0.10 ; 0.12) | 0.85 | 0.010 (-0.10 ; 0.12) | 0.85 |
| Delinquency (IMP only) | 6 months | 0.066 (-0.05 ; 0.18) | 0.27 | 0.055 (-0.07 ; 0.18) | 0.39 |
| | 12 months | 0.022 (-0.11 ; 0.15) | 0.75 | 0.013 (-0.12 ; 0.15) | 0.85 |
| Delinquency (SS only) | 6 months | -0.005 (-0.14 ; 0.13) | 0.94 | -0.003 (-0.14 ; 0.13) | 0.98 |
| | 12 months | -0.036 (-0.12 ; 0.05) | 0.42 | -0.036 (-0.12 ; 0.05) | 0.42 |

¹ Adjusted for sex, age, education and binge drinking.

² Unadjusted for sex, age, education and binge drinking.

β = Standardized logistic regression coefficient.

Post hoc analyses

Post-hoc analyses were conducted, by means of a latent-growth curve approach, to examine the effect of Preventure on a linear increase in the outcome variables. In this model, the slope of the outcome was regressed on the Preventure intervention variable. The fit between the linear model and the data was excellent for both depression [n = 699; $\chi^2 = 19.246$ (df=4), $p < 0.001$; RMSEA = 0.072, CFI = 0.959, TLI = 0.939], and anxiety

[$n = 699$; $\chi^2 = 6.361$ ($df=4$), $p < 0.10$; RMSEA = 0.029, CFI = 0.997, TLI = 0.995]. The intercepts and slopes were significant for both depression ($\beta_0 = 2.91$, $p < 0.001$ and $\beta_1 = -0.31$, $p < 0.001$), and anxiety ($\beta_0 = 7.73$, $p < 0.001$ and $\beta_1 = -.63$, $p < 0.001$), indicating that the participants scored greater than zero on the outcomes at baseline and that the levels of depression and anxiety decreased over time. A quadratic trend was also tested, but this resulted in a non-fitted model and therefore omitted. No significant intervention effects were found on the slopes for both outcome measures, nor for the full sample analyses, nor for the specific risk profile samples.

Per protocol analyses

Given that a substantial proportion of the respondents in the intervention condition did not attend one (29%) or both (20%) workshops, we reran all analyses on a sample including all students in the control condition and only those students in the experimental group who attended both workshops (71%, $n = 242$). These analyses showed similar results, with one exception. We found a significant negative effect ($\beta = 0.107$, $p = 0.045$) in delinquent behavior in the IMP group at 6 months follow-up, but only in the fully adjusted model and not in the unadjusted model.

Discussion

The aim of this study was to test the effectiveness of Preventure in the Netherlands on a range of mental health and conduct problems at 2, 6, and 12 months post intervention. We tested for a reduction in depression, anxiety and hyperactivity rates and delinquent risk taking behavior in the total experimental group as compared to the control group and expected these effects to be moderated by personality. Only one out of the fourteen hypothesis was supported; we found a positive intervention effect on anxiety rates in the AS group at 12-month follow-up. Contrary to our hypothesis, a negative intervention effect on depression rates was found at 12-month follow-up in the NT group. These significant effects were not found in the post-hoc growth curve analyses. Notwithstanding the intervention effects on psychological problems found in a previous study of Preventure (Castellanos & Conrod, 2006), as well as the teacher

delivered variant of Preventure, our results suggest that the effects of Dutch-Preventure on psychological problems are limited. We will discuss some possible explanations for our findings.

First, although overall effects on depression outcomes were found in the Adventure trial, the study did not identify any personality-specific effects on anxiety and depression. The authors suggest that an increased intervention dose may be needed for these specific subgroups (O'Leary-Barrett et al., 2013). Moreover, the literature suggests that effective interventions for internalizing and externalizing problem behavior are, mostly, multi-year programs, focusing on at-risk groups, targeting risk and protective factors, and focusing on multiple domains i.e. school, home environment, and children as well (Greenberg, 2000). Although Preventure also focuses on at-risk groups and targets risk and protective factors, it is also a very brief and single domain intervention. A more extended and multi-domain version of Preventure could probably enhance its effectiveness on mental health outcomes. However, this would make the intervention more expensive. Moreover it could make the intervention less attractive for schools to implement as it becomes more intensive and complex, and therefore more time consuming which may interfere too much with the day-to-day running of the school. Second, the prevalence of delinquency in our sample appeared to be low. Consequently, it may be that this study was underpowered to detect intervention effects on this outcome measure. Third, interventions that have proven to be effective in the original context are for a variety of reasons not always effective outside this context. Sample differences, other ways of delivering the intervention, cultural differences and many other factors can influence the outcomes of preventive interventions (Eisner, 2009; Flay et al., 2005).

Strengths and Limitations

A strength of this study is that the outcomes were measured at 2, 6, and 12 months. Intervention effects could therefore be tested shortly after implementation, as well as a full year later. A limitation in this study is that the fidelity of the implementation was not monitored by means of thorough measurements. Although the counselors received their training

directly from the intervention developers, and had observation and feedback sessions, it cannot not be ruled out that the implementation quality was not as high as during the other Prevention studies. Another limitation is that the use of self-reports might have led to measurement errors, due to situational and cognitive influences (Brener, Billy & Grady 2003). To overcome situational influences (e.g. social desirability) and to optimize measurement validity, we guaranteed full confidentiality (anonymity) to our participants (Koning et al., 2009; Del Boca & Darkes, 2003).

Conclusion

In conclusion, in the ongoing search for effective school-based health promotion interventions that are brief and target multiple problems, Prevention appears to be effective in the prevention of substance use problems, yet this study found no evidence for its effectiveness on other mental health problems. Thereby, this study demonstrates the importance of replication studies. Although much can be learned from interventions that were developed and studied abroad, it cannot be assumed that positive intervention effects found in one study will also be found in studies performed in other (cultural) contexts and under other circumstances. However, the personality-targeted prevention approach is quite novel and promising. Many school-based interventions suffer from severe implementation problems because of the length of the program, which in turn affects their effectiveness (Elliot & Mihalic, 2004), and it is therefore worth continuing to search for interventions that are very short, simple, targeted and effective.

Acknowledgements

This study was funded by grant number 120520011 from the Dutch Health Care Research Organization (ZonMw). Furthermore, we are grateful to all who participated in this study (counsellors, co-facilitators, teachers, children, and their parents).

References

- Arnett, J.J. (1994). Sensation seeking: A new conceptualization and a new scale. *Personality and Individual Differences*, 16, 289-296.
- Baerveldt, C. (1992). Schools and the prevention of petty crime: Search for a missing link. *Journal of Quantitative Criminology*, 8, 79-94.
- Bentler P. M., Bonett D. G. (1980). Goodness-of-fit procedures for the evaluation and selection of log-linear models. *Psych Bulletin*; 88: 588-606.
- Blackwell, E., Conrod, P.J., & Hansen, N. (2002, June). *Negative cognitions, hopelessness, and depression related drinking motives*. Abstract in Alcoholism: Clinical and Experimental Research, 26 (Supplement), 27A. Poster session presented at the Scientific Meeting of the Research Society on Alcoholism, San Francisco, CA.
- Bouma, J., Ranchor, A.V., Sanderman, R., & Van Sonderen, E. (1995). *Het meten van symptomen van depressie met de CES-D: een handleiding*. [Measuring symptoms of depression with the CES-D: Manual]. Groningen: Rijksuniversiteit Groningen.
- Brener N.D., Billy J.O., Grady W.R. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: evidence from the scientific literature. *J Adolesc Health*, 33, 436-57.
- Castellanos, N., & Conrod, P.J. (2006). Brief interventions targeting personality risk factors for adolescent substance misuse reduce depression, panic and risk-taking behaviours. *Journal of Mental Health*, 15(6), 645-658.
- Comeau, N., Stewart, S.H., & Loba, P. (2001). The relations of trait anxiety, anxiety-sensitivity, and sensation seeking to adolescents' motivations for alcohol, cigarette, and marijuana use. *Addictive Behaviors*, 26, 1-24.
- Conrod, P.J., Castellanos, N., & Mackie, C. (2008). Personality-targeted interventions delay the growth of adolescent drinking and binge drinking. *Journal of Child Psychology and Psychiatry*, 49(2), 181-190.
- Conrod, P.J., Castellanos-Ryan, N., & Mackie, C. (2011). Long-term effects of a personality-targeted intervention to reduce alcohol use in adolescents. *Journal of Consulting and Clinical Psychology*, 79(3), 296-306.
- Conrod, P.J., Castellanos-Ryan, N., & Strang, J. (2010). Brief, personality-targeted coping skills interventions and survival as a non-drug user over a 2-year period during adolescence. *Archives of General Psychiatry*, 67(1), 85-93.
- Conrod, P.J., Stewart, S.H., Comeau, M.N., & Maclean, A.M. (2006). Efficacy of cognitive-behavioral interventions targeting personality risk factors for youth alcohol misuse. *Journal of Clinical Child and Adolescent Psychology*, 35(4), 490-504.
- Conrod, P.J., & Woicik, P. (2002). Validation of a four factor model of personality risk for substance abuse and examination of a brief instrument for assessing personality risk. *Addiction Biology*, 7, 329-346.
- Cooper, M.L., Frone, M.R., Russell, M., & Mudar, P. (1995). Drinking to regulate positive and negative emotions: A motivational model of alcohol use. *Journal of Personality and Social Psychology*, 69, 990-1005.
- De Boer, M.R., Waterlander, W.E., Kuijper, L.D.J., Steenhuis, I.H.M., & Twisk, J.W.R. (2015). Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. *International Journal of Behavioral Nutrition and Physical Activity*, 12:4.
- De Graaf, R., Bijl, R.V., Smit, F., Vollebergh, W.A.M., & Spijker, J. (2002). Risk factors for 12-month comorbidity of mood, anxiety, and substance use disorders: Findings from the Netherlands Mental Health Survey and Incidence Study. *American Journal of Psychiatry*, 156, 620-629.
- Del Boca F.R., Darkes J. (2003). The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction*, 98, 1-12.
- Duncan T. E., Duncan S. C., Strycker L. A., Li F., Alpert A. (1999). *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications*. Mahwah, NJ: Erlbaum.
- Eisner, M. (2009). No effects in independent prevention trials: can we reject the cynical view? *Journal of Experimental Criminology*, 5, 163-183.
- Elliot, S.E., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science*, 5(1), 47-53.
- Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., Gottfredson, D., Kellam, S., Mocicki, E.K., Schinke, S., Valentine, J.C., & Ji, P. (2005). Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination. *Prevention Science*, 6, 151-75.

- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 38(5), 581-586.
- Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology*, 27(1), 17-24.
- Greenberg, M.T. (2000). *Preventing mental disorders in school-age children. A review of the effectiveness of prevention programs*. Prevention Research Center for the Promotion of Human Development College of Health and Human Development Pennsylvania State University.
- Hibell, B., Guttormsson, U., Ahlström, S., Balakireva, O., Bjarnason, T., Kokkevi, A., & Kraus, L. (2009). *The 2007 ESPAD Report. Substance use among students in 35 European countries*. Stockholm: The Swedish Council for Information on Alcohol and Other Drugs (CAN).
- Joiner, T.E., Jr. (2000). A test of hopelessness theory in youth psychiatric inpatients. *Journal of Clinical Child Psychology*, 29, 167-176.
- Kessler, R.C. (2004). Impact of substance abuse on the diagnosis, course, and treatment of mood disorders. The epidemiology of dual diagnosis. *Biological Psychiatry*, 56(10), 730-737.
- Kendler, K.S., Prescott, C.A., Myers, J., & Neale, M.C. (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Archives of General Psychiatry*, 60(9), 929-937.
- Koning I.M., Vollebergh W.A.M., Smit F., Verdurnen J.E.E., van der Eijnden R.J.J.M., ter Bogt T.F.M. et al. (2009). Preventing heavy alcohol use in adolescents (PAS): cluster randomized trial of a parent and student intervention offered separately and simultaneously. *Addiction*, 104, 1669-78.
- Krank, M., Stewart, S.H., Wall, A.-M., Woicik, P.B., & Conrod, P.J. (2011). Structural, concurrent, and predictive validity of the substance use risk personality scale in early adolescence. *Addictive Behaviors*, 36, 37-46.
- Lammers, J., Goossens, F., Conrod, P., Engels, R., Wiers, R.W., Kleinjan, M. (2015). Effectiveness of a selective intervention program targeting personality risk factors for alcohol misuse among young adolescents: Results of a cluster randomized controlled trial. *Addiction*, 110, 1101-1109.
- Lammers, J., Goossens, F., Lokman, S., Monshouwer, K., Lemmers, L., Conrod, P., ... Kleinjan, M. (2011). Evaluating a selective prevention programme for binge drinking among young adolescents: Study protocol of a randomized controlled trial. *BMC Public Health*, 11:126. doi:10.1186/1471-2458-11-126.
- Monshouwer, K., Smit, F., Ruiter, M., Ormel, J., Verhulst, F.C., Vollebergh, W.A.M., & Oldehinkel, A.J. (2012). Identifying target groups for the prevention of depression in early adolescence: the TRAILS study. *Journal of Affective Disorders*, 138, 287-294.
- Muthén, B., & Muthén, L. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén. Available at www.statmodel.com.
- Odgers, C.L., Moffitt, T.E., Broadbent, J.M., Dickson, N., Hancox, R.J., Harrington, H., et al. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology*, 20, 673-716.
- O'Leary-Barrett, M., Mackie, C.J., Castellanos-Ryan, N., Al-Khudairy, N., & Conrod, P.J. (2010). Personality-targeted interventions delay uptake of drinking and decrease risk of alcohol-related problems when delivered by teachers. *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(9), 954-963.
- O'Leary-Barrett, M., Topper, L., Al-Khudairy, N., Pihl, R.O., Castellanos-Ryan, N., Mackie, C.J., & Conrod, P.J. (2013). Two-year Impact of personality-Targeted, Teacher-Delivered Interventions on Youth Internalizing and Externalizing Problems: A Cluster-Randomized Trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52(9), 911-920.
- O'Malley, P.M., Bachman, J.G., & Johnston, L.D. (1983). Reliability and consistency in self-reports of drug use. *International Journal of the Addictions*, 18, 805-824.
- Maller, R.G., & Reiss, S. (1992). Anxiety sensitivity in 1984 and panic attacks in 1987. *Journal of Anxiety Disorders*, 6, 241-247.
- Malmberg, M., Overbeek, G., Monshouwer, K., Lammers, J., Vollebergh, W.A.M., & Engels, R.C.M.E. (2010). Substance use risk profiles and associations with early substance use in adolescence. *Journal of Behavioral Medicine*, 33, 474-485.
- Moeller, F.G., Barratt, E.S., Dougherty, D.M., Schmitz, J.M., & Swann, A.C. (2001). Psychiatric aspects of impulsivity. *American Journal of Psychiatry*, 158, 1783-1793.
- Moffitt, T.E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology*, 13, 355-375.
- Pulkkinen, L., & Pitkänen, T. (1994). A prospective study of the precursors to problem drinking in young adulthood. *Journal of Studies on Alcohol*, 55, 578-587.

Chapter 3

- Reiss, S.R., Peterson, R., Gursky, D., & McNally, R. (1986). Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. *Behaviour Research & Therapy*, 24, 11-18.
- Schulz, K.F., & Grimes, D.A. (2005). Multiplicity in randomized trials I: endpoints and treatments. *Lancet*, 365, 1591-95.
- Shall, M., Kemeny, R., & Maltzman, I. (1992). Factors associated with alcohol use in university students. *Journal of Studies on Alcohol*, 53, 122-136.
- Silverman, W.R., Fleisig, W., Rabian, B., & Peterson, R.A. (1991). Childhood Anxiety Sensitivity Index. *Journal of Clinical Child Psychology*, 20, 162-168.
- Stewart, S.H., Peterson, J.B., & Pihl, R.O. (1995). Anxiety sensitivity and self-reported alcohol consumption rates in university women. *Journal of Anxiety Disorders*, 9, 283-292.
- Van Widenfelt, B.M., Siebelink, B.M., Goedhart, A., & Treffers, D.A. (2002). The Dutch childhood Anxiety Sensitivity Index: Psychometric properties and factor structure. *Journal of Clinical Child Psychology*, 31(1), 90-100.
- Walsh, T.M., Stewart, S.H., McLaughlin, E., & Comeau, M.N. (2004). Gender differences in Childhood Anxiety Sensitivity Index dimensions. *Journal of Anxiety Disorders*, 18, 695-706.
- Weems, C.F., Hammond-Laurence, R., Silverman, W.R., & Ginsburg, G.S. (1998). Testing the utility of the anxiety sensitivity construct in children and adolescents referred for anxiety disorders. *Journal of Clinical Child Psychology*, 27, 69-77.
- Wiers, R.W., Ames, S.L., Hofmann, W., Krank, M., & Stacey, R.W. (2010). Impulsivity, impulsive and reflective processes and the development of alcohol use and misuse in adolescents and young adults. *Front Psychology*, 1, 144. doi: 10.3389/fpsyg.2010.00144. Epub 2010 Sep 30.
- Woicik, P.A., Stewart, S.H., Pihl, R.O., & Conrod, P.J. (2009). The substance use risk profile scale: A scale measuring traits linked to reinforcement-specific substance use profiles. *Addictive Behaviour*, 34, 1042-1055.

4

Effectiveness of an empowerment program for adolescent second generation migrants: a cluster randomized controlled trial

Ferry Goossens
Simone Onrust
Karin Monshouwer
Bram Orobio de Castro

Published as:

Goossens, F.X., Onrust, S.A., Monshouwer, R., & Orobio de Castro, (2016). Effectiveness of an empowerment program for adolescent second generation migrants: A cluster randomized controlled trial. *Children and Youth Services Review*, 64, 128-135.

Authors' contributions:

FG carried out the study (together with SO), analyzed the data (together with SO), and wrote the manuscript. RM and BODC revised the manuscript.

Conflict of interest:

FG, SO, and RM wish to confirm that the intervention is property of the Trimbos Institute. There are no other conflicts of interest associated with this publication, and there has been no significant financial support for this work that could have influenced its outcome.

Abstract

Dutch adolescent second generation migrants are at increased risk of becoming marginalized and developing problem behavior. We tested the effectiveness of the Dutch multi-component empowerment program POWER that aims to prevent such problems. We hypothesized a positive intervention effect on participants' sense of mastery, coping skills, activities, conduct problems, and prosocial behavior. We conducted a cluster randomized controlled trial including a pretest and posttest with an intervention condition ($n=132$) and a waitlist control condition ($n=116$). Analyses showed that POWER was only effective in influencing the youngsters' participation in activities like sport, hobbies, and casual work. When implemented with high fidelity, POWER also influenced the level of conduct problems as well as their coping style. However, a more accurate registration of the implementation process would have been helpful and the program can potentially be further improved by clearly specifying which components of the program can be adapted and which must be delivered as intended.

Introduction

Dutch adolescent second generation migrants - in particular those with a Moroccan, Antillean/Aruban, Surinamese, or Turkish background - show increased risk of developing behavioral problems and have poorer prospects than their autochthon peers. They are overrepresented in juvenile delinquency statistics (Blom, Oudhof, Bijl, & Bakker, 2005; Jenissen, 2007), drop out of school more often (De Boom, Weltevreden, Van Wensveen, Van San, & Hermus, 2011), and are overrepresented in forensic mental health care (Boon, De Haan, & De Boer, 2010). At the same time, these youth are underrepresented in prevention programs (Ince & Van den Berg, 2010) and mental health care (Boon et al., 2010), suggesting that opportunities to prevent problematic development of youth with a migrant background are being missed.

The Challenges of Adolescent Second Generation Migrants

A complex combination of individual, parental, school, peer, and community factors seem to influence the development of these problems (Paalman, 2013). Although the relation between risk factors and behavioral problems needs more research to arrive at a better understanding, some aspects seem to play an important role. Minorities often have a low socio-economic background (SCP, 2009), characterized by poor housing conditions, poverty, and limited education and career options (Dagevos, Gijsberts, & Van Praag, 2003; Dijkman, 1996; Eldering & Knorth, 1997; Martens, 1999; Schonberg & Shaw, 2007). They have to cope with discrimination (Berry, Phinney, Sam, & Vedder, 2010) and cultural incompatibilities between the home culture and the host culture (Guarnaccia & Lopez, 1998), and have specific family dynamics that can be risk factors for behavioral problems (Stevens, Vollebergh, Pels, & Crijnen, 2007). They often also lack sufficient social-emotional skills and problem-solving skills (Orobio de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002; Trentacosta & Shaw, 2009) and – albeit strongly dependent on the informant and outcome measures used – some researchers have found increased mental health problems including a variety of internalizing and externalizing problems in general (Stevens & Vollebergh, 2008) and psychotic disorders in specific (Veling, Selten, Mackenbach, & Hoek,

2007). Stigmatization, prior disappointing experiences, language/cultural differences, and limitations in resources also tend to be barriers to ethnic minority families searching for and accepting help for child behavior problems (Scheppers, Van Dongen, Dekker, Geertzen, & Dekker, 2006; Tolan & McKay, 1996).

Marginalization

One of the biggest challenges faced by these adolescents is overcoming their risk of exclusion, caused by an interaction of several economic, societal, and cultural processes. Growing up in disadvantage is strongly related to social marginalization; that is, exclusion from fulfilling social lives at individual, interpersonal, and societal levels (Bynner, 2006). Marginalized individuals experience relatively little control over their life and available resources, they risk becoming stigmatized. This risk of exclusion can be reinforced by cultural processes, such as acculturation. Acculturation refers to the process of cultural and psychological change that results following a meeting between cultures. Berry (2003, 2005) distinguished four different acculturation strategies: integration, assimilation, separation and marginalization. Individuals showing interest in maintaining the original culture as well as in learning and participating in the other culture (integration) display the fewest problems, whereas individuals showing both little involvement in maintaining the culture of origin and little interest in participating in the other culture (marginalization) appear to have the poorest mental health outcomes (Roneru, Weisman de Mamani, Flynn, & Betancourt, 2007).

Empowering second generation migrant youngsters

Prevention programs can teach second generation migrant youngsters to cope with the above challenges and prevent marginalization. Several review and meta-analytic studies have demonstrated that mainstream prevention programs appear equally effective for ethnic minority youth and majority youth (Barlow, Shaw, & Stewart-Brown, 2004; Huey & Polo, 2008; Leijten, Raaijmakers, Orobio de Castro, & Matthys, 2013; Weisz, JensenDoss, & Hawley, 2006; Wilson, Lipsey, & Soydan, 2003). However, mainstream programs are also less successful in reaching ethnic minority youth than majority youth. In addition, ethnic minority youth are much

more likely to drop out (Barlow et al., 2004; Wilson et al., 2003; Barrett & Ollendick, 2004).

Culturally sensitive programs, for example empowerment programs, appear more effective in reaching and retaining minority youth than mainstream programs (Kumpfer, Alvarado, Smith, & Bellamy, 2002). Empowerment programs also differ from more mainstream prevention programs in a way that they do not only want to influence the individuals psychological sense of personal control or influence, but also aims to have social influence (Rappaport, 1987). More precisely, empowerment programs aim to support individuals to develop a sense of mastery; that is, to become conscious of the effective fields of influence in their immediate surroundings in the context of their lives and to develop skills and abilities to gain reasonable control over their lives (Lee, 1992). Studies show that empowerment programs can lead to an increase in ethnic consciousness (Gutiérrez, 1990), a more positive feeling about prospects (Parsons, 1989), improvement in the internal locus of control (Parsons, 1989), better skills to analyze problems (Gutiérrez, 1990), more involvement in activities at school and in the neighborhood (Fertman & Chubb, 1992), and can have a positive impact on violent and provoking behavior, school delinquency, drug use, and condom use (Flay, Graumlich, Segawa, Burns, & Holliday, 2004).

Notwithstanding these promising results, the number of rigorous effectiveness studies of empowerment programs is still very limited, especially those including second generation migrants (Reischl et al., 2011; Flay et al., 2004). Along with more general reasons, this is probably due to difficulties in recruiting this group for preventive interventions. To date, no interventions available in the Netherlands have proved effective in preventing or reducing problem behavior in adolescent second generation migrants (Ince & Van den Berg, 2010).

The Dutch multi-component empowerment program POWER, however, proved to be successful in reaching and retaining adolescent second generation migrants at risk of marginalization. Specifically, POWER reaches out in deprived neighborhoods to migrant youngsters who seem

to be at risk of marginalization in terms of hanging around with criminal youth, carrying out acts of vandalism, dropping out of school, or being in frequent contact with the police. A pilot with 7 groups, each comprised of 9 to 14 young people, demonstrated that the youngsters were interested in participation and that most participants followed the course from beginning to end (van Diest, Wennink, & Uiterloo, 2005). The key element seems to be that trainers recruit participants themselves, mainly on the basis of their own observations and contacts with youngsters and with important key figures in the community, for example the local Imam for Islamic youth. Given the program's success in reaching and retaining these youngsters, it is important to know whether the intervention is also effective.

Aim and Hypothesis

The present study aimed to test the effectiveness of POWER in daily practice with a clustered randomized controlled trial. Primary outcome measures were problem behavior and social marginalization. Secondary outcome measures were the participants' sense of mastery and coping skills. We hypothesized positive intervention effects on all outcome measures. In moderator analyses, we tested whether intervention effects depended on gender, the level of problem behavior at baseline, ethnicity, or the implementation process.

