



Turun yliopisto
University of Turku

IMPLEMENTING THE KIVA ANTIBULLYING PROGRAM: WHAT DOES IT TAKE?

Anne Haataja

University of Turku

Faculty of Social Sciences
Department of Psychology

Supervised by

Professor Christina Salmivalli, Ph.D.
Department of Psychology
University of Turku
Finland

Docent Elisa Poskiparta, Ph.D.
Department of Psychology
University of Turku
Finland

Reviewed by

Professor Ersilia Menesini, Ph.D.
Department of Psychology
University of Florence
Italy

Researcher Kyrre Breivik, Ph.D.
Regional Centre for Child and Youth Mental
Health and Child Welfare
Uni Research Health
Norway

Opponent

Professor Ersilia Menesini, Ph.D.
Department of Psychology
University of Florence
Italy

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

ISBN 978-951-29-6552-6 (PRINT)

ISBN 978-951-29-6553-3 (PDF)

ISSN 0082-6987 (PRINT)

ISSN 2343-3191 (ONLINE)

Painosalama Oy - Turku, Finland 2016

*To all great teachers caring about
children's wellbeing today and
tomorrow*

Implementing The KiVa Antibullying Program: What Does It Take?

Anne Haataja

Department of Psychology

University of Turku

Finland

ABSTRACT

There is variation in how teachers and schools implement bullying prevention programs. Although this variation has been discussed, there has been little empirical research concerning the relationship between implementation fidelity and program outcomes. This thesis contains three studies, each of them in the context of implementing the KiVa antibullying program, and examines teachers' actions in preventing and intervening in school bullying. The first aim of this thesis is to examine implementation degree of the KiVa curriculum and its' association with reductions in victimization and bullying perpetration (Study I). The second aim is to clarify why teachers displayed different degrees of adherence to the KiVa curriculum during a school year (Study II). Thirdly, it is investigated whether recognizing victimization can be difficult for school staff (Study III). In addition to these peer-reviewed studies, the thesis includes a qualitative analysis (unpublished) of the teachers' open answers concerning their implementation experiences. The data were collected from elementary school teachers (Studies I–II; the unpublished study), elementary school students (Study I), and students on the elementary and middle school levels (Study III) during the evaluation of the effectiveness of KiVa antibullying program between 2007 and 2009.

The findings demonstrate that a larger reduction in victimization can be achieved in classrooms where teachers display higher levels of adherence to the KiVa curriculum and invest more time for preparing the lessons. Bullying perpetration, however, was not equally affected by the level of curriculum implementation. With respect to the implementation process over one year, there was significant variation between individual teachers' activity—ranging from systematic and high implementation to declining delivery from lesson to lesson. The sustained actions (high and moderate levels of implementation) were premised on principal support for antibullying work. Lesson preparation was associated with keeping implementation high throughout the school year. The findings also implied that the belief in the effectiveness of the program is important for a higher implementation degree at starting point of the process. Finally, there are severe flaws in teachers' ability to identify students who are victimized. As it turns out, it is possible that only one-fourth of chronically victimized students are helped by the school staff. Especially when the victims are middle-school-aged girls, when they bully others themselves, or when they do not tell adults about bullying, reaching out for them is difficult.

Implementation and dissemination of research-based interventions will take a good deal of time and effort. The findings demonstrate that active implementation is important for improving program outcomes. They also show how implementation can be sustained—there are both individual and interpersonal factors that facilitate or inhibit high-quality implementation. Thus, implications for future research regarding the implementation of school-based programs are suggested.