Method

This study was a cluster randomized controlled trial with an intervention condition and a waitlist control condition, both including 16 groups of youngsters comprised of between 5 and 12 youngsters per group (see Figure 1). In line with the randomization process, the trainers were instructed to recruit comparable groups each time with respect to ethnicity, gender, age, levels of problem behavior, and community: for example, two groups of Moroccan boys from 16 to 18 years of age from two different areas in a city, causing trouble in their neighborhood. Each time a trainer recruited two comparable groups, received written consent, and baseline data were collected, these two groups were randomized

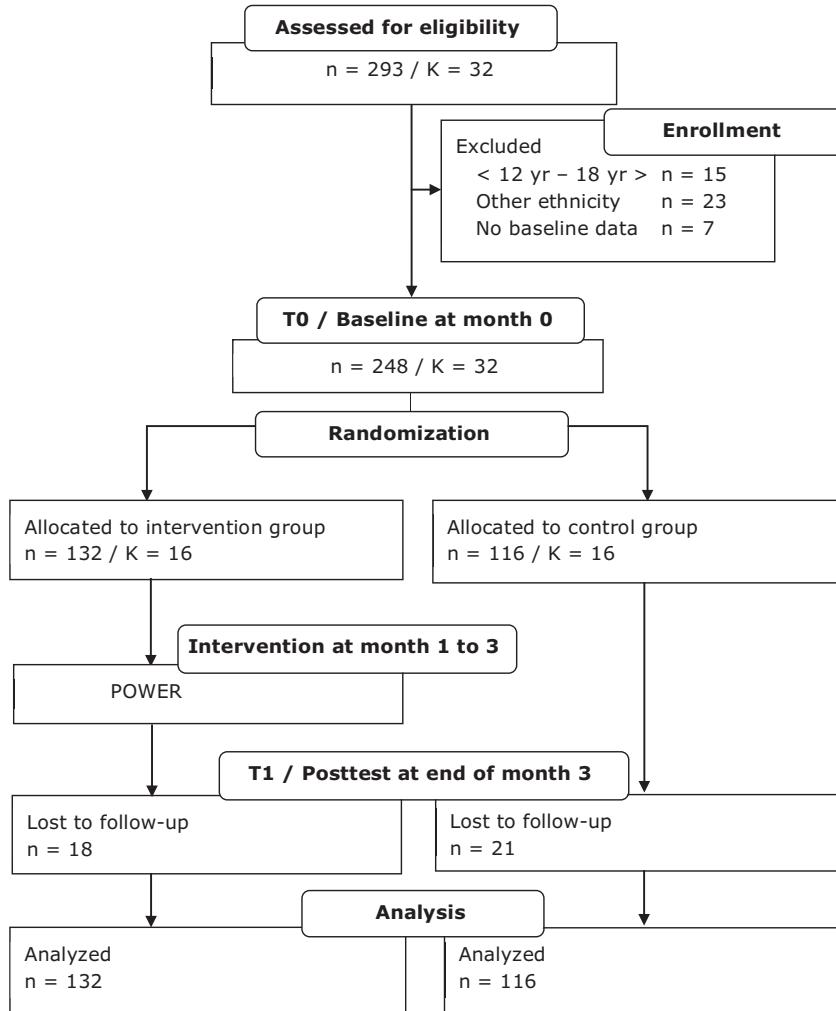
by the research staff. In the intervention condition, POWER was implemented directly after the randomization; in the waitlist condition, POWER started directly after the intervention group finished theirs (i.e. three months later). Some trainers provided the control condition group with one or two short activities (i.e. football or tenpin bowling) to keep them involved in the study during the implementation of POWER in the intervention group.

Participants were eligible to enter the study if they fulfilled the following inclusion criteria: 1) a Moroccan, Turkish, Surinamese, or Antillean/Aruban father and/or mother, 2) between 12 and 18 years of age, 3) no mental disorders, 4) not currently doing community service, and 5) at risk of marginalization. At risk of marginalization was defined as (a) living in a disadvantaged area, (b) poor social-emotional development (i.e. poor social-emotional skills/developing problem behavior), and (c) lack of structured social activities/hanging around in the neighborhood. Following the regular recruitment strategy of POWER no questionnaires or diagnostic instruments were used in the recruitment phase, because such measures might scare or annoy potential participants. Trainers were trained in how to select the intended target group.

Study Intervention

POWER consists of three elements (1) a culturally sensitive empowerment group course for youngsters, (2) a course for their parents, and (3) a community approach, i.e. involving relevant local organizations during the project. Although POWER tries to influence both the individual level as well as (the relation with) relevant stakeholders in the community, the main focus of the program is to influence the participants' feeling and coping strategies of personal control over their life and environment. POWER is designed as both a preventive and a curative intervention, depending on the youngsters' level of problem behavior at onset. The program is written in Dutch, however, an English version of the intervention, and more information regarding international implementation of POWER can be obtained from the authors.

Figure 1. Flow of participants



The objective of POWER is to reduce the risk of marginalization and problem behavior by means of the development and expansion of the feeling of personal control over one's life and environment (i.e. mastery), problem-solving abilities, and competencies. The program is based on the work of Lee (1992) on the empowerment of young black men in modern American society, and is focused around the following six basic principles. 1) The program can be used for groups that are either mono-cultural or multi-cultural. 2) The program is not neutral or value-free. There is a clear vision concerning what is allowed and what is not, what is right

or wrong. 3) As many exercises as possible allow the course members to reason out beforehand the consequences and implications of their actions, their attitude, their (behavioral) expression, etc. 4) The exercises also offer the opportunity for reflection and feedback of the actions after the fact: what went well/wrong and what could have been done better? 5) Without respect and appreciation of your own culture, you cannot have any respect for other cultures. 6) Responsibility, motivation and dealing with problems are the connecting themes throughout all of the modules.

The course for the youngsters includes 13 training modules of two to three hours with a frequency of at least once a week (Uiterloo, 2005a). The following themes are addressed: (cultural) identity, relationships/sexuality, dealing with frustration, criminality, health and drug use, residential and living environment, use of free time, education/work, and future prospects. All of the modules consist of a welcoming word and a retrospective on the preceding module (15 minutes), a workshop (120 minutes), and evaluation and assignments for the next week (30 minutes). The sessions include a workbook (Uiterloo, 2005b), writing exercises, discussion, role playing, watching movies, sex or drug education by prevention specialists, and homework. Using role-play, case histories and observations, the youngsters are able - under the guidance of a trainer and, if necessary, by using role models - to learn skills and attitudes in a realistic setting that are essential for social success. They can exchange experiences and receive the opportunity to practice their new behavior. The program intends to realize a shift from external to internal locus of control (the degree to which someone feels that he himself can exercise influence). The youngsters are regularly asked to make choices and - both before and afterwards - to think about the consequences of those choices. Alongside of that, an attempt is made to increase their communicative skills, their defensibility and their empathy, as well as their problem-solving abilities. The last session is the rite of passage: an official closing ceremony that symbolizes the transition to a new phase of life.

The family and parenting are important factors in the prevention, origin and continuation of problem behavior among young people. Therefore a parents' course is implemented parallel with the course for

youngsters and includes four theme meetings of two and a half hours, with particular emphasis on communication with their children, their style of parenting, and their involvement in their children's education. The objective of these meetings is to give the parents insight into the empowerment approach and to provide them with the skills to be able to react effectively to their children's new behavior (Uiterloo, 2005c). The first meeting is an introduction session. The second sessions aims to inform parents about the new behavior that the youngsters are being educated in order to create a safe environment at home in which this behavior can be practiced. The third and fourth session are focusing on the problems that are prevalent in their neighborhood and the ways to cope with these problems.

The community approach is intended to embed the intervention in the community. Course leaders can achieve this by involving the mosque, members of the youngsters' family, local healthcare, welfare, local government officials, and the police. Ideally a task group is formed of them during the POWER course who will meet regularly. The goal of the meetings is to discuss the progress of the program, to stimulate cooperation between the stakeholders, to facilitate the trainer during the implementation of the training, and to help to solve practical problems the youngster face. The people in this task group could, for example, provide psycho-education, additional support relating to education, family support, or join the rite of passage. Afterwards, the task group will decide what opportunities there are to support the youngsters who have successfully completed the program in the neighborhood in a positive manner (Uiterloo, 2005d).

Sample

The sample consisted of 248 Moroccan, Turkish, Surinamese, and Antillean/Aruban second generation migrants between 12 and 18 years of age from 16 lower-class neighborhoods in nine cities in the Netherlands. Participant groups were randomly assigned to 132 youngsters in 16 groups in the intervention condition and 116 youngsters in 16 groups in the control condition (see Table 1). Half of the courses were given in some of the 40 most problematic neighborhoods in the Netherlands as determined by

the Dutch Ministry of Housing, Neighborhoods and Integration in 2007. These neighborhoods are characterized by an accumulation of social, physical and economic problems.

Because of the method of working with groups of friends, both some younger and older peers, and some peers of another ethnicity were allowed to join the course when they were part of a group. They (n=38) were not, however, included in the analyses. Seven participants who did not fill out the baseline questionnaire were also excluded from the analyses. As the recruitment of youngsters for POWER is a dynamic process of several communication junctures at different (in)formal locations and times, it is not possible to give a number for the participants who were recruited but did not formalize their participation.

Table 1. Baseline characteristics of intervention and control condition

| | Intervention Condition | Control Condition |
|------------------|-----------------------------|-----------------------------|
| n | 132 | 116 |
| Male | 77 % | 65 % |
| Age | 14.8 years (SD = 1.5 years) | 15.2 years (SD = 1.5 years) |
| Low education | 72 % | 62 % |
| Origin | | |
| Moroccan | 49.2 % | 48.3 % |
| Turkish | 40.9 % | 40.5 % |
| Surinamese | 9.1 % | 11.2 % |
| Antillean/Aruban | 0.8 % | 0.0 % |

As POWER focuses both on youngsters already showing problem behavior and on youngsters at risk of developing problem behavior, and as participants were recruited without diagnostic measures, we checked the extent to which the participants were already showing problem behavior at baseline. To check the extent to which the youngsters were already showing problem behavior, the baseline scores from the conduct problem scale and the prosocial scale from the Strengths and Difficulties Questionnaire (SDQ) were compared with the SDQ scores from a subsample drawn from the Dutch Health Behavior in School-aged Children (HBSC) survey dataset (Van Dorsselaer et al., 2010). The age range in this HBSC subsample (n=5,214) was 12 to 18 years, and the

mean age was 14.8 years. The analyses revealed that girls showed less prosocial behavior compared to their norm group. Furthermore we compared the means of two activity scales from the Youth Self Report (YSR, see Measures below) with the means of these scale scores from a Dutch norm group. This norm group consisted of 1,124 Dutch youngsters from 11 to 18 years of age (Verhulst, Van der Ende, & Root, 1997). Simple one sample t-tests showed that both boys and girls from the POWER sample indeed participated less in activities like sport, hobbies, and casual jobs. On the social scale, measuring having (contact with) friends, the girls scored the same as their norm group and the boys scored better than their norm group, indicating that no significant self-reported social problems were present.

In sum, the sample consisted of migrant youth supposedly at risk of behavior problems and marginalization, mostly living in disadvantaged areas, and lacking structured social activities. They, however, consider themselves as not having many psychosocial problems. As POWER is designed as both a preventive and a curative intervention it is most likely that it would serve mainly as a preventive intervention in this study.

Loss to follow-up

The percentage of youngsters who did not complete the posttest was statistically equal between the two conditions, i.e. 18% in the intervention group and 14% in the control group. Missing data from participants who dropped out on posttest were imputed. Those who dropped out at posttest were more likely to be male ($F(1,246) = 27,862, p < .05$) than those who were present at the posttest. On the other socio-demographic or outcome variables, they did not differ. Those who dropped out in the intervention condition did not differ on socio-demographic or outcome variables from those who dropped out in the control condition.

Implementation

Using interviews, we were able to collect implementation data from 14 of the 16 courses. Eleven consisted of the full 13 sessions, two of 10 to 12 sessions, and one of eight sessions. In 12 groups, the mean duration was between two and three hours, in one group, less than two hours, and

in one group more than three hours. In one course on average 60-70% of the youngsters were present during the sessions, in three courses on average 70-80% was present, in eight courses on average 80-90% was present, and in two courses 90-100% of the participants was on average present during the sessions.

In total, 11 parent courses were given. Two trainers were unable to organize the parent course due to parents' lack of interest. They compensated for this by regularly keeping in touch with the parents personally. If the parent course was provided, between 40% and 90% of the parents attended it. The mean number of sessions per parent course was around two, and the mean duration was around two hours. Overall, half of the parent sessions were given according to the manual and half of the sessions were adapted.

The number of organizations involved in the community approach varied between 1 and 13 per course. In almost all courses, the police, municipal health service, mental health care, and migrant organizations were involved. The police were involved to discuss problems in the neighborhood. The health services were involved mostly in delivering information to the youngster on topics like sexual health or substance use. The migrant organizations were mainly involved in the recruitment of the youngsters and their parents.

Procedures

Recruitment

Ten youth workers with the same ethnic background and gender as the participants were recruited and trained in a three-day course (Uiterloo, 2005e, 2005f) by two experienced POWER trainers. All trainers had attended at least higher vocational education and were employed in addiction care, social work, mental health care, migrant organizations, or municipal health service. Each trainer received a set of intervention manuals (Uiterloo, 2005a, 2005c, 2005d), and the experienced trainers remained available for follow-up support. The trainers recruited the participants themselves through their personal and professional network, and using

their contacts with the local mosque and the community they normally work in. Much effort was put in the recruitment phase to gain the parents and youngsters trust to participate in the intervention and study. Within a timeframe of 2.5 years the ten trainers, as well as one of the experienced trainers, recruited a total of 32 groups of six to twelve youngsters.

Consent

Parents and participants were informed about the study in a letter distributed by the trainers on behalf of the research group. This formal way of informing, which was in accordance with the medical ethical commission's guidelines, was assisted by telephone calls and/or house visits by the trainers. Parents from all participating children gave their written active informed consent for the participation of their child(ren) in the study. Besides, youngster between 16 to 18 years of age had, and did, also give their own written active informed consent.

Data collection

Data collection took place in both groups by means of self-report questionnaires. The baseline measure was assessed in both conditions directly before the randomization, and the posttest directly after the implementation of POWER in the intervention condition.

Outcome measures

Primary outcome measures

Problem behavior

The SDQ is a self-report behavioral screening questionnaire measuring problem behavior in children (Goodman & Scott, 1999). In this study, we used the conduct problems score ($\alpha = .63$, item 7 "I mostly do what others tell me I have to do" was deleted because of its low internal consistency) and the prosocial scale ($\alpha = .62$). Both scales include five items. Items were measured on a 3-point scale and related to the frequency of children's positive or negative behaviors, resulting in one total score (range 0-10) for each scale. The SDQ has been shown to correlate highly with other well-established measures, including the Child Behavior Checklist

(Goodman & Scott, 1999). It has been shown that children from different cultural backgrounds are able to use the SDQ successfully (Achenbach et al., 2008).

Social marginalization

Social marginalization was operationalized as the number of societally oriented activities and social relations in which the participants were involved. These concepts were measured with two YSR-scales: the activities scale and the social scale (Verhulst et al., 1997). The activities scale includes four items and asks how much time the adolescents spend on sports, hobbies, and casual jobs, relative to their peers. For each item, 0, 1, or 2 points can be scored. The six-item social scale asks about contact with friends and other peers and membership of (sports) clubs relative to peers and is scored in the same way as the activities scale. Also, two open questions were used, ascertaining the nature and number of sports and other activities in which each participant involved, each question scoring 0 for no or one sport/activity, 1 for two sports/activities, and 2 for three or more sports/activities. Both scores were totaled, resulting in one score (range 0-4). To avoid contamination in the data, POWER could not have been filled in as a hobby or a club.

Secondary outcome measures

Mastery

The seven-item sense of mastery scale (Pearlin, Menaghan, Lieberman, & Mullen, 1981) was used to assess the extent to which individuals felt that they could successfully affect their environment and that their destiny was in their own hands. Items were rated on a 5-point Likert scale from *strongly disagree* to *strongly agree*. Internal consistency was $\alpha = .65$. The responses to all items were summed in one total score (range 7-35).

Coping

The Utrecht Coping List for Adolescents (Bijlstra, Jackson, & Bosma, 1994) measures seven different coping styles on a 4-point Likert scale, i.e. active way of tackling problems (seven items, $\alpha = .77$), palliative reaction (eight items, $\alpha = .71$), searching for social support (six items, $\alpha = .73$), passive

reaction (seven items, $\alpha = .75$), and reassuring thoughts (three items, $\alpha = .61$). The internal consistency of the expressing emotions (three items, $\alpha = .55$) and avoiding behavior scales (eight items, $\alpha = .59$) were poor and excluded from the analyses. The test-retest reliability varies between 0.45 and 0.85 (Shreurs, Tellegen, & Van de Willige, 1984; Schreurs & Van de Willige, 1988).

Moderators

Demographics

A self-report questionnaire assessed participants' age, gender, level of education, and father and mother's country of birth. The child's ethnicity was determined by the mother's country of birth. If the mother was not born in Morocco, Turkey, Surinam, or Antilles/Aruba, or was born in the Netherlands, the child's ethnicity was determined by the father's country of birth.

Implementation

In order to monitor the implementation process, trainers were interviewed after the implementation of POWER. With respect to the moderator analyses, we asked them to quantify a) the number of lessons they gave and b) the extent to which they followed the instructions in the manual for each of the 13 lessons of the course for youngsters: (a) 100–75%, (b) 75–50%, or (c) less than 50%.

Statistical Analyses

The study aimed to detect a standardized effect size d on the primary outcomes of 0.40, which is a medium effect size (Lipsey & Wilson, 1993). Tested one sided with an alpha = 0.05 and a statistical power (1-beta) of 0.80, 78 respondents were needed in each condition. However, as respondents were randomized and treated group-wise, respondents were not independent. To correct for this clustering, the sample size had to be raised ($n * C$) according to the following formula: $C = 1 + k * (\text{variance } B^2 / \text{variance } P^2)$. K is the number of respondents per course, variance B is the variance in effects between different POWER trainers, and variance P is the variance in effects between respondents (Van Houwelingen, 1998).

To estimate these unknowns, we used the data from a pilot phase of POWER, resulting in: variance $B = 0.230$, variance $P = 0.873$, youngsters per group = 9. This results in a correction factor C of $1.63 (C = 1 + 9 * (0.053/0.762))$, which results in a minimum of 127 youngsters per condition.

The primary analyses of interest were intervention main effects between the posttest scores of the intervention condition and the control condition. For these analyses, we used linear regression analyses as implemented in Mplus 6.11 (Muthén & Muthén, 1998–2010) and, for each of the outcome variables, tested whether changes over time varied as a function of the intervention, controlling for baseline assessments and nested data. Furthermore, we tested whether intervention effects depended on the level of problem behavior at baseline, gender, ethnicity, or level of implementation by adding interaction terms into the regression analyses (each interaction term was tested in a separate model). For problem behavior, we created a dummy coding 1 for the 20% of youngsters with the highest SDQ total score at baseline and for all others a 0. Gender was coded as 1 for boys and 0 for girls. For ethnicity, we created three dummies, one for being Moroccan, one for being Turkish, and one for being Antillean, Aruban, or Surinamese. For implementation, we coded the dummy 1 if the trainer followed 75–100% of the manual instructions for the group course for youngsters for at least ten lessons, and 0 if this criterion was not fulfilled.

Analyses were performed according to the intention to treat principle. Missing data from participants who dropped out on posttest was imputed, using multiple regression imputation as implemented in Mplus 6.11. All analyses were corrected for cluster effects as participants were nested in groups.

Results

Baseline Comparisons

Descriptive analyses showed that age, gender, education level, and ethnicity were statistically equally distributed between the two

conditions. Table 2 shows the means and standard deviations of all outcome variables at baseline and posttest for the intervention and control condition separately. Two significant differences in the outcome measures were found between the two conditions at baseline. The experimental condition had a lower baseline conduct problem score ($F(1,246) = 4,922, p < .05$) and a lower baseline social activities score ($F(1,246) = 0,042, p < .05$).

As it can be assumed that these variables (i.e. conduct problem score, and social activities score) are correlated to the development of our outcomes measures, these differences were statistically corrected by adding these variables as a covariate in the analyses. With that, we've also added age, gender and educational level as covariates in the model as, although not statistically significant, the intervention group included somewhat more boys, who were a couple of months younger and lesser educated. However, as it is discussed whether (significant) baseline differences has to be treated as covariates (De Boer, 2015), we have also run the unadjusted analyses, i.e. with only the baseline measure of the outcome variable as a covariate in the model. Both analyses lead to the same conclusion. In the following tables and the text we will present the unadjusted analyses.

Table 2. Means and standard deviations of outcome variables at baseline and posttest for intervention and control condition

| Measure | Scale | T | Intervention | | Control | |
|------------------------|-------------------------------------|----------|--------------|------|---------|------|
| | | | M | SD | M | SD |
| Problem behavior | Conduct problems * (Range 0-10) | Baseline | 2.22 | 2.37 | 1.64 | 2.10 |
| | | Posttest | 2.21 | 2.22 | 2.44 | 2.57 |
| | Prosocial behavior (Range 0-10) | Baseline | 7.33 | 2.18 | 7.35 | 1.87 |
| | | Posttest | 7.34 | 2.22 | 7.11 | 2.09 |
| Social marginalization | Activities scale (Range 0-8) | Baseline | 2.23 | 1.62 | 2.60 | 1.59 |
| | | Posttest | 2.46 | 1.71 | 2.20 | 1.57 |
| | Social scale * (Range 0-12) | Baseline | 7.07 | 1.74 | 6.55 | 1.79 |
| | | Posttest | 6.55 | 1.91 | 6.01 | 2.12 |
| Mastery | Number (Range 0-4) | Baseline | 0.75 | 1.04 | 0.56 | 0.87 |
| | | Posttest | 0.63 | 0.98 | 0.52 | 0.90 |
| | Total score (Range 7-35) | Baseline | 25.47 | 4.44 | 25.81 | 4.60 |
| | | Posttest | 25.66 | 4.27 | 25.16 | 4.48 |
| Coping | Active way (Range 7-28) | Baseline | 17.42 | 4.32 | 17.27 | 3.74 |
| | | Posttest | 17.87 | 3.71 | 17.33 | 3.82 |
| | Palliative reaction (Range 8-32) | Baseline | 17.46 | 4.38 | 17.70 | 3.69 |
| | | Posttest | 17.74 | 4.17 | 17.82 | 3.67 |
| | Social support (Range 6-24) | Baseline | 12.92 | 3.74 | 12.87 | 3.36 |
| | | Posttest | 13.29 | 3.60 | 13.40 | 3.41 |
| Reassuring thoughts | Passive reaction (Range 7-28) | Baseline | 12.37 | 3.82 | 12.47 | 3.33 |
| | | Posttest | 12.78 | 4.06 | 13.15 | 3.62 |
| | Reassuring thoughts (Range 5-20) | Baseline | 11.30 | 2.86 | 11.63 | 2.70 |
| | | Posttest | 11.69 | 2.53 | 11.58 | 2.78 |

* significant difference between intervention and control condition at baseline ($\rho < 0.05$)

Intervention Effects - Primary Outcome Measures

In Table 3, the *B* coefficient represents whether changes over time varied as a function of the intervention. The regression analyses showed one significant effect on one of the two indicators of social marginalization: a significant intervention effect ($B = 0.139$, $SE = .059$, $95\% CI = <.023 ; .255>$, $p = .019$, $d = .28$) was found for the activities scale, but no intervention effects were found for the social scale or number of activities. No significant effects were found for the levels of conduct problems and prosocial behavior.

Intervention Effects - Secondary Outcome Measures

In Table 3, the *B* coefficient represents whether changes over time varied as a function of the intervention. The regression analyses showed no

differences in mastery and coping styles at posttest between the two conditions.

Table 3. Regression analyses on outcome variables between intervention and control condition, standardized

| Measure | Scale | B | SE | (95% CI) | p |
|------------------------|----------------------|-------|------|----------------|-------|
| Problem behavior | Conduct problems | -.085 | .098 | (-.277 ; .107) | .383 |
| | Prosocial behavior | .055 | .076 | (-.094 ; .204) | .468 |
| Social marginalization | Activities scale | .139 | .059 | (.023 ; .255) | .019* |
| | Social scale | .052 | .072 | (-.089 ; .193) | .470 |
| Mastery | Number of activities | .024 | .090 | (-.152 ; .200) | .792 |
| | Total score | .076 | .064 | (-.049 ; .201) | .236 |
| Coping | Active way | .066 | .065 | (-.094 ; .204) | .314 |
| | Palliative reaction | .002 | .074 | (-.125 ; .129) | .908 |
| | Social support | -.020 | .064 | (-.145 ; .105) | .755 |
| | Passive reaction | -.043 | .081 | (-.202 ; .116) | .593 |
| | Reassuring thoughts | .044 | .065 | (-.105 ; .193) | .504 |

* significant intervention effect ($p < 0.05$)

Moderation - Gender, Level of Problem Behavior, and Ethnicity
As the target group of POWER is quite diverse in terms of gender, level of problem behavior at baseline, and ethnicity, we performed moderation analyses to study whether these factors are related to the outcomes. The analyses showed that they were not significant moderators.

Moderation - Implementation

In six courses, the content and instructions in the manual were followed 75–100% per session, for at least 10 sessions (high fidelity). In three courses, it was 50–75% followed per session, for at least 10 sessions. In five courses, less than 50% of the content and instructions in the manual were followed per session. Moderation analysis showed that the level of implementation was a significant moderator. Children participating in the six courses implemented with high fidelity benefitted from the course with respect to their conduct problems ($B = -0.115$, $SE = 0.043$, 95% CI = $< -0.199 ; -0.031 >$, $p = .007$, $d = .38$) as well as on their passive reaction coping style ($B = -0.125$, $SE = 0.050$, 95% CI = $< -0.223 ; -0.027 >$, $p = .012$, $d = .31$).

Discussion

In this study, we tested the effectiveness of the empowerment program for adolescent second generation migrants, POWER. Retention in the intervention was good, and POWER significantly increased the time adolescents spent on sports, hobbies, and casual jobs, which was one of our indicators of social marginalization. However, no intervention effects were found on the other primary outcome measures: involvement in social relations and problem behavior. Neither were there any intervention effects on the secondary outcomes; participants' sense of mastery and coping skills. However, when the intervention was implemented with high fidelity, an intervention effect was also found on the level of conduct problems as well on one of the coping skills. Thus, although the overall intervention effect is limited, it is promising that POWER, when implemented with high fidelity, appeared to have some positive effects on both a primary and secondary outcome measure. This indicates that the program might contribute to reducing participants' risk of social marginalization. An important question, however, is under which circumstances these outcomes can be achieved best?

The finding that the fidelity of implementation influences the effectiveness of the program contributes to the growing body of evidence that the implementation process can significantly influence intervention outcomes. It was evident that in this study the intervention as implemented deviated significantly from the prescribed intervention. Some trainers adapted substantial parts of the intervention, and this could have influenced intervention outcomes negatively. Although we cannot support it with objective data, we had the impression that in particular those trainers who were less used to work with systematically designed prevention programs, were the ones who were more inclined to add and remove elements from the course on base of what they believed was needed for their group and with that affected the effectiveness. On the other hand, adapting the intervention does not per se have to mean that the basic principles of the intervention are harmed: working with this specific target group also requires the trainers to be quite flexible. Skipping procedures in order to maintain trust and keep the youngsters

involved can sometimes be more beneficial to the process than maintaining the procedures. Reaching and retaining at-risk adolescent second generation migrants on a voluntary basis for quite an extensive intervention is a very complex task for many professionals. The question, however, is whether and how adaptations affected the intervention effects. A more accurate registration of the implementation process would have been helpful and the program can potentially be further improved by clearly specifying which components of the program can be adapted and which must be delivered as intended.

Although accurate information is missing from the specific program elements that have and have not been adapted, it was clear to us that the course for youngsters was implemented in a much better way than the parents course. The third element of POWER, the community approach, was implemented worst. This was probably affected by the fact that most trainers were trained specifically for this trial and, although they had training experience with this target group, it was the first time they implemented POWER and had to get use to the program. As it is known from the theoretical framework of empowerment that influencing the community is as much important as influencing the individual level this could be an explanation for the relatively modest findings (Rappaport, 1987).

We also like to address the fact that we worked with very different groups in terms of gender, age, ethnic background and difficulties they experience in their live, but used only one set of outcome measures that was chosen before the start of the courses and was the same for all the groups. Unlike an intervention for, let's say depression, POWER focusses on a wide range of problems. It could be a very interesting idea to determine the outcome measures on base of the specific problems in each group separately on base of the anamneses of each particular group and test whether the intervention was effective for the targets set at the beginning of the course. In this way the study design is more in line with the way the intervention is usually implemented and adapted.

Limitations and Strengths

The first limitation is that we used only one type of informant, the youngsters themselves. Information from trainers, parents, or police reports could have given a better insight into the youngsters' development during the intervention - for example, on the risk of learning to become deviant by interacting with deviant peers (Dodge, Dishion, & Lansford, 2006). However, trainers gave the intervention and would therefore not be independent, parents were hard to reach for their consent, let alone to motivate to fill out questionnaires, and collection of police reports would have deterred youngsters from participating. Another limitation is that we had no follow-up data for both conditions. As new behaviors have to be practiced before they become internalized, it cannot be excluded that intervention effects would become more fully developed later on. Table 2 indeed shows that changes for the intervention group are moving pretty consistently in the positive direction. A third limitation is that, despite our attempts to put together a questionnaire that is understandable for our sample, the internal consistency of some scales was quite low. Before the start of this study we tested whether the measures we want to use were comprehensible in a similar group of youngster. Questionnaires that seemed to be too difficult on base of this pilot were omitted. Nevertheless the internal consistency of some scales we finally used was quite low, which was probably affected by the youngsters' low education level, the small sample size and limited knowledge of the accuracy of questionnaires in ethnic groups (Okazaki and Sue, 1995).

Our study contains also several strengths. First, the number of controlled studies on the effectiveness of empowerment programs is very limited, making each additional controlled study a valid contribution to the knowledge on the effects of these programs. The second and third strengths of this study are our initial sample size and the low drop-out rate. Studying intervention effects of program with second generation migrants is usually considered as quite difficult. The fact that we were able to include 248 youngsters and retain 209 of them at posttest should be considered not only as an important achievement of the program but also as a major strength of our study.

Conclusion

In sum, considering that 1) there are as yet no interventions available in the Netherlands that have proved effective in reducing the risk of marginalization in adolescent second generation migrants, 2) the POWER strategy was successful in reaching and retaining the target group, and 3) POWER showed indications of effectiveness provided it was implemented with high fidelity, it is worth investing further in research on this intervention and its implementation. However, more emphasis should be placed on the implementation protocol, and better insight should be gained on the effective ingredients and adaptations.

Acknowledgements

This study was supported by the Dutch Medical Research Council, ZonMW (no. 15700.1002). We would like to thank Hamid Ouali and Ria van Dam for their dedication to the POWER project.

References

- Achenbach, T., Becker, A., Döpfner, M., Heiervang, E., Roessner, V., Steinhausen, H., & Rothenberger, A. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ findings: Research findings, applications and future directions. *Journal of Child Psychology and Psychiatry*, 49, 251–275.
- Alegria, M., Canino, G., Shrout, P. E., Woo, M., Duan, N., Vila, D., ... Meng, X.-L. (2008). Prevalence of mental illness in immigrant and non-immigrant U.S. Latino groups. *American Journal of Psychiatry*, 165, 359–369.
- Barlow, J., Shaw, R., & Stewart-Brown, S. (2004). The effectiveness of parenting programmes for ethnic minority parents. York, UK: Joseph Rowntree Foundation.
- Barrett, P. M., & Ollendick, T. H. (2004). Handbook of interventions that work with children and adolescents: From prevention to treatment. London, UK: John Wiley & Sons.
- Berry, J. W. (2003). Conceptual approaches to acculturation. In K. M. Chun, P. Balls Organista, & G. Marín (Eds.), *Acculturation: Advances in theory, measurement, and applied research* (pp. 17–37). Washington, DC: American Psychological Association.
- Berry, J. W. (2005). Acculturation: Living successfully in two cultures. *International Journal of Intercultural Relations*, 29, 697–712.
- Berry, J. W., Phinney, J. S., Sam, D. L., & Vedder, P. (2010). Immigrant youth: Acculturation, identity and adaptation. *Zeitschrift für Padagogik*, 55, 17–43.
- Bijlstra, J. O., Jackson, S., & Bosma, H. A. (1994). De Utrechtse Copinglijst voor adolescenten [The Utrecht Coping List for adolescents]. *Kind en Adolescent*, 15, 98–109.
- Blom, M., Oudhof, J., Bijl, R. V., & Bakker, B. F. M. (2005). Verdacht van criminaliteit. Allochtonen en autochtonen nader bekeken [Suspect of criminality. A closer look at migrants and autochthons]. Cahier 2005-2. The Hague, The Netherlands: WODC.
- Boon, A. E., De Haan, A. M., & De Boer, S. B. B. (2010). Verschillen in etnische achtergrond van forensische en reguliere jeugd-ggz-clients [Differences in ethnic background of forensic and regular youth mental health care clients]. *Kind en Adolescent*, 31, 16–28.
- Bynner, J. (2006). Childhood risks and protective factors in social exclusion. *Children & Society*, 15(5), 285–301.
- Dagevos, J., Gijsberts, M., & Van Praag, C. (2003). Rapportage minderheden 2003: Onderwijs, arbeid en sociaal-culturele integratie [Minority report 2003: Education, labour and social-cultural integration]. The Hague, The Netherlands: Sociaal en Cultureel Planbureau.
- De Boer, M.R., Waterlander, W.E., Ruijter, L.D.J., Steenhuis, I.H.M., & Twisk, J.W.R. (2015) Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. *J. Behav. Nutr. Phys. Activity*, 12:4
- De Boom, J., Weltevreden, A., Van Wensveen, P., Van San, M., & Hermus, P. (2011). Marokkaanse-Nederlanders 2010 [Moroccan-Dutch 2010]. Rotterdam, The Netherlands: Risbo, Erasmus Universiteit.
- Dijkman, T. (1996). Marginalisering bij jongeren uit etnische minderheden: Een beknopte literatuurstudie naar factoren in het marginaliseringsproces [Marginalization in ethnic minority adolescents: A brief literature study on the factors in the process of marginalization]. Ubbergen, The Netherlands: Uitgeverij Tandem Felix.
- Dodge, K. A., Dishion, T. J., & Lansford, J. E. (2006). Deviant peer influences in programs for youth. Problems and solutions. New York, NY: The Guilford Press.
- Eldering, L., & Knorth, E. J. (1997). Marginalisering van allochtone jongeren en risicofactoren in hun dagelijks leven [Marginalization among migrant youngsters and risk factors in their life]. *Kind en Adolescent*, 18, 2–14.
- Fertman, C. L., & Chubb, N. H. (1992). The effects of a psycho educational program on adolescents' activity involvement, self-esteem, and locus of control. *Adolescence*, 27(107), 517–526.
- Flay, R. F., Graumlich, S., Segawa, E., Burns, J. L., & Holliday, M. Y. (2004). Effects of 2 prevention programs on high-risk behaviors among African American youth. *Archives of Pediatrics & Adolescent Medicine*, 158, 377–384.
- Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology*, 27(1), 17–24.
- Guarnaccia, P. J., & Lopez, S. (1998). The mental health and adjustment of immigrant and refugee children. *Child and Adolescent Psychiatric Clinics of North America*, 7(3), 537–553.

- Gutiérrez, L. M. (1990). Working with women of color: An empowerment perspective. *Social Work*, 35(2), 149-153.
- Huey, S. J. J., & Polo, A. J. (2008). Evidence-based psycho-social treatments for ethnic minority youth. *Journal of Clinical Child and Adolescent Psychology*, 37(1), 262-301.
- Ince, D., & Van den Berg, G. (2010). Overzichtsstudie interventies voor migrantenjeugd [Interventions for young migrants]. Utrecht, The Netherlands: Nederlands Jeugdinstituut.
- Jenissen, R. P. W. (2007). Crimnaliteit, leeftijd en etniciteit. Over de afwijkende leeftijdsspecifieke criminaliteitscijfers van in Nederland verblijvende Antilliaanen en Marokkanen [Criminality, age and ethnicity. About the specific age-related crime figures from Antilleans and Moroccans living in the Netherlands]. *Onderzoek en beleid*, 277 (Report). The Hague, The Netherlands: Boom Juridische uitgevers.
- Konner, V. K., Weisman de Mamani, A. G., Flynn, P. M., & Betancourt, H. (2007). Acculturation and mental health: Current findings and recommendations for future research. *Applied and Preventive Psychology*, 12, 76-96.
- Kumpfer, K. I., Alvarado, R., Smith, P., & Bellamy, N. (2002). Cultural sensitivity in universal family-based prevention interventions. *Prevention Science*, 3(3), 241-244.
- Lee, C. C. (1992). Empowering young black males. Ann Arbor, MI: ERIC/CASS Publications.
- Leijten, P., Raaijmakers, M. A. J., Orobio de Castro, B., & Matthys, W. (2013). Does socioeconomic status matter? A meta-analysis on parent training effectiveness for disruptive child behavior. *Journal of Clinical Child and Adolescent Psychology*, 42, 384-392.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational and behavioral treatment. *American Psychologist*, 48, 1181-1209.
- Martens, E. P. (1999). Minderheden in beeld: SPVA-98 [Minorities examined: SPVA-98]. Rotterdam, The Netherlands: Instituut voor Sociologisch-Economisch Onderzoek.
- Muthén, B., & Muthén, L. (1998-2010). Mplus user's guide. 6th edition. Los Angeles, CA: Muthén and Muthén.
- Orobio de Castro, B., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development*, 73, 916-934.
- Okazaki, S., & Sue, S. (1995). Methodological issues in assessment research with ethnic minorities. *Psychological Assessment*, 7, 367-375.
- Paalman, C. (2013). Exploring behavioral problems in Dutch-Moroccan youth. Doctoral dissertation. Amsterdam, The Netherlands: VU University.
- Parsons, R. J. (1989). Empowerment for role alternatives for low income minority girls: A group work approach. Special Issue: Group work with poor and oppressed. *Social Work with Groups*, 11(4), 27-45.
- Pearlin, L. I., Menaghan, E. G., Lieberman, M. A., & Mullen, J. T. (1981). The stress process. *Journal of Health and Social Behavior*, 22, 337-356.
- Rappaport, J. (1987). Terms of Empowerment/Exemplars of Prevention: Toward a Theory for Community Psychology. *American Journal of Community Psychology*, 15(2), 121-148.
- Reischl, T. M., Zimmerman, M. A., Morrel-Samuels, S., Franzen, S. P., Faulk, M., Eisman, A. B., & Roberts, E. (2011). Youth empowerment solutions for violence prevention. *Adolescent Medicine: State of the Art Reviews*, 22(3), 581-600.
- Scheppers, E., Van Dongen, E., Dekker, J., Geertzen, J., & Dekker, J. (2006). Potential barriers to the use of health services among ethnic minorities: A review. *Family Practice*, 23, 325-348.
- Schonberg, M. A., & Shaw, D. S. (2007). Do the predictors of child conduct problems vary by high- and low-levels of socioeconomic and neighborhood risk? *Clinical Child and Family Psychology*, 10, 101-135.
- Schreurs, P. J. G., Tellegen, B., & Van de Willige, G. (1984). Gezondheid, stress en coping: de ontwikkeling van de Utrechtse Coping Lijst [Health stress and coping: Development of the Utrecht Coping List]. *Gedrag: Tijdschrift voor Psychologie*, 12, 101-117.
- Schreurs, P. J. G., & Van de Willige, G. (1988). Omgaan met problemen en gebeurtenissen: De Utrechtse Copinglijst (UCL) [Dealing with problems and events: The Utrecht Coping List (UCL)]. Lisse: Swets & Zeitlinger.
- SCP(2009). Jaarrapport integratie 2009 [Annual report integration 2009]. The Hague, The Netherlands: Sociaal en Cultureel Planbureau.
- Stevens, G. W. J. M., & Vollebergh, W. A. M. (2008). Mental health in migrant children. *Journal of Child Psychology and Psychiatry*, 49, 276-294.
- Stevens, G. W. J. M., Vollebergh, W. A. M., Pels, T. V. M., & Crijnen, A. A. M. (2007). Parenting and internalizing and externalizing problems in Moroccan immigrant youth in the Netherlands. *Journal of Youth and Adolescence*, 36, 685-695.

Chapter 4

- Tolan, P. H., & McKay, M. (1996). Preventing serious antisocial behavior in inner-city children: An empirically based family intervention program. *Family Relations: Journal of Applied Family and Child Studies*, 45, 145-155.
- Trentacosta, C. J., & Shaw, D. S. (2009). Emotional self-regulation, peer rejection, and antisocial behavior: Developmental associations from early childhood to early adolescence. *Journal of Applied Developmental Psychology*, 30, 356-365.
- Uiterloo, G. (2005a). POWER. Handboek [POWER. Manual]. Utrecht, The Netherlands: Trimbos Institute.
- Uiterloo, G. (2005b). POWER Werkboek [POWER. Workbook]. Utrecht, The Netherlands: Trimbos Institute.
- Uiterloo, G. (2005c). POWER Oudercursus [POWER. Course for parents]. Utrecht, The Netherlands: Trimbos Institute.
- Uiterloo, G. (2005d). POWER. Community Approach. Utrecht, The Netherlands: Trimbos Institute.
- Uiterloo, G. (2005e). POWER. Trainerscursus [POWER. Trainers' course]. Utrecht, The Netherlands: Trimbos Institute.
- Uiterloo, G. (2005f). POWER. Algemene inleiding [POWER. General introduction]. Utrecht, The Netherlands: Trimbos Institute.
- Van Diest, H., Wennink, J., & Uiterloo, G. (2005). De kracht van POWER. Ontwikkeling en uitvoering van een empowermentprogramma voor allochtone jongeren [The strength of POWER. Development and implementation of an empowerment program for migrant youngsters]. Utrecht, The Netherlands: Trimbos Institute.
- Van Dorsselaer, S., de Looze, M., Vermeulen-Smit, E., de Roos, S., Verdurnen, J., ter Bogaert, T., & Volleberg, W. (2010). HBSC 2009. Gezondheid, welzijn en opvoeding van jongeren in Nederland [HBSC 2009. Health, well-being and parenting of youth in the Netherlands]. Utrecht, The Netherlands: Trimbos Institute.
- Van Houwelingen, J. C. (1998). Dwalingen in de methodologie. III. Randomisatie op het niveau van behandelaren [Mistakes in methodology III. Randomization at practitioner level]. *Nederlands Tijdschrift voor Geneeskunde*, 142(29), 1662-1665.
- Veling, W., Selten, J. P., Mackenbach, J. P., & Hoek, H. W. (2007). Symptoms at first contact for psychotic disorder: Comparison between native Dutch and ethnic minorities. *Schizophrenia Research*, 95, 30-38.
- Verhulst F. C., Van der Ende J., & Koot H. M. (1997). Handleiding voor de Youth Self-Report (YSR) [Manual for the Youth Self Report (YSR)]. Rotterdam, The Netherlands: Sophia Kinderziekenhuis, Erasmus MC.
- Weisz, J. R., JensenDoss, A., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist*, 61(7), 671-689.
- Wilson, S. J., Lipsey, M. W., & Soydan, H. (2003). Are mainstream programs for juvenile delinquency less effective with minority youth than majority youth? A meta-analysis of outcome research. *Research on Social Work Practice*, 13(1), 3-26.

Part 2

Conflict of interest

5

Dual roles in research: independence of research within a Dutch database of effective youth interventions

Lars de Winter
Anna-Maria Verhagen
Ferry Goossens

Published as:

Winter, L. de, Verhagen, A.M., Goossens, F.X. (2016). Dubbelrollen in effectonderzoek. Over de onafhankelijkheid van onderzoekers in effectonderzoeken naar interventies uit de databank Effectieve Jeugdinterventies. *De Psycholoog*, 6, 40-53.

Authors' contributions:

LdW collected the data. FG designed the study. LdW and FG together analyzed the data and wrote the manuscript. AMV collected and double checked the data.

Conflict of interest:

LdW, AMV and FG declare they have no conflict of interest.

Abstract

The Dutch database of effective youth interventions shows the efficacy and effectiveness of educational programs for children and adolescents. This paper examines to what extent interventions in the database are based on Dutch studies with independent authors and if the declarations of conflict of interests are correct. In total 86 studies concerning 26 different interventions were examined, it was detailed to what extent the authors had had an interest through their involvement as program developers, license owners or as colleagues of the program developer. Results show that authors were fully independent in 16% of the cases. In 19% of the cases a conflict of interests was correctly declared. The authors argue that more independent studies are needed and stricter rules should be established to improve the transparency of the interest of authors.

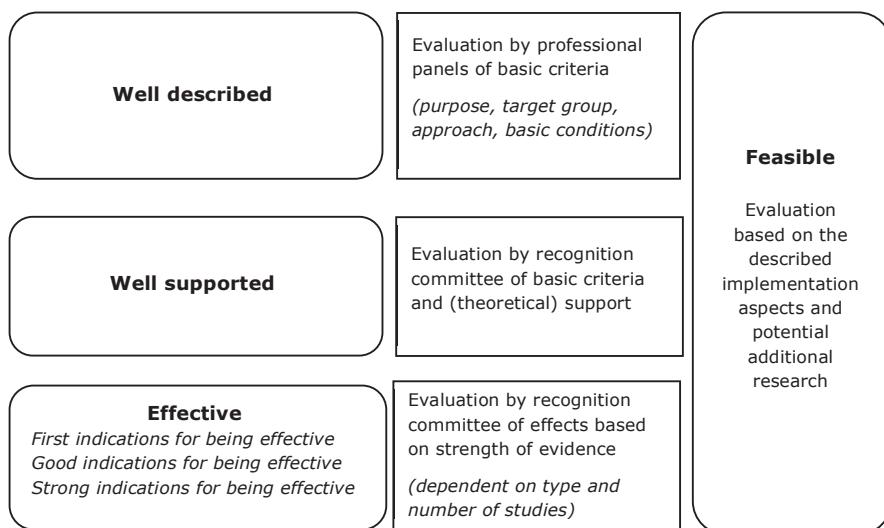
Introduction

The database 'Effectieve Jeugdinterventies' (www.jeugdinterventies.nl) contains a growing number of evaluated and recognized interventions that are focused on the safe and healthy upbringing and raising of youth. This includes programs that target the prevention of smoking (e.g., Actie Tegengif), that seek to improve social and emotional skills (e.g., Kanjertraining, PAD) or that support adolescents in their sexual development (e.g., Lang Leve de Liefde). Professionals working in prevention programs, education, parental support, or Child Services, as well as the funders for youth interventions, utilize this database to select independently evaluated interventions for their professional practice. The evaluation of intervention effects and feasibility takes place with the help of established criteria and is conducted by independent committees. Organizations that want to register an intervention for recognition have to complete an intervention description that substantiates why their intervention should be recognized, using academic (efficacy and effectiveness) studies to support their case. Based on the evidence and support provided, the accreditation committee will give one of the following evaluations: 'good description', 'well supported', or 'effective'. Within the last evaluation category, a distinction is made between 'first indications', 'good indications', and 'strong indications' (see figure 1). A detailed description of all evaluation criteria can be found in Van Dale, Zwikker, Dunnink, Bisseling and Rensen (2013).

The database has become an increasingly important instrument for (healthcare) professionals and (research) funders. In 2014, the database was viewed by 118,343 visitors, an increase of 31% in comparison to the year before (Nederlands Jeugdinstituut, 2014). Municipalities and other funders or buyers of interventions are encouraged to use recognized effective interventions. In addition, ZonMW, a big funding organization for Dutch research, also places a lot of value on the database, often only subsidizing research into interventions that are contained within the 'Effectieve Jeugdinterventies' database. Considering all this, it is of great importance that the evaluation of interventions is conducted with great care and done independently. The accreditation committee therefore adheres to

strict criteria (Van Dale, et al., 2013) in the evaluation of interventions and for their own procedures. One aspect, however, seems to be forgotten: the degree of independence of the authors of the trials that are used by the submitters of an intervention description to support their case and on which the accreditation committee bases their evaluation. The present study examines to what extent Dutch studies, that are submitted as evidence for intervention recognition, are conducted by completely independent researchers and whether a potential conflict of interests is disclosed in the research articles if there is a lack of independence.

Figure 1. Recognition criteria for interventions



- **Well supported:** An intervention is given this classification if the intervention is well described and plausible arguments are made that the intervention can achieve the stated goal.
- **Evidence based on first indications:** An intervention is evidence based on first indications if research with weak or only indicative evidence, such as change research, shows that there is sufficient effect perceived in the execution of the intervention, even if it cannot yet be conclusively proven that this effect is due to the intervention.

- **Evidence based on good indications:** An intervention is effective based on good indications if research with limited evidence shows that this intervention can achieve certain goals in daily practice better than other interventions or no intervention.
- **Evidence based on strong indications:** An intervention is effective based on good indications if based on sufficient research with strong or very strong evidence, it can be shown that this intervention can achieve certain goals in daily practice better than other interventions or no intervention.

(Van Dale, Zwikker, Dunnink, Bisseling en Rensen, 2013.)

Dependent researchers

There are several indications showing that research conducted by non-independent researchers leads to more positive results than research that is conducted by independent researchers (Beelman & Lösel, 2006; Farrington & Welsh, 2003; Gorman & Conde, 2007; Perlis et al., 2005; Petrosino & Soydan, 2005; Reyno & McGrath, 2006; St Pierre, Osgood, Mincemoyer, Kaltreider, & Kauh, 2006; Valentine, et al., 2011; Wilson, Lipsey & Derzon, 2003). In several intervention studies, this lack of independence is given as a possible explanation for the more positive results. There are two possible explanations for why the lack of independence leads to more positive results. The first possibility is that the authors consciously manipulate the research results or (un)consciously make choices during the study that increase the chance of positive intervention effects (*cynical view*). Generally this does not necessarily imply fraud, but rather the making of a series of (un)conscious choices during the study that will benefit the results. For instance, one could think of running multiple analysis-techniques and only reporting on the most 'favorable' technique, choosing to focus on favorable results within subgroups only, or selectively reporting only favorable outcome measures. The second possibility is that the authors or program developers are more aware of the characteristics and theoretical foundation for the intervention, and as a result they know better than others how they can implement the

intervention in the most faithful manner (*high fidelity view*) (Beelman & Lösel, 2006; Borman, Hewes, Overman & Brown, 2003; Eisner, 2007; Eisner 2009; Farrington & Welsh, 2003; Gandhi, Murphy-Graham, Petrosino, Chrismer & Weiss, 2007; Holder, 2009; Littel, 2005; Smith & Glass, 1977; St. Pierre, Osgood, Mincemoyer, Kaltreider & Kauh, 2005; Wilson, Lipsey & Derzon, 2003). However, regardless of which possible explanation is applicable, both explanations have consequences for the predictive value of the observed effects in daily practice. In the case of (un)conscious manipulation, it is clear, but also in the case of the high fidelity view, it is known that the results are difficult to replicate in daily practice. Program developers are generally not able to be present at every implementation of their intervention to ensure the same degree of faithfulness in intervention implementation as was reached during the time of the trial. It is a common fact that interventions are less faithfully implemented in daily practice (Elliot & Mihalic, 2004; Kendall & Southam-Gerrow, 1995; Petrosino & Soydan, 2005; Robinson, Berman & Neimeyer, 1990; Valentine, et al., 2011) and are often less effective than those interventions that are conducted in more controlled conditions (Chambless & Hollon, 1998; Elliot & Mihalic, 2004; Flay, et al., 2005; Petrosino & Soydan, 2005; Valentine, et al., 2011). The results of intervention studies that have been adapted in order to execute the intervention as perfectly as possible, are therefore often poor predictors for the effects of the intervention in the daily practice of professionals.