KiVa Koulu –ohjelman toteuttaminen: Onnistumisen Edellytykset

Anne Haataja

Psykologian oppiaine

Käyttäytymistieteiden ja filosofian laitos

Turun yliopisto

TIIVISTELMÄ

Koulujen ja opettajien välillä on vaihtelua siinä kuinka hyvin ne toteuttavat kiusaamisenvastaisia interventio-ohjelmia. Vaikka tämä vaihtelu on aiemmin tiedostettu, empiiristä tutkimusta on ollut vähän liittyen toteuttamisen tason ja sen tuloksellisuuden yhteydestä. Tämä väitöskirja sisältää kolme osatutkimusta, jotka liittyvät KiVa-ohjelman toteuttamiseen ja tarkastelevat opettajan roolia koulukiusaamisen ennaltaehkäisijänä ja vähentäjänä. Väitöskirjan ensimmäinen tavoite on tutkia KiVa-ohjelman oppituntien toteuttamisen vaikutusta kiusaamisen ja kiusatuksi joutumisen vähentymiseen (osatutkimus I). Toisena tavoitteena on selvittää miksi opettajien sitoutumisessa KiVa-oppituntien toteuttamiseen ilmeni vaihtelua lukuvuoden aikana (osatutkimus II). Kolmanneksi on tarkasteltu onko opettajien vaikeaa tunnistaa pitkäaikaista kiusatuksi joutumista (osatutkimus III). Näiden vertaisarvioitujen osatutkimusten lisäksi väitöskirja sisältää laadullisen tutkimuksen (julkaisematon) opettajien omista KiVa-ohjelman käyttöön liittyvistä kokemuksista. Tutkimuksissa käytetty aineisto on kerätty luokanopettajilta (osatutkimukset I–II ja julkaisematon tutkimus), alakoulun oppilailta (osatutkimus I) sekä ala- ja yläkoulun oppilailta (osatutkimus III) vuosien 2007 ja 2009 aikana, jolloin KiVa-ohjelman vaikuttavuutta on arvioitu.

Tulokset osoittavat, että kiusatuksi joutuminen voi vähentyä voimakkaammin niissä luokissa, joissa opettajat toteuttavat enemmän oppituntien sisältöjä ja käyttävät oppituntien valmisteluun enemmän aikaa. Sen sijaan kiusaamisen vähentymiseen, toteutuksen tasolla ei ollut merkittävää lisävaikutusta. Tulokset liittyen KiVa-oppituntien toteuttamiseen lukuvuoden aikana osoittavat merkittävää vaihtelua opettajien välillä—aina systemaattisesta ja korkeatasoisesta toteuttamisesta oppitunnista toiseen laskevaan. Systemaattisen oppituntien toteuttamisen lähtökohtana on rehtorin osoittama tuki kiusaamisenvastaiseen työhön. Oppituntien valmistelu on yhteydessä korkeatasoiseen oppituntien toteuttamiseen lukuvuoden aikana. Tulokset viittaavat myös, että usko ohjelman tehokkuuteen vähentää kiusaamisongelmia näyttää liittyvän korkeatasoiseen oppituntien toteuttamiseen ohjelman alkuvaiheessa. Lopuksi, opettajien kyvykkyydessä tunnistaa pitkäaikaisen kiusaamisen kohteeksi joutuneita oppilaita on selviä puutteita. Kuten tuloksista ilmenee, on mahdollista, että vain alle neljännes pitkäaikaisista kiusatuista saa apua koulun henkilökunnalta. Kiusatun tunnistaminen on vaikeaa erityisesti sellaisissa tapauksissa, joissa kiusattu oppilas on yläkouluikäinen tyttö, kiusaa itse muita tai ei kerro kiusaamista aikuisille.

Tutkimuspohjaisten interventio-ohjelmien käyttöönotto ja niiden levittäminen vaativat paljon aikaa ja ponnisteluja. Tutkimustulokset ilmentävät aktiivisen toteuttamisen tärkeyttä ohjelman vaikuttavuuden parantamiseksi. Ne osoittavat myös miten ohjelmaa voidaan johdonmukaisesti toteuttaa—on olemassa yksilökohtaisia ja yksilöiden välisiä tekijöitä, jotka edistävät tai ehkäisevät ohjelman laadukasta toteuttamista. Lisäksi tutkielmassa esitetään suuntaviivoja koulujen käyttöön suunniteltujen interventio-ohjelmien toteuttamista koskevaan tutkimukseen.