There are three common groups of non-independent researchers: 1) the researcher who developed the studied intervention him/herself; 2) the researcher who is employed within an organization that holds the license or patent for the studied intervention; and 3) the researcher that did not develop the intervention him/herself, but is employed within the same organization as the program developer (Borman, et al., 2003; Chambless & Hollon, 1998; Eisner, 2009; Elliot & Mihalic, 2004; Eisner & Humphreys, 2012; Petrosino & Soydan, 2005; Valentine, et al., 2011). International research (e.g., Eisner & Humphreys, 2012) has shown that the percentage of publications in which the program developer is one of the authors, is quite high, up to 78% (Gorman & Conde, 2007). This same study shows that in half of the 34 examined interventions for stimulant use or violence

prevention, the program developer (and thus often the researcher) is also the owner or director of the organization that disseminates the intervention, thereby clearly having financial interests (Gorman & Conde, 2007). Whether these percentages are comparable to the situation in the Netherlands is unknown.

Conflict of Interests

One way of providing clarity to the issue of the dual roles of authors is the 'conflict of interest', the 'declaration of interest', or the 'financial disclosure' statement, that researchers need to sign when they submit an article for publication to a journal. This statement provides insight into potential conflict of interests between the primary interests in a study, such as conducting research in an accurate, neutral and valid manner, and the secondary interests, such as the financial interests and the need for professional recognition (Eisner, 2009). Though the manner in which journals handle the potential conflict of interests might differ, disclosing this information is a common rule today.

Database as quality indicator

The involvement of program developers and license-holders is not uncommon in trials and is, among other things, due to the fact that interventions are generally first studied under optimal conditions. In doing so, program developers first study the theoretical concepts that underlie their intervention (the so-called efficacy studies). It is not until a later time that the effects of an intervention are studied in a more natural condition (effectiveness studies) or during the up-scaling phase when the intervention is implemented in daily execution conditions (dissemination studies) (Flay, 2005). All three research-types fulfill an important function in the development of interventions. It is definitely not unwise to first test the theoretical concepts underlying the intervention on a small scale for feasibility, before further developing the idea into intervention material and disseminating it on a greater scale. This, however, does not take away that when an accreditation committee is evaluating an intervention, not only the value of different types of research, but also the independence of the researchers, need to be carefully considered and weighed. Indeed, the ultimate goal of the database is to select interventions that under-

normal circumstances in daily practice would have an effect on the behavior on youth and their caregivers. This calls for a critical evaluation of the submitted evidence for the intervention to be recognized.

That results of research cannot always simply be accepted or adopted has been demonstrated before by Eisner (2007) in relation to a study examining Triple P. By using simple analysis-techniques and only utilizing the data that the authors had presented in the research report, Eisner arrived at a different conclusion. The authors of the original Triple P research report, claimed that Triple P was effective for several variables, however Eisner concluded that there was no systematic effect of the intervention. Eisner attributed the difference in their outcomes to the unclear and insufficient analysis-techniques and so-called selective outcome reporting, or in other words only publishing the significant positive intervention effects and not reporting the non-significant effects of outcome measures. A critical consideration of the submitted evidence by the researchers is thus not an unnecessary control-mechanism and the accreditation committee of the database plays an important role in this consideration.

Research questions

The basis of the present study concerns 26 interventions that have been recognized by the accreditation committee as effective, with first, good, or strong indications of being effective. To be recognized as an effective intervention, the submitters have (in addition to international research publications) presented 86 Dutch research articles in their intervention descriptions to support their case.

- 1) In terms of percentages, to what extent are evidence based preventive interventions for youth studied by independent researchers in terms of not also being the developer of the intervention, being his or her colleague, or working for the licensee?
- 2) What percentage of research articles contain a correct statement disclosing possible conflict of interests?

Method

The database contains research articles that were included in the description of all 26 interventions. The committee had recognized all of these interventions, on March 10th, 2015 for the database 'Effectieve Jeugdinterventies' as effective based on first, good, or strong indications of being effective. The present study only considered those trials that had been conducted within the Dutch practice, assuming thereby that these studies are most predictive for the actual effects of the interventions among Dutch youth. These selection criteria resulted in 86 articles concerning 26 different interventions for children and adolescents (up till 18 years of age) with differing purposes, target groups, and settings of conduction (see Table 1). All research articles can be found in the database descriptions or can be requested from the authors. (N.B. The intervention 'B.Slim beweeg meer, eet gezond', recognized as 'effective based on good indications', disclosed insufficient information regarding the authors of the research report that was presented as evidence, and was therefore not considered within the present study).

The dataset of research articles contained 48 published studies (55.8%), 18 research reports (20.9%), four unpublished studies (4.7%), nine dissertations (10.5%), four bachelor- or master-theses (4.7%), one concept version of a scientific article (1.2%), and two book chapters (2.3%). (N.B. All articles used in the present study can be found in the database descriptions or can be requested from the authors.)

Data collection

For every intervention, it was investigated who the developer(s) were, for which organization he and/or she worked at the time of development, and which organizations had the license or patent for the intervention, to determine if the authors of the research articles were indeed independent. To confirm this, several different sources were consulted, including the intervention description, the research articles provided therein as support, other research articles, the websites of the interventions and other sources on the Internet, such as international intervention databases. Two individuals conducted this process

Table 1. Characteristics of and number of research articles for the investigated interventions

| Intervention | Research articles | Topic | First indications | Age | Target group | Location |
|--------------------------|-------------------|-----------------------------------|-------------------|-------------------------|---------------------|----------|
| B-Fit | 2 | Obesity | 0-18 | Children & Adolescents | School | |
| Families First | 13 | Multi-problem families | 0-18 | Families | Home | |
| Gezond Gewicht Overvecht | 1 | Obesity | 0-18 | Children & Adolescents | Basic services | |
| ITGG | 5 | Behavior- and attachment problems | 0-18 | Children & Adolescents | Healthcare facility | |
| Weet en Beweeg | 1 | Obesity | 4-18 | Children & Adolescents | Healthcare facility | |
| Actie Tegengif | 2 | Alcohol, Drugs and Smoking | 12-18 | Parents | School | |
| Allés Kidzzz! | 5 | Behavioral problems | 4-12 | Primary school students | School | |
| BPTG | 1 | ADHD | 4-12 | Children and Parents | Healthcare facility | |
| DroogBedTraining | 4 | Bedwetting | 4-18 | Children | Home | |
| Kanjertraining | 3 | Social skills | 4-18 | Children & Adolescents | Healthcare facility | |
| Krachtvoer | 3 | Obesity | 12-18 | High school students | School | |
| Lang Leve de Liefde | 1 | Sexuality | 12-18 | High school | School | |
| Minder Boos en opstandig | 3 | Behavioral problems | 4-12 | Children and Parents | Healthcare facility | |
| Negen Maanden Niet | 1 | Alcohol, Drugs and Smoking | N.R. | Pregnant women | Home | |
| Ouder Baby Interventie | 1 | Attachment problems | 0-4 | Children and Parents | Home | |
| PLAYGrounds | 1 | Obesity | 4-12 | Primary school students | School | |
| Praten Met Kinderen | 3 | Parenting problems | 4-18 | Adolescents | Basic services | |
| PAS | 2 | Alcohol, Drugs and Smoking | 12-18 | Adolescents and Parents | School | |
| PAD | 5 | Behavioral problems | 4-12 | Primary school students | School | |
| RealFit | 7 | Obesity | 12-18 | Adolescents | Basic services | |
| Stemmingmakerij | 1 | Depression | 12-18 | Adolescents | Healthcare facility | |
| Vallen Is Ook een Sport | 1 | Falling | 4-12 | Primary school students | School | |
| | | Strong indications | | | | |
| De Dappere Kat | 3 | Anxiety | 4-18 | Children & Adolescents | Healthcare facility | |
| Incredible Years | 3 | Behavioral problems | 4-12 | Parents | Healthcare facility | |
| Taakspeel | 10 | Behavioral problems | 4-12 | Primary school students | School | |
| VIPP-SD | 4 | Behavior- and attachment problems | 0-4 | Parents | Home | |

independently of each other. Their results and findings were compared and when in doubt, the search was extended even further and/or the suspected developers and license-holders of the intervention were contacted to verify the findings. Based on this information, a list was then compiled, per intervention of the developers of the intervention, the organization for which they worked at the time of development and the names of the organizations that had a (Dutch) license or patent for the intervention. The names of these persons and organizations were then compared to the names and organizations of the authors as mentioned in the research articles. A distinction was made in the reporting between the first authors and the co-authors, however for 18 research articles there were no co-authors. For this reason, the total number of co-authors is lower than 86, specifically there were 68 co-authors.

Variables

Program developer

A variety of sources, such as submitted intervention descriptions, intervention material (e.g., handbooks), and intervention websites, were used to determine who had developed the intervention. For interventions developed in another country, both the original developers and the Dutch developers who had adapted or had continued developing the intervention for the Netherlands, were registered as program developers. If the intervention had only been translated into Dutch, the translators were not registered as program developers.

Another possibility was that it became apparent in the research article that the authors of the article had developed the intervention themselves and were the first to research the intervention, as development of an intervention often starts with testing a theoretical concept. In addition to that, often times a (public) handbook for the intervention is not yet available, but the intervention and results are detailed in a scientific article. In these specific cases, the authors of the article were registered as program developers.

License-holder / Intervention owner

An organization was registered as the license-holder or intervention-owner, if based on one of the search strategies utilized in the data collection, it could be determined that an organization either owned the intervention or had the license to conduct the intervention in the Netherlands. (N.B. For six interventions it could not be determined which organization had the license for the intervention.)

Colleague of program developer

For every research article, it was investigated if any of the (co-) authors were colleagues of the program developer. To determine this, it was investigated, using one of the search strategies utilized in data collection, for which (health) organization or university the program developer was employed at the time of the article's publication.

Conflict of Interests

The variable 'conflict of interests' was scored in two steps. First, it was examined whether a conflict of interests was mentioned in the research articles. This was often found in a separate paragraph in the article entitled 'conflict of interest', 'declaration of interest', or 'financial disclosure'. For cases in which such a statement was not present, this was considered as the statement 'no conflict of interest'. When such a statement was present, it was noted which authors had a conflict of interests as being the program developer, the license-holder, or colleague of program developer. Subsequently, it was examined whether the information contained in the statement regarding conflict of interests corresponded to the data that we had collected on the authors. Considering that it is only common/required to declare a possible conflict of interests in research articles, only the 48 published research articles were examined and not the research reports, the unpublished studies, dissertations, etc.

Results

Independence authors

In 72 of the 86 articles (84%) at least one of the (co-)authors was not independent, and in 14 of the 86 articles (16%) all (co-)authors were independent (see table 2). For 69% of the articles, at least one of the authors was also the program developer, for 46% of the articles one (co-)author was employed for the organization that had a license for the intervention or that owned the intervention, and for 57% of the articles one (co-)author was a colleague of the program developer.

Table 2. Research articles with non-independent authors in comparison to the total number of articles, in numbers (n) and percentages (%), per type of independence and according to authorship

| | N | % |
|--|---------|-----|
| Non-independent authors | 72 / 86 | 84% |
| First author non-independent | 68 / 84 | 81% |
| Co-author non-independent | 61 / 69 | 88% |
| Author is program developer | 59 / 86 | 69% |
| First author is program developer | 33 / 86 | 38% |
| Co-author is program developer | 38 / 69 | 55% |
| Author employed for license-holder | 37 / 80 | 46% |
| First author employed for license-holder | 31 / 80 | 39% |
| Co-author employed for license-holder | 26 / 64 | 41% |
| Author is colleague of program developer | 49 / 86 | 57% |
| First author is colleague of program developer | 24 / 86 | 28% |
| Co-author is colleague of program developer | 34 / 69 | 49% |

As is indicated in Table 3, the (co-)authors of all articles concerning interventions with strong indications for being effective were not independent. For 80% of the cases, the program developer was involved in the research article, in 50% of the cases, the license-holder was (co-)author of the research article, and in 70% of the cases, a colleague of the program developer was involved.

Table 3. Studies with non-independent authors in comparison to total number studies, in numbers (n) and percentages (%), according to type of indication for evidence

| | First indications | | Good indications | | Strong indications | |
|-----------------------------|-------------------|-----|------------------|-----|--------------------|------|
| | N | % | N | % | N | % |
| Non-independent authors | 20 / 22 | 91% | 32 / 44 | 73% | 20 / 20 | 100% |
| Program developer | 17 / 22 | 77% | 26 / 44 | 59% | 16 / 20 | 80% |
| Employed for license-holder | 8 / 22 | 36% | 19 / 38 | 50% | 10 / 20 | 50% |
| Colleague of program dev. | 17 / 22 | 77% | 18 / 44 | 41% | 14 / 20 | 70% |

Independence authors to intervention

When examining the intervention evaluations, 19 of the 26 intervention evaluations (73%) in the database were solely based on articles (in the Dutch setting) conducted by non-independent researchers (see Table 4). Three of the five interventions with first indications for being effective (60%) were based solely on dependent articles in the Dutch setting. For interventions with good indications for being effective, 12 of the 17 interventions (71%) were based on dependent articles. Lastly, all interventions with strong indications for being effective were based on dependent articles (100%).

Table 4. Distribution of (published) articles with non-independent authors per intervention

| | Non-independent studies | |
|---------------------------|-------------------------|------|
| | N | % |
| First indications | | |
| B-Fit | 1/2 | 50% |
| Families First | 12/13 | 92% |
| Gezond Gewicht Overvecht | 1/1 | 100% |
| ITGG | 5/5 | 100% |
| Weet en Beweeg | 1/1 | 100% |
| Good indications | | |
| Actie Tegengif | 2/2 | 100% |
| Alles Kidzzz?! | 1/5 | 20% |
| BPTG | 1/1 | 100% |
| DroogBedTraining | 0/4 | 0% |
| Kanjertraining | 3/3 | 100% |
| Krachtvoer | 3/3 | 100% |
| Lang Leve de Liefde | 0/1 | 0% |
| Minder Boos en Opstandig | 3/3 | 100% |
| Negen Maanden Niet | 1/1 | 100% |
| Ouder Baby Interventie | 1/1 | 100% |
| PLAYGrounds | 1/1 | 100% |
| Praten Met Kinderen | 3/3 | 100% |
| PAS | 2/2 | 100% |
| PAD | 5/5 | 100% |
| RealFit | 5/7 | 71% |
| Stemmingmakerij | 1/1 | 100% |
| Vallen Is Ook een Sport | 0/1 | 0% |
| Strong indications | | |
| De Dappere Kat | 3/3 | 100% |
| Incredible Years | 3/3 | 100% |
| Taakspel | 10/10 | 100% |
| VIPP-SD | 4/4 | 100% |

Conflict of interests

In 39 of the 48 published studies (81%), incorrect information had been disclosed concerning a conflict of interests (see Table 5). In addition, in 29 of the 48 articles (60%), there was no statement regarding a conflict of interest, even though there were indications for a conflict of interests. In 10 of the 14 articles (71%), there was a statement regarding conflict of interests (usually indicating that there was no conflict), however this statement proved incorrect as our information did reveal a conflict of interests. Lastly, there were five articles that did not include a conflict of interests statement, but were scored as being correct because our information also did not reveal a conflict of interest.

For 39 of the 43 articles (91%), in which at least one author was not independent, incorrect information was given about a possible conflict of interests. Of these, 29 of the 43 articles did not report a conflict of interest, even though there were indications for a conflict of interest, and in 10 of the 14 articles (71%) the conflict of interest was incorrectly reported (often there was a conflict of interest, but the authors reported that they had no conflict of interest).

Table 5. Frequency of incorrect information of conflict of interests in published studies

| | Incorrect information conflict of interests | | Conflict of interests not reported | | Conflict of interests incorrectly reported | |
|-----------------------------|---|-----|------------------------------------|-----|--|-----|
| | N | % | N | % | N | % |
| Total | 39 / 48 | 81% | 29 / 48 | 60% | 10 / 14 | 71% |
| Non-independent authors | 39 / 43 | 91% | 29 / 43 | 67% | 10 / 14 | 71% |
| Program developer | 37 / 40 | 93% | 27 / 40 | 68% | 10 / 13 | 77% |
| Employed for license-holder | 21 / 24 | 88% | 14 / 24 | 58% | 7 / 10 | 70% |
| Colleague of program dev. | 28 / 31 | 90% | 21 / 31 | 68% | 7 / 10 | 70% |

Discussion

In the present study it was examined to what extent Dutch studies regarding "effective" interventions, as evaluated by the database Effectieve Jeugdinterventies, were conducted by independent researchers. In addition, it was investigated to what extent a conflict of interest was correctly reported in the research articles. For only 16% of the research articles, the authors were completely independent, that is to say that they did not develop the intervention, they were not colleagues with the program developer, and were not employed at the organization that had a license or the patent for the intervention. This is a considerably low percentage, especially considering that this potential conflict of interest was (rarely) mentioned in a conflict of interest statement. The present study shows that in 81% of the published articles there was no statement regarding a conflict of interest, even though this (most probably) should have been included, or there was a statement but it incorrectly reported that there were no conflict of interests. In short, a lot of the studies that are submitted to support the effectiveness of an intervention are submitted by researchers who have a conflict of interests, without this being evident. As is argued in the introduction, this type of research is often a poor predictor for the actual effects of an intervention in daily practice. This is great cause for concern and requires improvement from researchers, research funders, journals, and the database Effectieve Jeugdinterventies (though presumably other intervention databases as well).

The database Effectieve Jeugdinterventies is designated to identify interventions that have an impact in daily practice on health-related problems. It was created to aid professionals and funders in choosing interventions that are actually effective. In this regards, the database serves as an additional control (read: quality test) in the interpretation of trials and in the selection of interventions; a task that requires a lot of expertise and is not feasible for many professionals in their daily work. We, therefore, want to advise the accreditation committees to closely examine the extent of independence present in the studies that are submitted as evidence for intervention recognition. It is important

that they do not base their evaluation solely on the arguments that the intervention owners present, but also actively seek for counterarguments regarding the efficacy and effectiveness. There are several ways in which this could be accomplished, such as the following practical adaptations:

- Request the submitters of intervention descriptions to indicate to what extent the submitted research articles were conducted by independent researchers.
- As a committee, check to what extent the submitted statement regarding independence is correct, as the present study has indicated that including a statement regarding a conflict of interest does not necessarily imply that this statement is correct.
- Ask the submitters if there are also (unpublished) studies that have been conducted in which counter-effects were found.
- Create greater insight into which potential conflict of interests are present for the different interventions and studies for the users of the database.

In addition, we advise the research funders, such as ZonMw, to be alert for possible conflict of interests when granting subsidies for trials. Make the independence of the researchers an explicit point of interest for the evaluation criteria. Are enough guarantees incorporated to ensure that the primary interests in a study, such as the accurate and neutral conduction of research, are the main focus, and that the secondary interests, such as financial interests or professional recognition, receive only minimal focus? Is there, for example, within the research group, at least one party that has no interests in either positive or negative outcomes? This should especially be the case when considering interventions that are already being conducted. In order to be able to predict the effects of an intervention in daily practice, it is not only important to test if (the theory behind) an intervention can lead to a

positive effect, but also that effectiveness studies are conducted that test whether the effects of an intervention are also maintained in daily practice. Too often it still appears as if the goal of efficacy studies is to prove that interventions in the social field do work and everything is geared towards providing evidence for that. However, the results that we then consequently expect to achieve in dealing with health-related problems such as depression, obesity, or alcohol abuse, are severely overestimated.

It is important to emphasize that the present study did not investigate to what extent the non-independence of the authors actually influenced the results of their research. What is interesting to note, however, is that all articles for interventions had strong indications for being effective, were solely based on dependent studies. A possible result of this is that the research outcomes are consequently positive and gain high scores for effectiveness. However, it is not possible to confidently conclude this based on the available data in the present study and future research is necessary to indicate to what extent this might be true. It is also necessary to mention that our plea should not be read as a motion of distrust towards the authors of the research articles. Within the Trimbos Institute, it is also common to do efficacy and effectiveness research into our own interventions and to publish these studies in international journals or to submit these interventions to the database for recognition. In some cases, we conduct the study and analyses ourselves and in other cases we are mainly involved in realizing the implementation of an intervention. Involvement of program developers and license-holders is often unavoidable, though greater transparency regarding everyone's role is definitely desirable. The present study did not take into consideration the strength or type of interest that an author or organization has or could have had. One organization might have a greater interest in getting positive results than another organization, as more money or time were invested in the development of the intervention or could be gained by disseminating the intervention. Moreover, the present study only investigated research articles that were based on Dutch studies, considering that it can be assumed that Dutch studies have the greatest predictive value of the effect of the intervention in the

Netherlands. However, this does mean that independent international studies were not considered in describing the results.

In sum, we consider it beneficial if the committee gave greater attention to the degree of independence in the submitted articles for intervention recognition. How to then incorporate this degree of independence into the evaluation of the intervention is a complicated issue considering that there are different degrees of dependence and conflict of interests. It does seem feasible, though, to include transparency concerning independence in quality evaluations as a first step. More independent intervention research should also be encouraged.

References

- Beelman, A. & Lösel, F. (2006). Child social skills training in developmental crime prevention: Effects on antisocial behavior and social competence. *Psicothema*, 18, 603-610.
- Borman, G. D., Hewes, G.M., Overman,L.T. & Brown, S. (2003). Comprehensive school reform: A meta-analysis. *Review of Educational Research*, 73, 125-230.
- Chamble, D.L. & Hollon, S.D. (1998). Defining Empirically Supported Therapies. *Journal of Consulting and Clinical Psychology*, 66, 7-18.
- Eisner, M. (2007). What Evidence for Prevention? A Critical Re-examination of the Prevention Study "Empowerment in Family and School (eifas): A Randomized Trial". <http://www.vrc.crim.cam.ac.uk/PDFs/re-examinationstudy>
- Eisner, M. (2009). No effects in independent prevention trials: can we reject the cynical view? *Journal of Experimental Criminology*, 5, 163-183.
- Eisner, M. & Humphreys, D. (2012). Measuring Conflict of Interest in Prevention and Intervention Research - A Feasibility Study, in: T. Bliesener, A. Beelmann & M. Stemmler (eds.), *Antisocial behavior and crime. Contributions of developmental and evaluation research to prevention and intervention*. Göttingen, London: Hogrefe international.
- Elliot, D.S. & Mihalic, S. (2004). Issues in Disseminating and Replicating Effective Prevention Programs. *Society for Prevention Science*, 5, 47-53.
- Farrington & D.P. Welsh, B.C. (2003). Family-based Prevention of Offending: A Meta- analysis. *The Australian and New Zealand Journal of Criminology*, 36, 127-151.
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S. G., ... Ji, P. (2005). Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination. *Prevention Science*, 6, 151-175.
- Gandhi, A.G., Murphy-Graham, G., Petrosino, A., Chrismar, S.S. & Weiss, C.H. (2007). The Devil Is in the Details Examining the Evidence for "Proven" School-Based Drug Abuse Prevention Programs. *Evaluation Review*, 31, 43-74.
- Gorman D. M. & Conde E. (2007). Conflict of interest in the evaluation and dissemination of 'model' school-based drug and violence prevention programs. *Evaluation and Program Planning*, 30, 422-429.
- Holder, H. (2009). Prevention programs in the 21st century: what we do not discuss in public. *Addiction*, 105, 578-581.
- Rendall, P.C. & Southam-Gerow, M.A. (1995). Issues in the Transportability of Treatment: The Case of Anxiety Disorders in Youths. *Journal of Consulting and Clinical Psychology*, 63, 702-708.
- Nederlands Jeugdinstituut (2014). *Year Report 2014: moving towards a sustainable and effective transformation*. Utrecht: Nederlands-Jeugdinstituut.
- Perlis, R.H., Perlis, C.S., Wu, Y., Hwang, C., Joseph, M. & Nierenberg, A.A. (2005). Industry Sponsorship and Financial Conflict of Interest in the Reporting of Clinical Trials in Psychiatry. *The American Journal of Psychiatry*, 162, 1957-1960.
- 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.
- Reyno, S. M., & McGrath, P. J. (2006). Predictors of parent training efficacy for child externalizing behavior problems - a meta-analytic review. *Journal of Child Psychology and Psychiatry*, 47(1), 99-111.
- Robinson, L.A., Berman, J.S. & Neimeyer, R.A. (1990). Psychotherapy for the Treatment of Depression: A Comprehensive Review of Controlled Outcome Research. *Psychological Bulletin*, 108, 30-49.
- Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.
- St. Pierre, T.L., Osgood, D.W., Mincemoyer, C.C., Kaltreider, D.L. & Kauh, T.J. (2005). Results of an Independent Evaluation of Project ALERT Delivered in Schools by Cooperative Extension. *Prevention Science*, 6, 305-317.
- Valentine, J.C., Biglan, A., Boruch, R.F., González Castro, F., Collins, L.M., Flay, B.R., ... & Schinke, S.P. (2011). Replication in Prevention Science. *Prevention Science*, 12, 103-117.
- Van Dale, D., Zwikker, M., Dunnink, T., Bisseling, R. & Rensen, P. (2013). *Recognition process for interventions: criteria voor shared quality evaluation 2013-2018*. Movisie/Nederlands Centrum Jeugdgezondheidszorg (NCJ)/Nederlands Instituut voor Sport en Bewegen (NISB)/Nederlands Jeugdinstituut (NJI)/RIVM Centrum Gezond Leven (CGL).
- Wilson, S.J., Lipsey, M.W. & Derzon, J.H. (2003). The Effects of School-Based Intervention Programs on Aggressive Behavior: A Meta-Analysis. *Journal of Consulting and Clinical Psychology*, 71, 136-149.

6

Conflict of interest in school based substance use prevention programs: a meta regression analysis

Ferry Goossens
Simone Onrust
Lars de Winter
Bram Orobio de Castro

Authors' contributions:

FG carried out the study, analyzed the data, and wrote the manuscript. SO collected data, helped analyze the data and revised the manuscript. LdW collected data. BOdC revised the manuscript.

Conflict of interest:

FG, SO, LdW and BOdC declare they have no conflict of interest.

Abstract

Intervention developers and licensees of preventive interventions are frequently involved in intervention evaluations. This study examined to what extent such a conflict of interest is prevalent in evaluation studies of school-based substance use prevention programs. It was tested, using meta-regression analyses of 288 programs, whether this conflict is associated with the study outcomes, and if this is then mediated by involvement of the developers or licensees in the implementation process (high fidelity view), or by the methodological quality of the study (cynical view). Results show that 75% of the evaluations were performed with at least one author having a conflict of interest. The effect size for smoking is significantly larger in studies in which one of the researchers is also the developer of the intervention ($d = -.18$) than in studies in which none of the researchers is the developer ($d = -.09$). The effect size for alcohol use is larger when at least one of the researchers is working for the licensee ($d = -.22$), than when none of the researchers are ($d = -.08$). Concerning the outcomes on drugs no significant differences in effect sizes were found. No evidence was found that the effect was mediated by the variables reflecting the high fidelity view or the cynical view.

Introduction

In order to detect effective school-based preventive interventions for alcohol, tobacco and drug use, a growing number of intervention evaluation studies have been conducted in the last decades (Faggiano et al., 2005; Foxcroft & Tsartsadze, 2011; Norberg, Rezelman, & Lim-Howe, 2013; Cuijpers, 2002). The aim of these studies is essentially to identify interventions that can influence the substance use of youngsters, and subsequently to choose and implement the best interventions on a large scale. However, it is known that program developers and program distributors of preventive interventions are frequently involved in these intervention evaluations (Gorman & Conde 2007; Holder, 2009; Eisner & Humphreys, 2012). Ten years ago, Gorman & Conde (2007) already found that in only 27 of the 246 publications they studied, regarding 34 interventions describing the effects of school-based preventive interventions for drug and violence prevention, the intervention developer was not one of the researchers. This is less than 11%. A close collaboration between intervention developers and intervention researchers can be very helpful in the ongoing process of improving the quality of the intervention. Yet, it has also been advocated that such a conflict of interest is an undesirable situation, as it can positively influence the study outcomes (Eisner, 2009; Petrosino & Soydan, 2005). Indeed, Petrosino and Soydan (2005) for example, found impressive positive effects in crime-prevention studies conducted by developers-as-evaluators, while no effects whatsoever were found in similar independent trials. However, to our knowledge it has never been studied whether this is also true in the field of school-based preventive interventions for alcohol, tobacco and drug use.

Conflict of interest

The term "conflict of interest" refers to "a set of conditions in which professional judgment concerning a primary interest, such as the validity of research, might be influenced by a secondary interest, such as financial gain" (Tobin, 2003, page 1161). Such financial interests may be linked to copyrights, royalties, research funding and income generated from the distribution of programs (Resnik, 2000). Additionally, personal credits for detecting or creating an evidence based intervention can also be a

competing interest (Eisner, 2009). Note, however, that a distinction has to be made between a conflict of interest and a bias. The latter occurs when a researcher's judgment concerning his primary interest (i.e., the production of objective knowledge) has been clearly influenced by a competing interest. In contrast, a conflict of interest exists irrespective of whether the researcher's judgment and behavior can be demonstrated to have been adversely influenced (Resnik, 2000). A conflict of interest exists simply as a condition of the researcher having two competing interests. Whether it affects the study outcomes, is another question.

Cynical view, and high fidelity view

There are basically two possible explanations how a conflict of interest can affect study outcomes: the cynical view and the high fidelity view. The first view holds that the researchers consciously manipulate the research results or (un)consciously make choices during the study that increase the chance of positive intervention effects. This may be fraud, but it is more likely the making of a series of (un)conscious choices during the study that benefit the results. For instance, using *post-hoc* outcome variable selection and then reporting only those outcomes which show positive and statistical significance at any follow-up point, choosing to only evaluate statistical significance rather than practical significance as well, or displaying a selection bias selecting subjects and thereby limiting generalizability. The second possibility is that the program developers are more aware of the characteristics and theoretical foundation of the intervention than independent researchers, and as a result are more knowledgeable on how to implement the intervention in the most faithful manner. The problem here, however, is that program developers are generally not able to be present at every implementation of their intervention to ensure the same degree of faithfulness in intervention implementation as was reached during the time of the trial (Beelman & Lösel, 2006; Borman, Hewes, Overman & Brown, 2003; Eisner, 2007; Eisner 2009; Farrington & Welsh, 2003; Gandhi, Murphy-Graham, Petrosino, Chrismer & Weiss, 2007; Holder, 2009; Littel, 2005; Smith & Glass, 1977; St. Pierre, Osgood, Mincemoyer, Kaltreider & Rauh, 2005; Wilson, Lipsey & Derzon, 2003). Obviously, both cynical and high fidelity processes can co-occur in the same study.

The present study

The first objective of this study was to describe to what extent school-based interventions aiming to prevent or reduce substance use in youth, were studied by researchers that have a conflict of interest. A conflict of interest hereby implying that the researcher was also the developer of the intervention and/or working for the licensee. The second objective of this study was to test if this conflict of interest is associated with the study results, i.e. if there is a positive relation between being a developer-as-researcher and/or licensee-as-researcher and the effect size described in the article. The third aim of this study was to test whether this relation is mediated by the high fidelity view or the cynical view. To study which view is of greater influence, we examined four variables that can be used to characterize one of the views and which could be retrieved from the research articles. Regarding the cynical view: (a) selecting outcome reporting, and (b) not performing intention-to-treat analyses. Regarding the high fidelity view: (a) direct or (b) indirect influence of the researchers in the implementation.

We further tested whether the effect is moderated by the developmental stage of the intervention or the sample size. We did so because the developmental stage of the intervention is related with having a conflict of interest. We distinguish efficacy, effectiveness, and dissemination studies that test a (well-developed) intervention with (commercially) available intervention materials, from studies that mainly test a theoretical model of behavioral change, without the use of such intervention materials. For the latter type of intervention studies, the authors are always scored as also being the intervention developer, and never having a conflict of interest by also being the licensee, as a license model is by definition not applicable to this type of study. We further assume that these theoretical models produce better outcomes. By testing moderation, we check whether the relation will remain significant when only taking 'real' interventions into account. We further test whether the relation depends on the study's sample size, as we assume that studies with smaller sample sizes more often report positive intervention effects than studies with larger sample sizes. We also assume that they are performed more often by researchers with a conflict of interest.

Method

Selection of studies

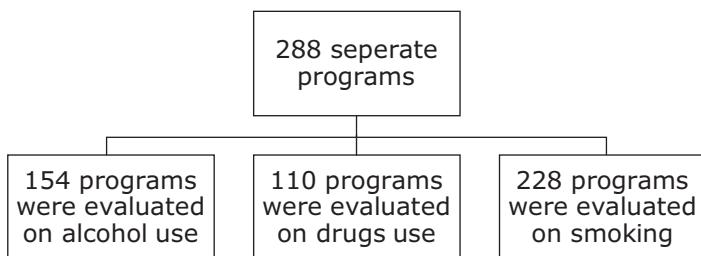
The first step in this research was to create a dataset of studies testing the effects of school-based preventive interventions aiming to prevent or reduce alcohol, tobacco and/or drug use. We extended the dataset of a recent study on school-based programs to prevent or reduce substance use (Onrust et al., 2016). For this study, PsycINFO, Pubmed, ERIC, and COCHRANE were searched for controlled studies of school-based programs, which were evaluated on their effect regarding substance use. Studies were eligible for review when they (a) examined programs delivered in the school setting, (b) targeted elementary, middle or high school students, (c) evaluated behavioral outcomes in smoking, alcohol use and/or drug use, (d) compared the intervention with a control condition, and (e) reported sufficient data to calculate effect sizes. The computer search was restricted to studies published between January 1966 and April 2013.

Studies listed in PsycINFO were retrieved by a combination of key and text words referring to substance use and school-based programs. Results of the computer search were limited by methodology and language (English, German, and Dutch). This resulted in 2839 records. Pubmed was examined using MeSH terms related to substance use in combination with a title and abstract search for school-based programs. Filters were used for methodology and language. This resulted in 1669 records. Results of the computer searches of PsycINFO and Pubmed were combined in Reference Manager. Removal of duplicates resulted in 3848 unique records. Studies registered in ERIC were retrieved by Thesaurus descriptors referring to substance use and Thesaurus descriptors referring to school-based programs. The computer search of the ERIC database resulted in 951 records. Finally, COCHRANE was searched for evaluations of school-based programs. Relevance of the records was evaluated manually, which resulted in two additional titles.

All studies identified by computer searches of the electronic databases were further examined by means of a stepwise screening process. The

first step consisted of evaluating the relevance of the identified records based on the title of the publication. In total, 4801 titles were screened on relevance, resulting in the exclusion of 3426 records. The second step consisted of screening 1416 abstracts, resulting in the exclusion of 959 papers that did not meet the inclusion criteria. The remaining publications were retrieved (32 publications were not available) and the full-text was studied. We also examined the reference lists of all publications of interest, which resulted in 5 additional titles. After studying all publications full text, 144 publications were excluded. Most of them were excluded because these publications provided insufficient data to calculate effect sizes. The remaining 286 publications were included in the analyses. These publications reported on 241 studies, evaluating the effects of 288 distinct programs (see figure 1).

Figure 1. Number of school-based substance use prevention programs in analyses



Note that we distinguish between publications, studies, programs and contrasts between programs and comparator conditions. A single publication can describe multiple studies, and a single study can be described in multiple publications. What interests us most are the evaluations of the contrast between a preventive program and a control condition. Effect sizes are related to the contrast between programs and their control conditions, and ultimately the effect sizes are the units of our analysis. Therefore, multiple programs described in the same publication were coded and analyzed separately. In the same vein, multiple publications evaluating the same contrast were combined into a single effect size. For more details concerning the methodology, see Onrust, Otten, Lammers & Smit (2016).

Variables

Independent variables

Conflict of interest

For every intervention, it was investigated who the developers were and which organizations had the license or patent for the intervention. To do this, several sources were consulted, including the research article, the website of the intervention, and international intervention databases such as NREPP. Based on this information, a list was compiled with the names of the interventions, the names of the developers of the intervention, and the names of the organizations that had a license or patent for the intervention. The names of these persons and organizations were compared to the names and organizations of the authors as mentioned in the research articles. Based on this information we created two variables: (a) at least one author of the research article was also (1) or was not (0) the intervention developer, and (b) at least one author of the research article was also working (1) or not working (0) for the licensee. From these variables we then created a third variable: (c) at least one author was also the developer and/or working for the licensee (1) or none of the authors have a conflict of interest (0). We further registered whether it was the first or a co-author who had the conflict of interest.

Dependent variables

Effect sizes

The effect sizes were examined regarding three different student behavioral outcomes, (a) smoking, (b) alcohol use, and (c) drug use. We included several measures, ranging from the number (or percentage) of participants using substances to the number of cigarettes smoked or alcoholic beverages consumed. If a single study reported multiple outcome measures per outcome category, these results were combined into a single effect size. For every comparison between a school-based program and a control condition, we calculated one effect size (standardized mean difference, also known as Cohen's d) per outcome

category (smoking, alcohol use, and drug use), if these corresponding outcome measures were available. Cohen's d was preferably calculated using the means and standard deviations of both the program group and the control group (at post-test). If means and standard deviations were not reported, we used statistics that were reported for the test between the conditions (for instance a p or t-value). In case of dichotomous outcomes, odds ratios were calculated, and these were converted to standardized effect sizes (Chinn, 2000). All effect sizes were calculated using procedures implemented in the software program Comprehensive Meta-Analysis (version 2.0; Biostat, Englewood, New Jersey).

Effect sizes of zero indicated that there was no difference between the included program and the control condition. Negative effect sizes indicated that students in the program condition were less engaged in substance use than students in the control condition. According to Lipsey and Wilson (1993), a standardized effect size of less than -.32 corresponds to a small effect, effect sizes between -.32 and -.55 correspond to medium effect sizes and effect sizes larger than -.55 correspond to large effects.

Mediators

Direct and indirect involvement in the implementation

To check if the intervention was implemented by the researchers, we read every article full text, including the acknowledgements, searching for information describing whether the authors had implemented the intervention themselves. We scored the variable direct implementation (1) when the researchers implemented the intervention directly to the target group, and (0) if this criteria was not fulfilled. We scored the variable indirect implementation (1) when the intervention was implemented from the researchers to an intermediate group (who in turn implemented the intervention to the target group).

Selective outcome reporting

We checked in every article if there was selective outcome reporting, checking whether all pre-specified substance use outcome measures were also reported in the results section. We additionally checked the

design article, if it was available for the intervention study, as well as trial registers when these were referred to in the article. We scored this variable (1) if there was clearly selective outcome reporting in the article. We scored (0) if it was free of selective outcome reporting, or if we couldn't verify that there was selective outcome reporting.

No intention-to-treat

We coded this variable (0) if analyses were conducted according to the intention-to-treat principle, and less than twenty percent of data was missing. We scored this variable (1) if this condition was not fulfilled.

Moderators

Developmental stage of the intervention

We coded this variable (0) if the study clearly described an efficacy, effectiveness or dissemination trail in which a (well-developed) intervention was tested with (commercially) available intervention materials. This means that the intervention had for instance a name, an intervention website, and/or materials that could be bought or were available at the researchers. We coded this variable (1) if the study tested a theoretical model of behavioral change without the use of such an intervention and/or intervention materials.

Sample size

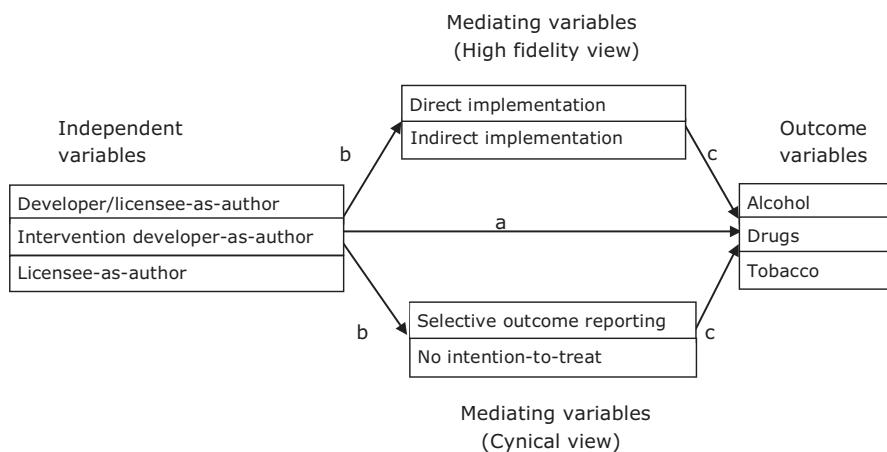
This variable was coded (1) when the total sample size was smaller or equal to 250 participants, and (0) when the sample size was larger than 250 participants.

Analysis

We first calculated the percentages of programs in which one of the researchers was also the developer of the intervention and/or working for the licensee. As a second step, we calculated the pooled effect sizes per outcome variable (i.e., alcohol use, drug use, and smoking), and tested whether the outcome variables could be predicted by the independent variables, using meta-regression analyses (line a in figure 2).

For those relations in which the regression coefficient (β) was significantly different from zero, we performed mediation analyses in three steps (Baron & Kenny, 1986). For each independent variable that predicted an outcome measure, we first tested whether it predicted one of the mediating variables, using logistic regression analyses (lines b). If this was found, we then tested whether this mediating variable predicted the outcome variable, using meta-regression analyses (lines c). If this was found, we as a final step tested whether the relation between the predicting independent variable and the outcome variable was mediated by the mediating variable, using meta-regression analyses that included the outcome variable, the independent variable, and the mediating variable. By doing so we tested whether the significant relations between the independent variables and the outcome variables (line a) were mediated by the high fidelity view and/or the cynical view.

Figure 2. Model of mediation



It was further tested whether the relations between the independent variables and the outcomes variables were moderated by the developmental stage of the intervention or the sample size. Before doing so, we first checked whether studies with a small sample sizes were more often performed by researchers with a conflict of interest, and whether studies with a small sample size and studies testing behavioral change models reported higher effect sizes. If this was found, we created an interaction variable for the dichotomous moderating variable and the

dichotomous independent variable. The outcome variable, moderating variable, independent variable, and interaction variable were then examined using a meta-regression model.

Statistical software

All analyses were performed in Stata (version 12; StataCorp, Texas) using the downloadable procedure "meta-reg".

Results

Conflict of interest

In table 1, the percentage of researchers with a conflict of interest in studies regarding school-based substance use prevention programs is presented. In total, 75 percent of the evaluations were performed with the first author and/or the co-author also being the developer of the intervention and/or working for the licensee. In 69 percent of the studies the intervention developer was one of the researchers and in 11 percent someone working for the licensee was one of the researchers.

Table 1. Percentage of conflict of interest in studies regarding school-based substance use prevention programs

| | All programs N=288 | | Alcohol N=154 | | Drugs N=110 | | Tobacco N=228 | |
|------------------------------------|-----------------------|-----|------------------|-----|----------------|-----|------------------|-----|
| | N | % | N | % | N | % | N | % |
| Conflict of interest | 215 | 75% | 116 | 75% | 80 | 73% | 165 | 72% |
| First author is developer/licensee | 178 | 62% | 93 | 60% | 63 | 57% | 134 | 59% |
| Co-author is developer/licensee | 167 | 58% | 82 | 53% | 59 | 54% | 132 | 58% |
| Developer-as-researcher | 200 | 69% | 106 | 69% | 75 | 68% | 154 | 68% |
| First author is developer | 161 | 56% | 82 | 53% | 57 | 52% | 122 | 54% |
| Co-author is developer | 148 | 51% | 73 | 47% | 53 | 48% | 118 | 52% |
| Licensee-as-researcher | 32 | 11% | 19 | 12% | 14 | 13% | 25 | 11% |
| First author is licensee | 21 | 7% | 13 | 8% | 9 | 8% | 16 | 7% |
| Co-author is licensee | 31 | 11% | 18 | 12% | 13 | 12% | 24 | 11% |

Relation between conflict of interest and effect size

Table 2 presents the pooled effect sizes, reported as Cohen's d , including the 95% confidence interval per outcome variable. In general, the interventions have a mean effect size of $d = -.10$ for alcohol, $d = -.12$ for drugs, and $d = -.15$ for tobacco, indicating that there are small intervention effects overall (Lipsey & Wilson, 1993).

Table 2. Pooled effect sizes in Cohen's d and the 95% confidence interval per outcome and per type of conflict of interest

| | Alcohol | | Drugs | | Tobacco | |
|----------------------------|---------|---------------|-------|---------------|---------|---------------|
| | d | (95% CI) | d | (95% CI) | d | (95% CI) |
| All programs | -.10 | (-.07 ; -.13) | -.12 | (-.09 ; -.16) | -.15 | (-.12 ; -.18) |
| No conflict of interest | -.11 | (-.06 ; -.16) | -.16 | (-.09 ; -.22) | -.08* | (-.05 ; -.11) |
| Conflict of interest | -.10 | (-.06 ; -.13) | -.11 | (-.07 ; -.15) | -.18* | (-.14 ; -.21) |
| No developer-as-researcher | -.13 | (-.08 ; -.17) | -.14 | (-.09 ; -.21) | -.09* | (-.06 ; -.13) |
| Developer-as-researcher | -.09 | (-.05 ; -.12) | -.11 | (-.06 ; -.15) | -.18* | (-.14 ; -.21) |
| No licensee-as-researcher | -.08** | (-.05 ; -.11) | -.12 | (-.08 ; -.17) | -.14 | (-.11 ; -.17) |
| Licensee-as-researcher | -.22** | (-.13 ; -.29) | -.10 | (-.05 ; -.14) | -.19 | (-.11 ; -.27) |

* Significantly different from mean effect size of all tobacco programs at $p > .05$

** Significantly different from mean effect size of all alcohol programs at $p > .01$

Analyses showed that the effect size of smoking is significantly larger ($d = -.18$ versus $d = -.08$) when one of the researchers is also the developer and/or working for the licensee of the intervention, than in studies where none of the researchers are also the developer and/or working for the licensee ($\beta = -.08$; $SE = .03$; 95% $CI = < -.13 ; -.02 >$; $p = .012$). The effect size of smoking is significantly larger ($d = -.18$ versus $d = -.09$) when at least one of the researchers is the developer of the intervention ($\beta = -.06$; $SE = .03$; 95% $CI = < -.12 ; -.01 >$; $p = .029$). The effect size of alcohol use ($d = -.22$ versus $d = -.08$) is significantly larger when at least one of the researchers is working for the licensee ($\beta = -.13$; $SE = .04$; 95% $CI = < -.21 ; -.05 >$; $p = .002$). No significant differences in the effect size were found between having or not having a conflict of interest for the drugs outcome.

Mediation of the relation between conflict of interest and effect size

Table 3 shows whether being involved in the implementation as an author, practicing selective outcome reporting, and not performing intention-to-treat analyses, is in general more prevalent in studies with

authors who have a conflict of interest than in studies with authors who not have a conflict of interest. The table shows that indirect implementation was predicted by having a conflict of interest for the smoking outcome variable ($\beta = 1.25$; $SE = .44$; 95% CI = < .39 ; 2.10 >; p = .004). Indirect implementation was also predicted by the developer being the researcher for the smoking outcome variable ($\beta = 1.35$; $SE = .41$; 95% CI = < .54 ; 2.16 >; p < .001). Not performing intention-to-treat analyses was predicted by having a conflict of interest for the smoking outcome variable ($\beta = 0.79$; $SE = .39$; 95% CI = < .24 ; 1.55 >; p = .043). The variables direct implementation and selective outcome reporting were not predicted by any of the independent variables. Only indirect implementation and not performing intention-to-treat analyses were therefore tested as mediators in the next step.

Table 3. Independent variables per mediating variables

| | High fidelity view | | Cynical view | | |
|-------------------------|-----------------------|-------------------------|------------------------|-----------------------|------|
| | Direct implementation | Indirect implementation | Sel. outcome reporting | No intention to treat | % |
| | N | % | % | % | % |
| <i>Smoking</i> | | | | | |
| No conflict of interest | 63 | 6% | 11%** | 10% | 78%* |
| Conflict of interest | 165 | 10% | 30%** | 15% | 88%* |
| <i>Smoking</i> | | | | | |
| No developer-as-res. | 74 | 5% | 11%*** | 9% | 80% |
| Developer-as-res. | 154 | 11% | 32%*** | 15% | 88% |
| <i>Alcohol</i> | | | | | |
| No licensee-as-res. | 135 | 8% | 23% | 10% | 76% |
| Licensee-as-res. | 19 | 0% | 32% | 21% | 90% |

* Significant at $p < .05$

** Significant at $p < .01$

*** Significant at $p < .001$

Table 4 shows that the effect sizes of smoking did not significantly vary depending on whether the authors were or were not indirectly involved in the implementation. The effect sizes also did not vary depending on the intention-to-treat variable. It can be concluded that none of the variables reflecting the high fidelity view or the cynical view mediate the relation between having a conflict of interest and variance in the effect sizes.

Table 4. Outcomes variables (smoking and alcohol effect sizes) per mediating variables (implementation & selective outcome reporting)

| | Smoking | | | Alcohol | | |
|--------------------------------|---------|------|---------------|---------|------|---------------|
| | N | d | 95% | N | d | 95% |
| <i>High fidelity view</i> | | | | | | |
| Direct implementation | 21 | -.17 | (-.11 ; -.23) | 11 | -.11 | (-.03 ; -.18) |
| No direct implementation | 207 | -.15 | (-.12 ; -.18) | 143 | -.10 | (-.07 ; -.13) |
| Indirect implementation | 57 | -.18 | (-.10 ; -.25) | 37 | -.10 | (-.02 ; -.19) |
| No direct implementation | 171 | -.13 | (-.11 ; -.16) | 117 | -.10 | (-.07 ; -.13) |
| <i>Cynical view</i> | | | | | | |
| Selective outcome reporting | 30 | -.14 | (-.07 ; -.21) | 17 | -.10 | (-.04 ; -.14) |
| No selective outcome reporting | 198 | -.15 | (-.12 ; -.18) | 137 | -.10 | (-.07 ; -.13) |
| No intention-to-treat | 195 | -.15 | (-.12 ; -.18) | 119 | -.11 | (-.08 ; -.15) |
| Intention-to-treat | 33 | -.13 | (-.07 ; -.18) | 35 | -.06 | (-.01 ; -.12) |

Moderation of the relation between conflict of interest and effect size

Developmental stage of the intervention

Our study included 61 interventions (21%) that were basically testing theoretical models of behavioral change. In these studies the researchers were always the intervention developers. Contrary, in these studies there was never a conflict of interest with regards to the licensee being the researcher, as a licensee model was not applicable for this type of interventions. We therefore tested whether the intervention's developmental stage was a moderator for the relation between conflict of interest and effect size (see table 5).

The analyses further showed that the relation between a conflict of interest and the effect size was moderated by the developmental stage of the intervention for the drugs outcome ($\beta = .12$; $SE = .05$; $95\% CI = < .023 ; .22 >$; $p = .016$). The direction was, however, contrary to our hypotheses, as studies with the theoretical models for the drugs outcome had lower effect sizes than the more developed interventions.

Table 5. Pooled effect sizes in Cohen's d and the 95% confidence interval per outcome, for studies testing theoretical models versus interventions

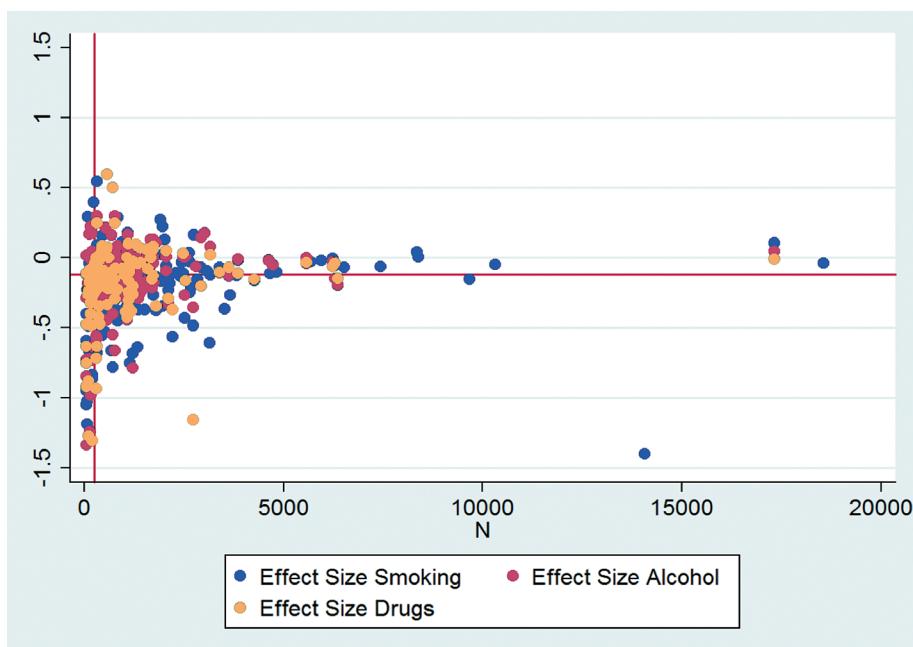
| | Alcohol | | Drugs | | Tobacco | |
|-------------------|---------|---------------|-------|---------------|---------|---------------|
| | d | (95% CI) | d | (95% CI) | d | (95% CI) |
| Theoretical model | -.07 | (.00 ; -.13) | -.01* | (.07 ; -.10) | -.21 | (-.12 ; -.29) |
| Intervention | -.11 | (-.07 ; -.14) | -.14* | (-.10 ; -.18) | -.14 | (-.11 ; -.17) |

* Significant at $p < .05$

Sample size

Figure 3, in which the effect sizes are shown on the y-axis and the sample sizes on the x-axis, shows that in studies with a sample size of $n \leq 250$, larger effect sizes are more often reported than smaller or positive effect sizes (= a counterproductive intervention effect). Instead of a normal distribution around the mean effect size (as a reference the mean effect size for drugs is presented by the horizontal line), there are relatively few studies in the group of studies with a $n \leq 250$ (left of the vertical reference line) that report an effect size smaller than the mean effect size.

Figure 3. Funnel plot, effect sizes (d) on the y-axis and the sample size (n) on the x-axis



The effect sizes in studies with $n > 250$ (see table 6) are in general lower than the effect sizes of all studies together (see table 2). The mean effect sizes for the alcohol programs, the drugs programs, and the tobacco programs are all .02 lower, relative to those in table 2. All 95% confidence intervals of the effect sizes remain lower than 0, indicating that all effects sizes remain statistically significant.

Table 6. Pooled effect size in Cohen's d and 95% confidence interval, $n > 250$

| | Alcohol | | Drugs | | Tobacco | |
|----------------------------|---------|---------------|-------|---------------|---------|---------------|
| | d | (95% CI) | d | (95% CI) | d | (95% CI) |
| All programs | -.08 | (-.05 ; -.11) | -.10 | (-.06 ; -.14) | -.13 | (-.10 ; -.16) |
| No conflict of interest | -.10 | (-.05 ; -.16) | -.16 | (-.08 ; -.23) | -.08 | (-.05 ; -.11) |
| Conflict of interest | -.08 | (-.04 ; -.11) | -.08 | (-.04 ; -.12) | -.15 | (-.11 ; -.18) |
| No developer-as-researcher | -.12 | (-.07 ; -.17) | -.14 | (-.08 ; -.20) | -.09 | (-.06 ; -.12) |
| Developer-as-researcher | -.07 | (-.03 ; -.10) | -.08 | (-.04 ; -.13) | -.15 | (-.11 ; -.18) |
| No licensee-as-researcher | -.06 | (-.03 ; -.09) | -.10 | (-.06 ; -.14) | -.12 | (-.09 ; -.15) |
| Licensee-as-researcher | -.20 | (-.13 ; -.29) | -.09 | (-.04 ; -.13) | -.19 | (-.11 ; -.26) |

The results of the analyses revealed that sample size did not moderate the relation between conflict of interest and effect size (no results shown).

Discussion

The first purpose of this study was to examine to what extent school-based interventions, that aim to prevent or reduce substance use in youth, are studied by researchers that have a conflict of interest. In total 75% of the evaluations were performed with one of the authors also being the intervention developer and/or working for the licensee. In 11% of the studies, one of the licensees was also one of the researchers, and in 69% of the studies the intervention developer was also one of the researchers. Percentages were pretty consistent among all three outcome measures, i.e. alcohol use, drug use, and smoking. Importantly, conflict of interest is related with the effect size. The mean effect size of smoking drops from $d = -.18$ to $d = -.09$ when the none of the researchers are also the developer of the intervention. For alcohol use the mean effect size drops from $d = -.22$ to $d = -.08$ when none of the researchers are working for the licensee.

Having a conflict therefore seems to be related to intervention study outcomes in the field of school-based substance use prevention programs. Yet, this effect was only found for one of the smoking and alcohol outcomes, and not for the drugs outcome. We have no logical explanation why these effects were found on the alcohol and tobacco

outcomes and not on the drugs outcomes. The subsample of drug interventions was indeed the subsample with the smallest number of studies, relative to the studies with outcomes for alcohol and tobacco use, yet it is not likely that this is the reason for not finding an association as descriptive data in table 2 also did not support our hypothesis with regards to the drugs outcome. On the contrary, the descriptive data provided evidence that was directly opposite to our hypotheses. It appears that having a conflict of interest does not always influence intervention study outcomes equally. Nonetheless, the variance in intervention outcomes is still remarkable. This raises two important questions 'How is this possible?' and 'Is this a problem?'

In order to answer the first question, we tested whether the effect could be explained by the high fidelity view or the cynical view. Having a conflict of interest does indeed significantly predict being indirect involved in the implementation, but only with regards to smoking. Having a conflict of interest did also predict not performing intention-to-treat analyses, but again only for smoking. Although direct implementation and selective outcome reporting were not predicted by conflicts of interest, the descriptive data was, however, found to consistently be in the hypothesized direction. That having a conflict of interest is mediated by variables reflecting the high fidelity or the cynical view, is therefore not an illogical idea, yet, none of these variables predicted variance in the outcome variables. Moreover, an unmeasured third variable (such as funding source) might also account for any relations between conflict of interest and study design. Therefore it cannot be concluded, based on this study, that these variables mediate the relation between having a conflict of interest and finding larger effect sizes. Considering this, the question "is this a problem" is a little harder to answer.

There are now two lines of argument that can plausibly be followed. The first argues that the higher effect sizes in studies where the intervention developer and/or licensees are involved, are caused by a better quality and quantity of the implementation. There is evidence that interventions implemented with, for example, more fidelity, completeness or dosage, will produce better outcomes (Durlak & DuPre, 2008). Therefore, if

intervention developers and/or licensee-as-authors produce a better delivery of the intervention, and thereby better intervention outcomes, then we have to be sure that regular implementation of the intervention is at the same level as it was during the trial. If this is not the case, the effect sizes as found in the effect studies of the intervention will not be reproduced in daily practice. Realizing high levels of implementation quantity and quality may be expensive or time-consuming, but it will possibly also pay a profit as it produces better outcomes. In this scenario, studies with a conflict of interest indeed produce outcomes that can be expected during high quality routine implementation of the interventions.

However, it is also known that during routine implementation of interventions, the quantity and quality of the implementation is lower than during trials (Elliot & Mihalic, 2004; Kendall & Southam-Gerrow, 1995; Petrosino & Soydan, 2005; Robinson, Berman & Neimeyer, 1990; Valentine, et al., 2011). It can be questioned whether the same levels of ownership, compassion and expertise regarding the intervention can reasonably be expected from health professionals who also have many other demanding tasks, as can be expected from the intervention developer, who is an expert regarding the intervention. Following this line of reasoning, studies in which none of the authors have a conflict of interest, are probably better estimators of the effects that can be expected during routine implementation of school-based substance use prevention programs. The effects sizes in these studies are indeed lower but still significantly different from 0, even when checked for publication bias, which was prevalent in the studies with a sample size equal to or smaller than 250.

More research on this topic, however, is definitely needed to provide evidence that can verify one of these two explanations. Or at least to be sure that the cynical view is not true. If that view is true, the field of school-based substance use prevention programs has a much bigger problem that needs to be solved then raising the quality of the implementation.

Strengths and Limitations

Although we studied 288 programs, this number of observations is still limited for moderation and mediation analyses. This could possibly explain why indicators of neither the high fidelity view nor the cynical view were detected as a mediating variable. Considering the relatively small numbers of studies in which selective outcome reporting was detected for example, it is possible that there was not enough power for differences to gain significance. Moreover, our indicators of both views were fairly limited as they only concerned readily available proxies of implementation and study quality. Studying this phenomenon with more direct indices would be helpful in answering the question regarding mediation. Another limitation of this study was that the number of variables tested supporting one of the two views was limited, as was their accuracy. More information regarding the quantity and the quality of the implementation process would have been helpful herein. Yet, the implementation process is, in general, poorly described in research articles. This is unfortunate, as the implementation is equally important for the trial outcomes as the intervention itself. The combination of the intervention and the implementation is what in the end produces the intervention effects. We also have to mention that the cut-off point for the funnel plot was made based on visual inspection of the plot. To our knowledge this is, however, not uncommon in reading out funnel plots.

A strength of this study is that this is the first attempt, in school-based substance use prevention programs, to not only report the prevalence of conflicts of interest, but also to study the underlying mechanisms. Although this last goal didn't result in significant results, the results from this study do remain interesting. The differences in effect sizes between studies with and without a conflict of interest were statistically significant. It was additionally found that the variance in outcomes remained significant even when excluding studies with a small sample size and studies that reported on interventions that were at a premature stage of intervention development. More research on this topic is needed to better understand what the benefits and disadvantages are of the involvement of intervention developers and licensees in intervention evaluations.

Although it could not be confirmed whether the high fidelity view or the cynical view mediated the variance in intervention outcomes due to having a conflict of interest, there are enough indications to continue putting effort in both enhancing the fidelity of the implementation of interventions and reducing biases in research methodology. It is known, in general, that interventions that are implemented with low fidelity lose effectiveness, and that a series of methodological choices in favor of the intervention does translate into finding interventions that are really effective. It can also be concluded that it does indeed matter who the researcher of the intervention is. Full transparency regarding the implementation process of the intervention during the study, as well regarding whether the authors of intervention trials have a conflict of interest or not, is therefore of great importance.

Conclusions

Having a conflict of interest is very common in studies at school-based interventions that aim to prevent or reduce substance use in youth. In particular intervention developers are often one of the authors of the study paper. Moreover, such a conflict of interest is related with the effect size. The mean effect size of smoking is significantly lower when none of the researchers is also the developer of the intervention, and the mean effect size for alcohol use is significantly lower when none of the researchers is working for the licensee. The effect sizes remain, however, significant from zero. More research on this topic is needed to get a better idea of the actual effects of, in particular, well-developed and widely available school-based interventions for the prevention and reduction of substance use.

References

- Baron, R.M. & Kenny, D.A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research - Conceptual, Strategic, and Statistical Considerations, *Journal of Personality and Social Psychology*, Vol. 51(6), pp.1173-1182.
- Beelman, A. & Lösel, F. (2006). Child social skills training in developmental crime prevention: Effects on antisocial behavior and social competence. *PsycotHEMA*, 18, 603-610.
- Borman, G.D., Hewes, G.M., Overman, L.T. & Brown, S. (2003). Comprehensive school reform: A meta-analysis. *Review of Educational Research*, 73(2), 125-230.
- Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, 19, 3127-3131.
- Cuijpers, P. (2002). Effective ingredients of school-based drug prevention programs. A systematic review. *Addictive Behaviors*, 27, 1009-1023.
- Durlak, J.A. & DUPRE, E.P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am. J. Community Psychol.* 41, 327-350.
- Eisner, M. (2009). No effects in independent prevention trials: can we reject the cynical view? *Journal of Experimental Criminology*, 5, 163-183.
- Eisner, M. & Humphreys, D. (2012). Measuring Conflict of Interest in Prevention and Intervention Research - A Feasibility Study (draft). Final version published in: T. Bliesener, A. Beelmann & M. Stemmler (eds.), *Antisocial behavior and crime. Contributions of developmental and evaluation research to prevention and intervention*. Göttingen, London: Hogrefe International.
- Faggiano, F., Vigna-Taglianti, F.D., Versino, E., Zambon, A., Borraccino, A., & Lemma, P. (2005). School-based prevention for illicit drugs' use. Cochrane Database of Systematic Reviews, CD003020.
- Farrington & D.P., Welsh, B.C. (2003). Family-based Prevention of Offending: A Meta-analysis. *The Australian and New Zealand Journal of Criminology*, 36, 127-151.
- Foxcroft, D.R., & Tsirtsadze, A. (2011). Universal school-based prevention programs for alcohol misuse in young people. Cochrane Database of Systematic Reviews, CD009113.
- Gandhi, A.G., Murphy-Graham, G., Petrosino, A., Chrismer, S.S. & Weiss, C.H. (2007). The Devil Is in the Details Examining the Evidence for "Proven" School-Based Drug Abuse Prevention Programs. *Evaluation Review*, 31, 43-74. Doi: 10.1177/0193841X06287188.
- Gorman, D. M. (2006). Conflicts of interest in the evaluation and dissemination of drug use prevention programs. In J. Kleinig, & S. Einstein (Eds.), *Intervening in drug use: Ethical challenges* (pp. 171-187). Huntsville: Office of International Criminal Justice.
- Gorman, D.M. & Conde, E. (2007). Conflict of Interest in the Evaluation and Dissemination of "Model" School-based Drug and Violence Prevention Programs. *Evaluation and Program Planning*, 30, 422-429. Doi: 10.1016/j.evalprogplan.2007.06.004.
- Holder, H. (2009). Prevention programs in the 21st century: what we do not discuss in public. *Addiction*, 105, 578-581. doi:10.1111/j.1360-0443.2009.02752.x
- Kendall, P.C. & Southam-Ceron, M.A. (1995). Issues in the Transportability of Treatment: The Case of Anxiety Disorders in Youths. *Journal of Consulting and Clinical Psychology*, 63, 702-708.
- Lipsey, M.W., & Wilson, D.B. (1993). The efficacy of psychological, educational and behavioral treatment. *American Psychologist*, 48, 1181-1209.
- Littel, J.H. (2005). Lessons from a systematic review of effects of multisystemic therapy. *Children and Youth Services Review*, 27, 445-463. Doi: 10.1016/j.chlyouth.2004.11.009.
- Norberg, M.M., Kezelman, S., & Lim-Howe, N. (2013). Primary prevention of cannabis use: A systematic review of randomized controlled trials. *Plos One*, 8, e53187.
- Onrust, S.A., Otten, R., Lammers, J., & Smit, F. (2016). School-based programs to reduce and prevent substance use in different age groups: What works for whom? Systematic review and meta-regression analysis. *Clinical Psychology Review*, 44, 45-59.
- 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.
- Resnik, D.B. (2000). Financial interests and research bias. *Perspectives on Science*, 8(3), 255-285.
- Robinson, L.A., Berman, J.S. & Neimeyer, R.A. (1990). Psychotherapy for the Treatment of Depression: A Comprehensive Review of Controlled Outcome Research. *Psychological Bulletin*, 108, 30-49.
- Smith, M.L., & Glass, G.V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.
- St. Pierre, T.L., Osgood, D.W., Mincemoyer, C.C., Kaltreider, D.L. & Kauh, T.J. (2005). Results of an Independent Evaluation of Project ALERT Delivered in Schools by Cooperative Extension. *Prevention Science*, 6, 305-317. Doi: 10.1007/s11121-005-0015-0.
- Tobin, M.J. (2003). Conflict of interest and AJRCCM: restating policy and a new form to upload. *American Journal of Respiratory and Critical Care Medicine*;167:1161-1164.
- Valentine, J.C., Biglan, A., Boruch, R.F., González Castro, F., Collins, L.M., Flay, B.R., & Schinke, S.P. (2011). Replication in Prevention Science. *Prevention Science*, 12, 103-117.
- Wilson, S.J., Lipsey, M.W. & Derzon, J.H. (2003). The Effects of School-Based Intervention Programs on Aggressive Behavior: A Meta-Analysis. *Journal of Consulting and Clinical Psychology*, 71, 136-149.

7

General discussion

General Discussion

Main findings and interpretation of the findings

Intervention studies

The first aim of this dissertation was to test the effectiveness of three preventive mental health interventions for children and adolescents (PATHS, Preventure, and POWER), both for the entire target group, as well as for subgroups in terms of gender, age, ethnicity, levels of problem behavior, and personality characteristics. In the PATHS and POWER study we also tested whether the intervention effects were moderated by the quantity and the quality of the delivery of the intervention. The objective of the trials was to provide information for others that could be helpful in deciding whether it is recommendable to implement these interventions on a broader scale in the Netherlands, as was done during the studies.

PATHS and the influence of the implementation strategy

The first trial we conducted (see chapter 2) revealed that the combination of a universal prevention program shown to be effective in several studies (PATHS), and using a health service system for implementing such an intervention on a broad scale, does not automatically result in positive effects for public health. This is a disappointing finding, as universal prevention programs only have a significant impact on the public health when they are effective *and* implemented on a large scale. It is therefore crucially important that we gain a better understanding on how to implement interventions on a large scale in a manner that maintains their effectiveness, or, how to develop interventions that are robust enough in their effectiveness that they survive variation in the delivery of the intervention.

The first objective can be achieved, for example, by a thorough support, training, and supervision system. It is known that such systems are very helpful in implementing interventions with, for example, more fidelity, and thereby result in better intervention outcomes (Durlak & DuPre, 2008). However, such systems can be either expensive or time consuming, and thus limit the dissemination of the intervention. Indeed, this was exactly

one of the reasons we tested a new implementation strategy for PATHS. The original, more thorough, implementation strategy provided no more than 5% of the elementary schools with PATHS in the Netherlands over the last twenty-five years. Its impact on society was, as a consequence, limited. Therefore, we tested a less intensive implementation strategy through health services, which could possibly then be used in the future to scale up the implementation of the intervention. However, the new strategy that was tested also did not seem to do the job successfully. This was, among other things, also reflected in the level of program completeness, which was 50% in the first year and 49% in the second year, and was probably also the reason no intervention effects were found. In sum, although the health service system used in this study could in theory be helpful in implementing the intervention on a broader scale, it was not able to do this without losing the interventions effectivity. Continuing implementation models with high levels of support, training, and supervision systems therefore remains of great importance, and the search for strategies to implement effective interventions on larger scales has to be continued.

Not only from a qualitative point of view, but also from a quantitative point of view, the complete implementation process have to be taken very serious from the very beginning of the development of an intervention. The impression I have, based om my personal experiences in the field of public health, is that in developing interventions the primary focus is too often centered on the theory of behavioral change within the target group, and that (the importance of) the implementation process is underestimated. This is a burden, as the implementation process can affect the effectiveness of the behavioral change model significantly and/or limit the dissemination of the intervention, as is demonstrated in this dissertation. The product of the theory of change behavior and the implementation process together determine the added value to the public health, both have to therefore be equally important when designing interventions. As stated in the introduction, an intervention can be seen as a theory about behavioral change, packed in an intervention that consists of manuals, workbooks and exercises. The same applies to the implementation process, it can be viewed as a theory concerning

the best way to implement an intervention. When an intervention is implemented for example from a licensee to a group of trainers, from this group of trainers to a group of teachers, and from these teachers to the actual target group of students, then there are three moments in which significant variance in the delivery of the intervention can occur and that can then influence the intervention outcomes. It seems important to me that intervention developers consider this process of implementation equally important as the theory of behavioral change. Fortunately there are methods available that can be very helpful herein, such as the method Intervention Mapping (Bartholomew et al., 2016). This method describes clearly how to incorporate both the model of change behavior and the implementation model in developing interventions.

Preventure and the importance of measuring the implementation

The second trial revealed that Preventure, an intervention that has shown to be effective in a Western country, is not per se also effective in another Western country (see chapter three). That is to say, not effective concerning the secondary outcomes, a range of mental health problems. Possible reasons for the differences in the outcomes between this study and other studies regarding mental health outcomes were considered in the discussion section of the chapter, such as sample differences, other ways of delivering the intervention, and cultural differences. However, we could not test these hypotheses. Considering the development of binge drinking and binge drinking frequency however, significant intervention effects were found (Lammers et al, 2015). An explanation for why intervention effects were found on these outcomes and not on other mental health outcomes, could be that the findings in the other Preventure trials were very modest (Castellanos & Conrod, 2006). Reproducing these secondary outcomes was probably more difficult than reproducing the outcomes regarding alcohol, which is after all the primary focus of the intervention.

In retrospect the use of implementation measures could and would probably have been helpful in understanding these different outcomes. In more general terms, this underlines the importance of combining outcomes measures with implementation measures. Without a good

understanding of the implementation process, it remains a guessing matter why an efficacious intervention is, or is not, effective when implemented in a specific manner. Or metaphorically speaking, we have to not only take a snapshot of the results, but also a 'making of' movie of the implementation process. Ultimately, it is not the theory of behavior change or the intervention itself that differ between an efficacy trial and an effectiveness trial. It is the same theory, the same anti-smoking classroom intervention, the same self-defense training, or the same anti-bullying project that is tested. It is therefore plausible that the differences have something to do with the implementation process of the intervention, at least, under the condition that the intervention is implemented in the same (cultural) context and within the same target group. Therefore it is not only important to look into the black box of the intervention, searching for the effective ingredients of it, but equally important to look into the black box of the implementation process. That we did not do so in the Preventure trial was a considerable limitation of our study.

POWER and the balance between fidelity and flexibility

The study of the POWER intervention (see chapter 4) underscores the idea that variation in the implementation process is related to variation in intervention outcomes. When implemented better, more effects can be achieved (Durlak & DuPre, 2008). This again pleads for greater investment into the implementation process of interventions. Without a proper implementation structure, and corresponding funding, it can be hard to achieve actual behavioral change within the target group. However, especially during this intervention, we also experienced how difficult it can be to take an accurate picture of the outcomes, to make a roentgen scan of the effective ingredients, and to make a movie of the implementation process. Describing the theory of behavioral change, describing the exact goals, and even describing the target group of the intervention was quite difficult in this study. This is, however, not unique for interventions within the social domain. A dynamic social context, multiple outcomes, and a heterogeneous target group all seem to make it a challenge to take accurate pictures. We, however, have to be aware that, although an intervention maybe hard to describe or to be proven effective, we do

have continue (the research into) this kind of interventions. Problems in the social domain will not be solved by waiting for the perfect picture, or by arguing that there are so few effective interventions. Ultimately, the fact that there are no studies demonstrating the effectiveness of an intervention, does not prove that nothing effective happens during the implementation of interventions. The encouraging thing is that the field of social interventions can easily learn from other fields, such as youth care or public health. It has only been during the last decade(s) that interventions in these fields are labeled as effective, and only a small number of interventions have done so till now. I am optimistic that we will find more interventions in the field of social interventions that are effective.

Conflict of interest

Our second aim was to describe to what extent being a developer-as-researcher, being his or her colleague, or being a licensee-as-researcher is common in intervention trials for children and adolescents. Additionally, we wanted to know to what extent this type of conflict of interest was associated with intervention trial outcomes, and whether this relation was especially be mediated by the cynical view or the high fidelity view.

Dutch database of effective youth interventions: catalyst or judge?

In the fourth study we found that a significant number of the studies that were used in a Dutch database of effective interventions to support the effectiveness of interventions, were conducted by researchers who have a conflict of interest. This is an important finding as it is known in general (Petrosino & Soydan, 2005), and shown in chapter six of this manuscript, that studies conducted by intervention developers and employees of the licensee of the intervention, show more positive intervention effects than studies performed by independent researchers. We will discuss the consequences of this finding in the next paragraph, as we would first like to highlight another important result of this study. Despite the efforts journals have invested into reporting this, a conflict of interest was rarely mentioned (correctly) in the articles we studied. In 39 of the 48 published studies (81%), incorrect information had been disclosed concerning a conflict of interest. A possible explanation is that,

in particular, intervention developers possibly do not see themselves as having a conflict of interest, considering that in general little money is earned in the field of preventive medicine. It also likely that probably a part of the articles was already published before the rules regarding reporting conflicts of interest were in effect. However, it still is a conflict of interest. Receiving personal credits and acquiring funding for the (governmental) organization one works for is, after all, also in benefit of the developer. Although the stakes are probably not as high as in the field of medicine, it still has to be clear that whether the 'pharmaceutical company' of the intervention studied the effects of the 'medicine', or if independent researchers studied the effects. More transparency about conflicts of interests is therefore needed.

Another point I would like to address is the lack of funding that is available in the Netherlands for the ongoing implementation (and modernization) of (public health) interventions and the extent to which funders explicitly require conflicts of interest to fund research. In reaction to the publication of our article, the argument was presented that it is desirable for intervention developers and licensees to be involved in effect studies (NJI, 2016). Moreover, inclusion of a developer or licensee of an intervention is an explicit criteria for funding of intervention research in the recent large funding schemes by ZonMw in the Netherlands (Zorg voor Jeugd 1 & 2). However, not only from a cynical point of view but also from a high fidelity point of view, this raises the question why this would be important in testing the effectiveness of interventions. Both explanations, the cynical view and high fidelity view, have been shown to have consequences for the predictive value of the observed effects. In the case of (un)conscious manipulation it is clear, but also in the case of the high fidelity view, it has to be taken into account that the results found in trials are not per definition also the effects that can be expected in daily practice.

We therefore have to realize that studies performed by authors with a conflict of interest probably overestimate the effects that are achieved in daily life, unless the interventions are implemented with the same high level of implementation quality during the day-to-day implementation

of the intervention. This, however, is exactly where the problem occurs. To the best of my knowledge, few organizations are able to implement (public health) interventions on a large scale while simultaneously also keeping the quality of the implementation on a high level. This has to do with, among other things, limitations in financial resources that are available. Therefore, as there is currently a central funding system for intervention (development and) research in the Netherlands (ZonMW), a central funding system is needed for the implementation (and modernization) of interventions. This is not only because funding for the (ongoing) implementation of (public health) interventions is generally very hard to acquire, but it will also bring more transparency into the funding process of the implementation of interventions that are already funded. Financing the implementation of interventions will then become less sensitive to the issues of the day and become more a matter of scientific knowledge about what works and what does not. More investments in the implementation (and modernization) are thus greatly needed and they will be a logical step after the investments that have been done in the recent decades to develop (and study the effectiveness of) interventions.

School-based substance use prevention programs: effective according to who?

The fifth study taught us that it is good to hand over the camera to independent researchers from time to time, regardless of whether developers/licensees-as-researchers can be accused of retouching the photos or not. This study revealed that in the field of school-based preventive interventions targeting alcohol and tobacco use, the effect size of the study is dependent on the researcher having a conflict of interest or not. This is an important finding, as the default assumption in research is that it does not matter who the researcher is for the study results. This, however, appears to be untrue. Unfortunately, we could not verify whether the results, in general, were better because of a more thorough implementation, manipulation of the research, or an another, unknown reason. More research on this topic is needed, as it can help in understanding the bias that arises when researchers with a conflict of interest study interventions. Nonetheless, one way or the other, the

influence of intervention developers and licensees in research trials blurs the sight on the 'real' intervention effects. More intervention studies by researchers without a conflict of interest are therefore also needed.

In addition, it is important that we keep building on transparency systems in which researchers are obliged to describe their initial objectives and (statistical) methods. Trial registers will be very helpful in this, particularly because not only intervention developers and licensees may have reasons to manipulate the study, as researchers that have no conflict of interest in relation to the intervention may also have reasons to search for positive intervention outcomes. As shown in this study, there seems to be a publication bias in particular in studies with small sample sizes. It is known that studies with positive outcomes are in general more easily accepted by journals than studies with null findings. Moreover, as is also known, having publications on the researcher curriculum vitae is beneficial in the academic field. More transparency in intervention research is therefore needed and initiatives like the Open Science Framework offers opportunities in doing so.

Methodological limitations

With regard to the effectiveness studies our pallet of implementation measures was limited. More insight in the implementation process would have been helpful in the interpretation of the study outcomes, and to compare them with the other pictures made of the intervention outcomes. Looking back I should have paid greater attention to the use of such implementation measures, which is an important lesson I've learned by conducting the studies described in this manuscript.

In the study regarding the Dutch database of interventions, it would have been extra informative if we had also collected data concerning the effect sizes found in the studies. Regardless of the limited number of studies, it would have been helpful to know whether the dual-roles indeed were related to better study outcomes. However, the small proportion of studies with strong experimental designs in the Dutch database would have limited the sample size too much to reliable test for moderation by researcher involvement. Considering the study focused on school-

based preventive interventions targeting alcohol, drugs, and tobacco use, it would probably have been better to use more indicators regarding both the high fidelity and the cynical view. This would have been helpful in answering the question whether the high fidelity view or the cynical view mediated the variance in intervention outcomes.

Generalizability of the results

With regard to the intervention studies, their outcomes are, of course, limited to these specific interventions. The lessons learned from these studies regarding the importance of replicating intervention studies, and the importance of using implementation measures are not limited simply to the interventions studied. This manuscript demonstrates that it cannot be assumed that what may work in one specific setting or under certain conditions, also will work in other settings and conditions (Green, 2001). To understand the reasons behind this, a better view on the implementation process is needed. Fortunately, a great amount of work is currently being conducted from this perspective and the field of implementation research is growing, as can be seen in the recent founding of the Global Implementation Initiative, the European Implementation Collaborative, and the Netherlands Implementation Collective, as well as the publication of a specific journal on this topic, *Implementation Science*.

With regards to the conflict of interest studies, the topics addressed in this manuscript are also of importance for other fields in which interpersonal contact is the effective ingredient (or nowadays also via e-health). We have already demonstrated this in this manuscript for a range of youth interventions targeting for example, problems with alcohol, drugs and tobacco, obesity, behavioral problems, sexuality, or attachment; and we have limited reasons to believe that this problem does not also occur in studies in other fields of psychological and social interventions such as criminal justice, social work, child welfare, or mental health care.

Implications for future research

The first implication is that it is worthwhile to keep investing in intervention research in the Netherlands, even when interventions have already been proven to be effective elsewhere. Both PATHS and Preventure had shown effectiveness in, for example, respectively the US and Canada, however these intervention effects could not be replicated during our studies. (That is to say, the intervention effects of Preventure on the secondary outcomes could not be replicated.) This is an important reason to keep testing the effects of interventions that are adopted from abroad in the cultural context where they are aimed to be implemented. Indeed, the number of evidence based interventions available in The Netherlands is still limited. The database 'Effectieve Jeugdinterventies' for example only includes seven "evidence based" interventions as of mid-2016. Therefore more intervention research is needed in the upcoming decades to raise this number of effective interventions. I would also advocate that there is a need to change the focus from efficacy trials 'Is this intervention effective?' to effectiveness trials 'Is this intervention effective in daily practice?', and dissemination trials 'Is this intervention effective when implemented on a large scale?'. There is after all a growing number of interventions that have been shown to be efficacious, yet, what is currently needed are interventions that also work in daily life, when implemented on a broad scale. Only then significant impact on the public health can be achieved with preventive interventions. In favor of this, it is probably also helpful to take pictures of, or monitor, interventions that are currently implemented on a broad scale, yet are not per se evidence based. Regardless of whether the intervention is effective or not, it is helpful to understand how some licensees have managed to implement their intervention on a large scale. What are the conditions for implementing an intervention on a broad scale, and what can we learn from this for interventions that are effective but have not yet been implemented on a broad scale?

With that in mind, it is necessary that describing the implementation process in articles and using implementation measures in trials, becomes equally important as describing the intervention and outcomes measures. In our conflict of interest studies we have experienced how

hard it is to find information regarding the implementation process in the research articles. This is a pity, as it is essential information to understand the trial outcomes. Observations, questionnaires, and logs can all be used in measuring the implementation. Berkel et al. (2011), for example, wrote a paper describing eight dimensions of implementation that can be measured; differentiation (distinctiveness of a program's theory and practices from other available programs), dosage (number of program sessions delivered), reach (extent to which participants of the program are representative for the target population), monitoring (documenting what services members of the intervention and control groups receive beyond program evaluation), fidelity (whether prescribed program components were delivered as instructed in program protocol), quality (teaching and clinical skills with which the program is implemented), adaptation (additions made to the program during implementation), and responsiveness (involvement and interest in the program). It is important to understand which of these implementation variables does indeed influence the effectiveness, and what needs to be changed in the implementation process to reproduce the effects found in efficacy trials. These questions have to be answered in order to create interventions and delivery systems that do contribute to the public health.

Implications for public health

The studies in this manuscript were helpful in making the decision whether the interventions under study should be implemented on a broaderscale. It was decided to not continue the strategy of implementing PATHS by the Municipal Health Services, and remain with the common implementation strategy including dedicated PATHS trainers. Ultimately this seems to be a better strategy to achieve the intervention effects. It was further decided to not implement Preventure as an intervention for mental health outcomes. The results of the intervention effects of Preventure on alcohol outcomes were, however, more encouraging (Lammers et al, 2015), and it is currently under consideration whether Preventure should be implemented as a strategy in reducing binge drinking. The third study in this manuscript was helpful in deciding how to continue the implementation of POWER. Based on the study it was decided to first improve the theoretical base of the intervention,

improve the handbooks, and to tighten up the admission requirements for professionals to become a POWER trainer. Additionally, the license to implement the intervention nationwide for the upcoming four years was given to an organization that is probably able to do this with greater fidelity than the Trimbos Institute.

The implications for public health from the conflict of interest studies are more difficult to quantify, and are a long shot. The aim of these studies was to address a problem that is common in intervention research, stimulate more independent research, get a better idea of which interventions do really work, subsequently enhancing the chance that these interventions in particular will be implemented and thereby improve mental health. However, describing the improvements that can be made to intervention research, can also be seen as a disqualification of all the integer work that is done by an intervention researcher. Local governments, the main source of funding for these type of interventions in the Netherlands, could distrust intervention researchers as a result of the conflict of interest studies done in this manuscript. Consequently they may choose to finance interventions they think will work, instead of relying on research conducted. This would, of course, never be a satisfying outcome as science is still the most precise way to determine what works and what doesn't.

Conclusions

In this manuscript we have shown that it is fruitful to take pictures of an intervention during several stages of its development phase. For example, when it is well nurtured at home, when it is on vacation in another country, and when someone else holds the camera. All these pictures together add to the general picture of the intervention. The first objective of this manuscript was to perform studies that could help decide whether it is recommendable to implement PATHS, Preventure, and POWER on a broad scale in the Netherlands using the implementation strategy specified for each. It can be concluded that this is not a good idea. That is, it is not a good idea for PATHS to be implemented using the less intensive implementation strategy tested, for Preventure to be implemented when aiming to prevent mental health problems, and

for POWER to be implemented when not done with enough fidelity. The other conclusion we can draw is that it is important to study the efficacy and effectiveness of interventions multiple times and in multiple ways. The effects of an intervention can vary depending on several influencing factors, and as shown specifically in this manuscript, depending on the delivery of the intervention and a conflict of interest of the researcher. Therefore, keep on taking pictures, do not forget to make the 'making of' movies, and occasionally hand over the camera to someone else!

References

- Bartholomew Eldridge, L.R., Markham, C.M., Ruiter, R.A.C., Fernández, M.E., Kok, G., & Parcel, G.S. (2016). *Planning health promotion programs; an Intervention Mapping approach*, 4th Ed. San Francisco, CA: Jossey-Bass.
- Berkel, C., Mauricio, A.M., Schoenfelder, E., & Sandler, I.N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science*, 12, 23-33.
- Castellanos, N., & Conrod, P.J. (2006). Brief interventions targeting personality risk factors for adolescent substance misuse reduce depression, panic and risk-taking behaviours. *Journal of Mental Health*, 15(6), 645-658.
- Durlak, J.A. & DuPre, E.P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am. J. Community Psychol.* 41, 327-350.
- Green, L.W. (2001). From Research to "Best Practices" in Other Settings and Populations. *Am J Health Behav* 2001;25(3):165-178.
- Lammers, J., Goossens, F., Conrod, P., Engels, R., Wiers, R.W., Kleinjan, M. (2015). Effectiveness of a selective intervention program targeting personality risk factors for alcohol misuse among young adolescents: Results of a cluster randomized controlled trial. *Addiction*, 110, 1101-1109.
- NJI (2016). Reaction of the Netherlands Youth Institute in: De Winter, L., Verhagen, A.M., Goossens, F.X. (2016). Dubbelrollen in effectonderzoek. Over de onafhankelijkheid van onderzoekers in effectonderzoeken naar interventies uit de databank Effectieve Jeugdinterventies. *De Psycholoog*, 6, 40-53.
- 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.



Summary
Samenvatting
Dankwoord
Over de auteur

Summary

During the past decades in the Netherlands, as in other Western countries, there has been a lot of investment into research concerning the effectiveness of preventive interventions in public mental health. At the core of many of these research studies is the question whether the intervention is effective in attaining the goals of the intervention. This dissertation seeks to answer this question regarding three preventive interventions that focus on improving the social and emotional development of children in primary school (PATHS), on decreasing the alcohol use and psychological problems of adolescents in secondary school (Prevention), and on empowering second-generation immigrant youth (POWER), respectively. Not only were the outcomes of the effectiveness studies examined, but it was also examined for PATHS and POWER to what extent their outcomes are related to variation in the implementation of these interventions. The underlying goal of the present studies is to collect information that can be of aid in determining if, and how, these interventions can be widely implemented in the Netherlands.

Additionally, this dissertation will examine the double roles that researchers may play. It is known that studies conducted by researchers, who are simultaneously also the developers, their co-workers, or license-holders of the intervention they are studying, show greater effect sizes than independent studies. This dissertation will examine to what extent double roles occur in the studies that were the basis in judging the effectiveness of interventions included in the database Effectieve Jeugdinterventies, as well as in an international dataset of studies concerning school-based interventions for the prevention of substance use. Furthermore, I examine whether the independence of researchers does indeed have an influence on the effect sizes, and if so, if this effect can be explained by the cynical view or the high fidelity view. The high fidelity view entails that developers and license holders are very aware of how the intervention should be conducted and seek to ensure that it is indeed conducted properly during implementation. This increases the chance that the intervention does achieve the goals set. The first explanation, the cynical view, assumes that the researchers (un)

consciously influence the results, thereby increasing the possibility of positive results. A more detailed introduction, as well as the design of this dissertation are described in *chapter one*.

In the *second chapter*, the quasi-experiment regarding PATHS is discussed. This intervention provides lessons concerning social and emotional development for children in all eight years of primary school in the Dutch education system. The intervention was originally developed in the United States, where extensive research has been conducted regarding its effectiveness and it was evaluated as effective by several institutions. The intervention has already been offered in the Netherlands for several decades, however up until today its effectiveness for children in regular primary schools in the Netherlands has not been studied. To add to this, the Dutch license holder was only able to reach 5% of the targeted population during the past decades and thereby the impact on the psychosocial health of the Dutch youth has been limited. The quasi-experiment described in chapter 2 examines the effectiveness of this intervention over a period of two years among 1,294 students in kindergarten, grades 1, 3, and 5 of regular primary education. The training and support provided to the teachers was not conducted, as is customary, by trainers associated with the license holder, but instead by youth health care nurses and prevention workers in municipal health organizations. This was done to examine whether the intervention would also be effective when implemented in this manner, keeping in mind that this could lead to greater dissemination (and thereby also greater impact) of the intervention on the long term. The results, however, revealed that the intervention was not effective under these conditions. Considering that the extent of the implementation of various program elements was also measured, we could determine that on average only half of the program elements were actually executed in both the first and the second year of implementation. Though we could not determine to what extent the intervention would have been executed if it had been implemented by trainers associated to the license holder, we could determine that the chosen method of implementation had led to a relatively low degree of completeness of implementation. It is plausible that this, consequently, led to the lack of intervention effects.

This is not only a disappointing result, but also an important argument to be cautious with broad implementation of preventive interventions, that have only been able to demonstrate their value when implemented under favorable conditions, but not yet under the conditions that make broad implementation possible.

The *third chapter* describes the study regarding Preventure, originally a Canadian intervention, that focusses primarily on reducing binge-drinking, and subsequently on reducing psychosocial mental health problems. In this cluster randomized control trial we examine the effect of this intervention on secondary outcome measures, among 699 students in the first years of secondary school. More specifically, the effects on depressive symptoms, anxiety problems, hyperactive behavior and delinquent behavior are examined. In contrast to a previous international study regarding the effectiveness of Preventure, we did not find any evidence of this intervention's effectiveness on the outcomes measures in the present study. It is still unclear what could explain these differences. The intervention developer was involved within not only the translation of the intervention material, but also the implementation and the effectiveness study. Additionally, the long-term (12 months), middle-term (6 months), and short-term (2 months after intervention was completed) effects were all measured and sufficient resources were available to implement the intervention well. However, the actual extent of implementation was not measured and several outcome measures, such as delinquent behavior, were already very limited in prevalence in the sample population at baseline. Nevertheless, this study once again revealed that effects found in one implementation context do not necessarily also have to be found in another context, even when the conditions of the implementation of the intervention do not in any exceptional way differ between the two contexts.

The *fourth chapter* discusses a Dutch intervention that seeks to prevent marginalization of second generation adolescents of Moroccan, Turkish, Surinamese, or Antillean descent. In total 132 adolescents participated in a POWER course in several groups, each containing 13 sessions of 2.5 hours, while 116 adolescents participated in the control group during the

three months of implementation, after which they were also given the opportunity to participate in a POWER course. The results of this cluster randomized control trial displayed several positive intervention effects, though this was found especially for those groups in which the extent of implementation was relatively high. An important question that this study raises is: which conditions actually lead to higher effectiveness? It thereby cannot be said that implementing the intervention exactly according to the manual will per definition lead to better results, especially when it concerns a sample population that is difficult to reach and retain during preventive interventions. This study underscores the importance not only of measuring outcome variables, but also of keeping an accurate record of implementation variables. Without this information it remains difficult to determine which elements of the intervention are crucial in achieving the intervention effects in daily practice.

Chapter five explores the concept of double roles in effectiveness studies. As described in the introduction, the independence of researchers is one of the elements that play a role in what results are found in effectiveness studies of preventive interventions. More precisely, it is known that research conducted by researchers holding a double role leads to more positive results than research that is conducted by independent researchers. Independence hereby implying that the researcher was not the developer of the intervention (or their colleague), and also not the owner or license holder of the intervention in a specific country. In the fifth chapter, we examine to what extent the evaluations of the database Effectieve Jeugdinterventies regarding the effectiveness of preventive interventions are based on studies that were conducted by independent researchers. For that purpose, 86 research articles were investigated that were used by the evaluating committee of the database to grant 26 interventions the designation of 'effective'. These articles only pertain to research conducted within the Netherlands, presuming that those studies are most predictive of the effects that could be obtained in the Dutch society. The results show that in 84% of the research articles, at least one (co-)author was not independent. Viewed from the perspective of the intervention, this means that in total 19 of the 26 evaluations of the interventions were based solely on articles conducted by dependent

researchers. Regardless if the more positive outcomes that are often found in dependent effectivity studies can be explained by the high degree of involvement (high fidelity view) or (un)conscious manipulation (cynical view) of the researchers, there is a realistic possibility that a large percentage of the interventions in the database are not as effective in daily practice as could be assumed based on the effectivity studies.

Chapter six seeks to provide an answer to the question to what degree double roles occur in an international dataset of 288 school-based interventions for the prevention of stimulant use. It appeared that for 75% of the examined interventions, double roles occurred. Additionally, this chapter examined whether (in)dependence is related to more positive outcome measures using meta-regression analyses. For the outcome measure of smoking, this was found to be the case if one of the researchers was also the intervention developer. For alcohol use, it was found that this was the case when one of the researchers was employed by the license holder. Lastly, it was examined if this association could be explained because the researchers, who were also the developers or owners of the intervention, knew the intervention very well and were therefore able to implement it well (high fidelity view) or because the researchers (un)consciously manipulated the study to achieve more positive results (cynical view). Yet, the analyses did not provide evidence that one of these mechanism mediated the relation between having a conflict of interest and finding larger effect sizes.

The *final chapter* summarizes and discusses the main findings of the first chapters. In addition, I formulate several recommendations that strive to extract more value from studies regarding preventive interventions. It is argued that more independent effectivity studies for preventive interventions are needed, as well as greater use of measuring instruments to monitor the implementation of the intervention during effectivity studies. Finally it is of great importance to be aware the role implementation plays in the search for interventions that are effective not only under optimal conditions, but also when implemented on a larger scale.

Samenvatting

De afgelopen decennia is er in Nederland, evenals in andere Westerse landen, veel geïnvesteerd in het onderzoeken van de effectiviteit van preventieve interventies voor de publieke geestelijke gezondheid. In deze onderzoeken staat veelal de vraag centraal of de interventie effectief is in het behalen van de interventiedoelen. In dit proefschrift stel ik die vraag voor drie preventieve interventies die respectievelijk beogen de sociale en emotionele ontwikkeling van kinderen in het basisonderwijs te bevorderen (PAD-leerplan), alcoholgebruik en psychische problemen terug te dringen bij jongeren in het voortgezet onderwijs (Prevention), en tweede generatie allochtone jongeren te empoweren (POWER). Daarbij is voor PAD en POWER niet alleen gekeken naar de uitkomsten van het effectonderzoek, maar ook in hoeverre deze uitkomsten samenhangen met variatie in de implementatie van de interventie. Het onderliggende doel van de onderzoeken is om informatie te verzamelen die behulpzaam is bij het bepalen of, en hoe, deze interventies op grotere schaal geïmplementeerd zouden kunnen worden in Nederland.

Daarnaast ga ik in dit proefschrift in op dubbelrollen van onderzoekers. Het is in zijn algemeenheid bekend dat onderzoeken die zijn gedaan door onderzoekers die tevens de interventieontwikkelaar zijn, een collega van hem of haar, of werken voor de licentiehouder van de interventie, grotere effectgroottes laten zien dan onafhankelijke onderzoeken. In dit proefschrift bekijken we in welke mate dubbelrollen voorkomen in onderzoeken die ten grondslag liggen aan de effectiviteitsoordelen van interventies in de databank Effectieve Jeugdinterventies, en in een internationale dataset van onderzoeken naar schoolgerichte interventies voor genotmiddelenpreventie. Daarnaast bekijken we of die afhankelijkheid inderdaad invloed heeft op de effectgroottes en of dit effect verklaard wordt door de cynical view of de high fidelity view. Dit laatste gezichtspunt houdt in dat de ontwikkelaars en licentiehouders bijzonder goed weten hoe de interventie uitgevoerd moet worden en hier tijdens het onderzoek ook zorg voor dragen. De kans dat de interventie haar doelen bereikt zou daardoor vergoot worden. De eerste verklaring gaat er vanuit dat de onderzoekers de resultaten (on)bewust positief

beïnvloeden en daarmee de kans op positieve uitkomsten vergroten. Een uitgebreidere inleiding op dit onderwerp en de opzet van dit proefschrift staan beschreven in het *eerste hoofdstuk*.

In het *tweede hoofdstuk* bestuderen we door middel van een quasi-experiment het PAD-leerplan. Dit lesprogramma voor sociaal-emotionele ontwikkeling is bedoeld voor kinderen in alle acht de leerjaren van het Nederlandse basisonderwijs. De interventie is ontwikkeld in de Verenigde Staten, daar uitgebreid onderzocht op effectiviteit en door verschillende instanties als effectief beoordeeld. De interventie wordt daarnaast ook al enkele decennia in Nederland aangeboden, maar was hier tot op heden nog niet op effectiviteit onderzocht onder leerlingen in het reguliere basisonderwijs. Daar kwam bij dat de Nederlandse licentie-eigenaar slechts vijf procent van de beoogde doelgroep wist te bereiken in de afgelopen decennia en de impact op de psychosociale gezondheid van de Nederlandse jeugd derhalve beperkt was. In het quasi-experiment dat beschreven staat in dit hoofdstuk wordt de effectiviteit van de interventie gedurende twee jaar bestudeerd onder 1.294 leerlingen uit de groepen 2, 3, 5 en 7 van het reguliere basisonderwijs. De training en ondersteuning aan leerkrachten werd daarbij niet zoals gebruikelijk uitgevoerd door trainers die verbonden zijn aan de licentiehouder, maar door JGZ-verpleegkundigen en preventiewerkers van GGD-en. Het doel hiervan was om te toetsen of de interventie ook effectief was als hij langs deze lijn werd geïmplementeerd, met daarbij in het achterhoofd de gedachte dat dit optermijnto een hogere disseminatiegraad (en daarmee meer impact) van de interventie kon leiden. De resultaten lieten echter zien dat de interventie onder deze condities niet effectief was. Omdat de mate van implementatie van de verschillende programmaonderdelen in dit onderzoek ook gemeten was, kon vastgesteld worden dat gemiddeld genomen slechts de helft van de programmaonderdelen was uitgevoerd in zowel het eerste als het tweede jaar van de uitvoering. Hoewel niet vastgesteld kon worden in welke mate de interventie compleet uitgevoerd zou zijn als deze geïmplementeerd was met behulp van de trainers van de licentiehouder, kon wel vastgesteld worden dat deze manier van implementeren leidde tot een betrekkelijk lage mate van compleetheid van implementatie. Het is aannemelijk dat dit vervolgens

heeft geleid tot het ontbreken van interventie-effecten. Dit is niet alleen een teleurstellende uitkomst, maar ook een belangrijk argument om behoedzaam te zijn met het breed implementeren van preventieve interventies die hun waarde wel onder gunstige omstandigheden hebben bewezen, maar nog niet onder omstandigheden die een brede implementatie mogelijk maken.

In *hoofdstuk drie* beschrijven we een onderzoek naar Preventure, een van oorsprong Canadese interventie die zich primair richt op het terugdringen van binge-drinken en secundair op het terugdringen van psychosociale ongezondheid. In deze cluster randomized controlled trial onderzoeken we het effect van deze interventie onder 699 leerlingen in de eerste jaren van het voortgezet onderwijs op de secundaire uitkomstmaten, meer precies op het effect op depressieve symptomen, angstproblemen, hyperactief gedrag en delinquent gedrag. In tegenstelling tot een eerder buitenlands effectonderzoek naar deze interventie, vonden we in dit effectonderzoek geen bewijzen voor de effectiviteit van het programma op deze uitkomstmaten. Waar dit aan ligt is onduidelijk gebleven. De interventie-ontwikkelaar was betrokken bij zowel de vertaling van de interventiematerialen als de implementatie als het onderzoek. De effecten zijn zowel op de lange (12 maanden), middellange (6 maanden) als de korte termijn gemeten (2 maanden na afronding van de interventie) en er waren voldoende middelen beschikbaar om de interventie goed te implementeren. Aan de andere kant, de daadwerkelijke mate van implementatie was niet gemeten en sommige uitkomstvariabelen, zoals delinquent gedrag, waren in de steekproef bij aanvang al zeer beperkt aanwezig. Desalniettemin heeft het onderzoek wederom laten zien dat een effect dat is gevonden in de ene uitvoeringscontext niet vanzelfsprekend ook gevonden wordt in een andere context. Ook niet als de omstandigheden waarin de implementaties plaatsvinden niet op uitzonderlijke wijze lijken te verschillen van elkaar.

Het *vierde hoofdstuk* gaat in op een Nederlandse interventie die marginalisatie beoogt te voorkomen onder tweede generatie jongeren van Marokkaanse, Turkse, Surinaamse of Antilliaanse afkomst. In totaal 132 jongeren volgden in verschillende groepen de POWER cursus bestaande

uit 13 sessies van 2,5 uur en 116 jongeren zaten gedurende de drie maanden van uitvoering in de controlegroep, waarna ook zij de POWER cursus konden volgen. De uitkomsten van deze cluster randomized controlled trial lieten enkele positieve interventie-effecten zien, maar vooral in de groepen waarin de mate van uitvoering meer interventiegetrouw was. Een belangrijke vraag die dit onderzoek oproept is welke omstandigheden exact zorgden voor die hogere effectiviteit. Het is daarbij namelijk niet gezegd dat het exact volgen van de handleiding per definitie leidt tot betere resultaten, in het bijzonder bij een doelgroep waarvan bekend is dat ze lastig te bereiken en vast te houden is voor preventieve interventies. Het onderzoek onderstreept daarmee het belang van niet alleen het meten van de uitkomstvariabelen, maar ook van een accurate registratie van de implementatievariabelen. Zonder deze informatie blijft het lastig om aan te geven welke elementen van een interventie cruciaal zijn om ook in de dagelijkse praktijk interventie-effecten te realiseren.

In hoofdstuk vijf is nader onderzoek gedaan naar dubbelrollen in effectonderzoeken. Zoals beschreven in de introductie, is de onafhankelijkheid van onderzoekers één van de factoren die een rol speelt bij de resultaten die gevonden worden in effectonderzoeken naar preventieve interventies. Onafhankelijk wil daarbij zeggen dat de onderzoeker niet degene is die de interventie ontwikkeld heeft (of een collega daarvan), noch een licentie heeft om deze te implementeren in een bepaald land. In het vijfde hoofdstuk bekijken we in hoeverre de oordelen van de databank Effectieve Jeugdinterventies over de effectiviteit van preventieve interventies gebaseerd zijn op onderzoeken die uitgevoerd werden door onafhankelijke onderzoekers. Daartoe zijn 86 onderzoeksartikelen bestudeerd die ten grondslag lagen aan het oordeel van de beoordelingscommissie van de databank om 26 interventies het predicaat effectief toe te kennen. Het gaat daarbij enkel om onderzoek dat is uitgevoerd in de Nederlandse context, ervan uitgaande dat die onderzoeken het meeste voorspellend zijn voor de effecten die behaald kunnen worden in de Nederlandse samenleving. De resultaten lieten zien dat in 84 procent van de onderzoeksartikelen minimaal één (co-) auteur niet onafhankelijk was. Vanuit het perspectief van de interventie: in totaal 19 van de 26 kwaliteitsoordelen van de interventies in de

databank zijn gebaseerd op artikelen die enkel onderzoeken beschrijven die zijn uitgevoerd door afhankelijke onderzoekers. Ongeacht of de gunstigere uitkomsten die doorgaans gevonden worden in afhankelijk effectonderzoek verklaart wordend ooreenhogemate van betrokkenheid door de onderzoekers of manipulatie, er is een reële kans dat een groot deel van de interventies uit de databank in de dagelijkse praktijk niet zo effectief is als op basis van de effectonderzoeken verondersteld zou kunnen worden.

In *hoofdstuk zes* is een antwoord gezocht op de vraag in welke mate dubbelrollen voorkomen in een internationale dataset van 288 schoolgerichte interventies voor genotmiddelenpreventie. Dit bleek het geval bij 75 procent van de onderzochte programma's. Tevens is in dit hoofdstuk door middel van meta-regressie analyses bestudeerd of (on) afhankelijkheid samenhangt met positievere onderzoeksuitkomsten. Dat was in het bijzonder het geval voor de uitkomstmaat roken als één van de onderzoekers tevens de interventieontwikkelaar was en voor de uitkomstmaat alcohol als één van de onderzoekers tevens voor de licentiehouder werkte. Ten slotte is gekeken of deze samenhang verklaard kon worden doordat de onderzoekers, die tevens interventie-ontwikkelaar of interventie-eigenaar zijn, de interventie als geen ander kennen en hem daardoor zeer goed konden uitvoeren (*high fidelity view*) of omdat ze (on) bewust het onderzoek beïnvloedden om positieve uitkomsten te behalen (*cynical view*). De mediatie-analyses hebben echter geen evidentie geleverd dat één van deze twee mechanismes de relatie medieert.

In het *laatste hoofdstuk* zet ik de bevindingen uit de eerste hoofdstukken nog eens op een rij. Daarnaast formuleer ik een aantal aanbevelingen die beogen om uiteindelijk meer rendement te halen uit de onderzoeken naar preventieve interventies. Zo pleit ik voor meer onafhankelijk effectonderzoek naar preventieve interventies, meer ook voor meer gebruikmaking van meetinstrumenten om de implementatie te monitoren tijdens effectonderzoeken. Ten slotte beschrijf ik het belang om meer oog te hebben voor de rol van implementatie in de zoektocht naar interventies die niet alleen in een optimale setting effectief zijn, maar ook wanneer ze op grote schaal geïmplementeerd worden.

Dankwoord

Voor wie de rest van dit proefschrift niet heeft gelezen, het gaat over preventieprogramma's. In de onderzoeken die ik er naar heb gedaan werken ze eigenlijk niet echt (hoofdstuk 2 t/m 4). In de onderzoeken die anderen er naar hebben gedaan werken ze soms wel, maar betwijfel ik of dat wel echt zo is (hoofdstuk 1 en 7). Ik heb nog geprobeerd om te kijken of ik er iets meer over kon zeggen, maar daar kwam ik niet echt uit (hoofdstuk 5 en 6). Zo, dat scheelt je weer een boel leeswerk.

Lieve familie, vrienden en collega's, dit is in een notendop waar ik me de afgelopen 12,5 jaar mee bezig heb gehouden en ik vond het fantastisch. Nou ja, soort van, dat wil zeggen, ik doe het nooit meer. Ik herinner me eigenlijk vooral heel veel vloeken tegen mijn kamergenoot. "Waarom zijn we hieraan begonnen?", "Wie beweerde ook alweer dat promoveren leuk is?" en "Geloof jij er nog in dat je proefschrift ooit afkomt?" waren regelmatig te horen uitspraken op onze kamer. Maar, zelfs tegen mijn eigen verwachting in is het er toch gekomen en mag ik jullie nu het meest gelezen deel van proefschriften presenteren: het dankwoord. Voor wie zichzelf snel wil opzoeken, ik heb een handige indeling voor je gemaakt: promotor en copromotoren, collega's, vrienden, Tjerk en familie. Op het einde nog wat emotionele woorden.

Promotor en copromotoren

Bram Orobio de Castro, wat mij betreft de hoogleraar met de mooiste achternaam. Iedere keer dat wij afspraken kwam ik vol zelfvertrouwen en goede moed thuis, het was volkomen duidelijk welke kant het op moest. Dat ik vervolgens bij thuiskomst echt geen flauw benul had waar ik moest beginnen, schijnt er dan bij te horen. Je hebt me met jouw onuitputtelijke enthousiasme en optimisme door mijn promotie heen geloodst. Bram, veel dank voor je hulp en de waardevolle discussies.

Karin - *Het staat in dit artikel, ik stuur het je op.* - Monshouwer, jij was de precieze van mijn twee copromotoren. Op de momenten dat ik de analyses of de discussie wel af vond, zorgde jij ervoor dat ik toch nog een slag maakte. Schaven, verder uitzoeken, meer artikelen lezen, niet zeuren



en nauwkeuriger opschrijven. Karin, dankjewel voor het aanleren van het handwerk van wetenschappelijk onderzoek doen.

Simone - Bakkie doen? - Onrust, jij was degene bij wie ik kon uitblazen. Samen moppen op de wereld en verzochten dat anderen moeilijk doen. Maar jij was ook degene die me hielp bij de veelal ingewikkelde analyses en hielp oplossingen zoeken voor methodologische vraagstukken. Simone, dank voor al je hulp en je onvoorwaardelijke vertrouwen in mij als promovendus. Hoe druk je het zelf ook had, je was er voor me als ik je nodig had.

Collega's

Elke promovendus heeft een kamergenoot en meestal is dat een bijzonder iemand voor hem of haar. In mijn geval is dat niet anders. Jeroen Lammers is niet alleen mijn collega, kamergenoot, mede-promovendus op het Trimbos-instituut en mijn voormalig afstudeerbegeleider bij de GGD Midden-Nederland, maar bovenal mijn maatje. Dank voor je luisterend oor, de dagelijkse gezelligheid en de enorme lol die we samen hebben. Met jou als paranimf, ben ik verzekerd van een enorme steun tijdens mijn verdediging. Dank ook aan mijn andere collega's, in het bijzonder die van mijn programma Jongeren & Riskant Gedrag en het secretariaat. Brigitte, Carola, Jeroen, Joost, Lex, Lonneke, Lotte, Marloes, Martha, Myranda, Neeltje en Pien, jullie maken het een plezier om naar het werk te komen. Ninette, dank voor de ruimte die je voor me hebt gemaakt om mijn droom te verwesenlijken. Aukje, dank voor alle leuke en lieve kaartjes en berichtjes die ik, ondanks al mijn liefdevolle gepest, de afgelopen jaren van je heb mogen ontvangen.

Vrienden

Micha, Jos, Erik, Robbert, Michiel, Evert, Thijs, Hester en Ruben, jullie hebben volgens mij nog steeds geen flauw idee wat mijn werk inhoud, of 'vrijwilligerswerk', zoals Raoul het eens noemde. Jullie vinden me vooral zeuren over de gevaren van alcohol en drugs. Maar vanaf nu weten jullie het wel, bij deze het bewijs dat ik de afgelopen jaren echt aan het promoveren was. Dank voor jullie vriendschap, jullie zorgen voor de balans in mijn leven. Bijzondere dank voor Robbert die de prachtige voorkant heeft ontworpen.

Twee vrienden wisten wel waar ik mee bezig was. Allereerst Wijnand die in hetzelfde werkveld actief is en met wie het heerlijk is om eindeloos te bomen over het werk. Over wat er niet aan deugt en over hoe het beter kan. Dankjewel voor je vriendschap en het doet me veel plezier dat je mijn paranimf wilt zijn. Dan Célestine, met wie ik samen gestudeerd heb en die ontelbare keren haar vertrouwen in mij heeft uitgesproken. Hoewel je je er misschien niet bewust van bent, het heeft me enorm gesteund. Thanks pop!

Tjerk

Tjerk, de man met wie ik in de toekomst gedoe wil maken. In jou heb ik mijn grote liefde gevonden. Het was even zoeken, maar dan heb je ook wat! De vrolijkheid die je me elke dag brengt en de plannen die we samen maken, mijn leven is absoluut leuker met jou aan mijn zijde. Ik kijk uit naar een mooie tijd met jou in onze prachtige werfkelderloft.

Familie

Mam, dankzij jou mag ik mijzelf nu doctor noemen. Jij hebt er voor gezorgd dat ik heb kunnen studeren en door jouw liefdevolle opvoeding het vol kon houden op de moeilijke momenten in ons leven. Je hebt er echt alles aan gedaan om ervoor te zorgen dat ik me volledig heb kunnen ontwikkelen tot de gelukkige volwassen man die ik nu ben. Pap, door jou heb ik mijn karakter kunnen ontwikkelen. Dankzij jou weet ik wat het is om hard te werken, door te zetten en dat ik zelf iets van mijn leven moet maken. Ik hoop dat je daarboven een beetje trots op me bent. Maikel, mijn grote broer, we kletsen niet zo vaak, maar bij deze wil ik je mijn dank uitspreken voor alles wat je voor me gedaan hebt. Dat is meer dan je soms denkt. Channel, mijn knappe kleine zusje, zonder jou was het leven simpelweg saai geweest. Niet veranderen hoor, gewoon af en toe te laat blijven komen, dat is goed voor je grote broer om los te leren laten. Daarnaast een hele dikke kus voor Cor, Kim, Yermo en natuurlijk mijn grote kleine helden Stijn en Kick.

Ten slotte, of het gepast is weet ik niet, maar ik ben ook best een beetje trots op mezelf. Het schrijven van een proefschrift heeft me bewust gemaakt van zowel mijn sterke kanten als mijn beperkingen. Het was

een proces van doorzetten en teleurstellingen overwinnen, maar ook van academische ontwikkeling en persoonlijke groei. Het was de taaiste uitdaging in mijn werkzame leven, maar daarmee ook die waar ik het meeste trots op ben!

Ferry

Over de auteur

Curriculum Vitae

Ferry Goossens (1978) ontving in 2000 zijn Bachelor diploma Sociaal Pedagogische Hulpverlening aan de Hogeschool van Amsterdam. Het mag daarbij niet onvermeld blijven dat hij in de eerste jaren van die opleiding, zo rond zijn twintigste levensjaar, tevens twee jaar onderdeel was van de succesvolle Nederlandse boyband Velvet. Na afronding van zijn opleiding werkte hij een half jaar als administratief medewerker om na te denken over zijn toekomst. In het voorjaar van 2001 vervulde hij deze functie voor die van steward bij Martinair waarin hij een half jaar in zonnigere oorden kon nadenken over zijn ambities. In het najaar van 2001 startte hij met de (verkorte) opleiding Algemene Sociale Wetenschappen aan de Universiteit Utrecht. Zijn afstudeeronderzoek vervulde hij bij de GGD Midden-Nederland en betrof een effectonderzoek naar *De Gezonde School en Genotmiddelen* in het basisonderwijs. Daarnaast werkte hij als preventiewerker op dit project en gaf daartoe enige tijd voorlichting over alcohol, tabak en drugs aan leerlingen uit groep 8. Met het behalen van zijn Master diploma in 2004 en het volgen van de opleiding voor trainer mentale en fysieke weerbaarheid op het Marietje Kessels project, was de basis gelegd voor een (academische) carrière in het werkveld van de preventieve gezondheidszorg.

In de zomer van 2004 begon hij als wetenschappelijk medewerker bij het Trimbos-instituut. De landelijke implementatie van *De Gezonde School en Genotmiddelen* en de begeleiding van de implementatie van het PAD-leerplan gedurende een effectonderzoek, waren daarbij zijn eerste taken. Daarmee werden ook de eerste stappen gezet in een promotietraject dat met de voltooiing van dit proefschrift na 12,5 jaar tot een einde is gekomen. In deze periode deed Goossens niet alleen effectonderzoek naar de preventieve interventies PAD-leerplan, Preventure en POWER, maar stond hij tevens aan de basis van het depressiepreventieproject Happyles (www.happyles.nl), de Gamen Infolijn (www.gameninfo.nl) en

de Gokken Infolijn (www.gokkeninfo.nl). Hij werkte daarnaast aan een scala van projecten op het terrein van opvoedingsondersteuning (Triple P), KOPP/KVO (Kinderen van ouders met psychische problemen/Kinderen van verslaafde ouders), game- en kansspelverslaving, maar bovenal alcohol- en drugspreventie. De rol van implementatie op interventie-effecten had daarbij zijn bijzondere interesse. Dit resulteerde onder andere in de (mede-)oprichting van het Nederlands Implementatie Collectief.

Vanaf 2014 werkt Goossens als projectleider van het preventieprogramma *Uitgaan, Alcohol & Drugs* bij het Trimbos-instituut. Vanuit die functie levert hij een bijdrage aan het voorkomen en terugdringen van alcohol- en drugsgerelateerde gezondheidsschade onder uitgaande jongeren. Daartoe schreef hij samen met collega's *Het Grote Uitgaansonderzoek 2013*, de *Strategische Verkenning Uitgaansdrugs 2015* en de leidraad *Alcohol en drugs bij evenementen* voor gemeenten. Tevens is hij kwartiermaker gamen en gokken, met als opdracht om de preventie van problematisch gamen en gokken op een hoger plan te tillen, zowel binnen als buiten het Trimbos-instituut. In 2016 bevorderde hij tot senior wetenschappelijk medewerker en werd binnen het Trimbos-instituut benoemd als Director of the Centre of Implementation.

zijn toekomstdroom is om één keer in zijn leven een grote wereldreis te maken.

Other publications

Journal articles (English)

- Lammers, J., Goossens, F., Conrod, P., Wiers, R., Engels, R., & Kleinjan, M. (subm.). Effectiveness of a selective prevention program targeting personality risk factors for youth alcohol use and misuse: results of subgroup analyses.
- Lammers, J., Goossens, F., Conrod, P., Engels, R., Wiers, R., & Kleinjan, M. (2015). Targeting personality risk factors for alcohol use and binge drinking among young adolescents: results of the Dutch Preventure trial. *Addiction*, 110(7), 1101-1109.
- Quigg, Z., Hughes, R., Bellis, M.A., Hasselt, N. van, Calafat, A., Košir, M., Duch, M., Juan, M., Voorham, L. & Goossens, F.X. (2014). Incidents of harm in European drinking environments and relationships with venue and customer characteristics. *International Journal of Alcohol and Drugs Research*, 3, 269-275.
- Hughes, R., Quigg, Z., Bellis, M.A., Calafat, A., Van Hasselt, N., Rosir, M., Voorham, L., Goossens, F.X., Duch, M., & Juan, M. (2012). Drunk and disorganized: relationships between bar characteristics and customer intoxication in European drinking environments. *International Journal of Environmental Research and Public Health*, 9, 4068-4082.
- Lammers, J., Goossens, F., Lokman S., Monshouwer, R., Lemmers L., Conrod, P., Wiers, R., Engels, R., & Kleinjan, M. (2011). Evaluating a selective prevention programme for binge drinking among young adolescents: Study protocol of a randomized controlled trial. *BMC Public Health*, 11: 126.

Journal articles (Dutch)

- Goossens, F.X., & De Graaf, I.M. (2010). Campagne maakt stellen van opvoedvragen normaal. *Jeugd en Co-Kennis*, 4, 24-33.
- Hoekstra, F., Klein Velderman, M., Wiefferink, R., Van Overveld, C., Ter Beek, J., Buijs, G., Goossens, F., & Paulussen, T. (2009). De adoptiefase van schoolgerichte preventieprogramma's: van boodschap naar beslissing. *Tijdschrift voor Gezondheidswetenschappen*, 87(1), 27-36.

Chapter in dissertation (English)

- Gooren, E.M.J.C., Lier, P.A.C. van, Goossens, F.X., Overveld, C.W. van, Paulussen, T.G.W.M., & Koot, H.M. (2012). *Testing the Role of Social and Emotional Skills in Early School-Age Development of Peer Relations*. In: Gooren, E.M.J.C., *Peers Matter: Social Problems in Kindergarten and the Development of Psychopathology (Dissertation)* (pp. 97-126). Amsterdam: VU Amsterdam.

Chapter in book (Dutch)

- Van Santvoort, F., & Goossens, F.X. (2013). *Preventieve ondersteuningsgroep voor kinderen (8-12 jaar) van ouders met psychische problemen en/of verslavingsproblemen: de Doe-praatgroep*. In: C. Braet, & S.M. Bögels (Eds.), *Protocolaire behandelingen voor kinderen en adolescenten met psychische klachten, deel II* (pp. 467-491). Amsterdam.

Research reports (Dutch)

- Van Leeuwen L. & Goossens, F.X. (2016). *Problematisch gamen: aanbevelingen voor preventie*. Utrecht: Trimbos-instituut.
- De Greeff, J., Goossens, F.X., Cooke, H., Sannen, A.M.L., & Van Hasselt, N.E. (2016). *Alcohol en drugs bij Evenementen*. Leidraad voor gemeenten. Utrecht: Trimbos-instituut.
- Van Leeuwen L., Goossens, F.X., & Lammers, J. (2015). *Preventieplan Kansspelen op Afstand. Preventiekoersen voor het voorkomen van (online) kansspelverslavingsproblematiek*. Utrecht: Trimbos-instituut.
- Sannen, A. & Goossens, F.X. (2015). *Preventie van alcohol-, tabaks- en drugsgebruik onder studenten. Inventarisatie van effectieve preventieve interventies voor het MBO, HBO en WO*. Utrecht: Trimbos-instituut.
- Van Dorsselaer, S. & Goossens, F.X. (2015). *Alcohol-, tabaks- en drugsgebruik door studenten. Inventarisatie van (onderzoeken naar) prevalentieschattingen onder MBO-, HBO- en WO-studenten in Nederland*. Utrecht: Trimbos-instituut.

A

- Goossens, F.X. & Van Hasselt, N.E. (2015). *Strategische Verkenning Uitgaansdrugs 2015. Denkrichtingen voor de preventie van (gezondheidsschade door) het gebruik van uitgaansdrugs onder jongeren en jongvolwassenen*. Utrecht: Trimbos-instituut.
- Hoefnagels, C., Onrust, S., Goossens, F., Van Haastrecht, H., Staal, J., Bartelds, S., Van Goor, M., De Graaf, I., & Speetjens, P. (2013). *Nemen kinderen het woord? Effectevaluatie van Klokhuis over kindermishandeling*. Utrecht: Trimbos-instituut.
- Goossens, F.X., Frijns, T., Van Hasselt, N.E., & Van Laar, M. (2013). *Het Grote Uitgaansonderzoek 2013. Uitgaanspatronen, middelengebruik en risicogedrag onder uitgaande jongeren en jongvolwassenen*. Utrecht: Trimbos-instituut.
- Goossens, F.X. (2013). *SCHERP: Lespakket over wapenbezit. Theoretische en kwalitatieve evaluatie*. Utrecht: Trimbos-instituut.
- Valkenberg, H., Van Leeuwen, L., Klein Wolt, R., & Goossens, F. (2012). *Alcohol en jongeren: een vervolgonderzoek onder Spoedeisende Hulpbezoekers*. Amsterdam: VeiligheidNL / Utrecht: Trimbos-instituut.
- Goossens, F.X. (2012). *Verslaving: Maatschappelijke gevolgen. Overlast, geweld, verwervingscriminaliteit, verkeersongevallen en ziekteverzuim*. Utrecht: Trimbos-instituut.
- Goossens, F.X. & Speetjens, P. (2011). *Financiering KOPP/KVO. Een inventarisatie*. Utrecht: Trimbos-instituut.
- Goossens, F.X. & De Graaf, I.M. (2010). *Groepstraining Triple P voor ouders van tieners. Evaluatie van een pilot study*. Utrecht: Trimbos-instituut.