CONTENTS

ABSTRACT	4
TIIVISTELMÄ	6
CONTENTS	8
PREFACE	10
LIST OF ORIGINAL PUBLICATIONS	13
1. INTRODUCTION	14
1.1 The KiVa Antibullying Program: Development and Implementation	15
1.2 Evaluating Fidelity to Antibullying Programs.....	16
1.3 Measuring Fidelity: Adherence and Implementation Quality	17
1.4 Looking at Fidelity as a Process	18
1.5 Teacher Competence and Curriculum Implementation.....	19
1.6 Recognition of Victimization: A Challenge for Intervention	20
2. AIMS OF THE THESIS	22
3. METHODS	23
3.1 Data Collection.....	23
3.1.1 Participants in Study I.....	24
3.1.2 Participants in Study II.....	25
3.1.3 Participants in Study III	25
3.2 Measures.....	26
3.2.1 Teacher-reported measures	27
3.2.2 Student-reported measures	29
3.3 Statistical Analyses.....	32
4. OVERVIEW OF THE EMPIRICAL STUDIES	35
5. TEACHERS' IMPLEMENTATION EXPERIENCES	38
5.1 Program Instrumentality.....	40

5.2	Perceived Benefits and Rewards	43
5.3	External Factors: Context and Situation	44
5.4	The Link Between Implementation Problems Expressed and Fidelity	45
5.5	Conclusion.....	45
6.	DISCUSSION	47
6.1	Effectiveness of Implementation: Importance of Fidelity Examined.....	48
6.2	Maintaining Fidelity: The Process and Teacher Experiences Examined	49
6.3	Recognition of Victimized Students: The Challenge Examined	51
6.4	Strengths.....	53
6.5	Limitations and Directions for Future	53
7.	REFERENCE	58
	APPENDIX: THE ORIGINAL PUBLICATIONS	63

PREFACE

This dissertation marks the end of my academic journey as a doctoral student at the University of Turku. The five-year journey helped me to grow both professionally and as a person. I have been fortunate to have many people who have made working meaningful and fun, supported me when the journey started to be full of ups and downs—great achievements and unpleasant setbacks instead of nice and steady progress.

Foremost, I am grateful to my two supervisors, Professor Christina Salmivalli and Docent Elisa Poskiparta, for the opportunity to this work in the KiVa project. Christina, you have been a visionary and an empowering supervisor. I would never have progressed further than the start without your support in every step of the publication process. You also managed to generate faith in me in the success of the entire process. Elisa, you have been a resilient supervisor. Your patience and the constructive comments helped me to complete the work. Christina and Elisa, I have felt heard—also beyond the scope of doing research.

I want to address my thanks to people with whom I have had the privilege to author the scientific publications that form this thesis: Dr. Marinus Voeten and Dr. Aaron J. Boulton, you have provided me with your valuable knowledge and expert advice in statistics when working with the first paper. Moreover, your help in data management and an encouraging attitude had a great significance not only in getting research published but also in my coping. Dr. Miia Sainio, your help with the research tools as well as with academic writing were invaluable. You also knew the road ahead, and helped me to continue instead of returning. Dr. Annarilla Ahtola, you helped me to read the map when it turned out that some of the concepts in evaluation research were too ambiguous to be applied to the data collected. Mira Turtonen, thank you for the

idea to examine recognition of victimized students. Ulla Talvenheimo and Tuomas Kaukoranta have proofread the parts of this thesis. Thank you for the excellent comments.

I was honoured to have Professor Ersilia Menesini and Dr. Kyrre Breivik as the preliminary reviewers of this thesis. I respectfully thank them for providing very insightful and valuable reviews which helped me finalize my dissertation. Professor Menesini has agreed to serve as my opponent which I greatly appreciate.

I want to express my sincere appreciation and gratefulness to all the people in the KiVa and Kamu projects and at the Department of Psychology over the years. In particular I want to thank Johanna Alanen, Claire Garendeau, Sanna Herkama, Sari Hesselgren, Ernest Hodges, Marita Kantola, Ari Kaukiainen, Mari Kontio, Oskari Lahtinen, Kätlin Peets, Virpi Pöyhönen, Sanna Roos, Henna Ruohonen, Silja Saarento, Tiina Turunen, Tuija Tuomisto, Katja Valli, Minna Varjonen, and An Yang. For the funding of my work, I thank the KiVa project, funded by the Academy of Finland, and the Faculty of Social Sciences in the University of Turku for awarding me a grant for finalizing the dissertation.

I have been fortunate to work with the teachers and other school personnel in the following elementary schools: Kantokaski (before the academic journey started), Mäntymäki (before and during the journey), and Aurora (during). I thank them all for the eye-opening discussions, collaboration, laughter and their interest for my research. Thus, I owe my gratitude to those teachers who openly shared their experiences in the KiVa trainings over the years. I wish that even a bit of their knowledge and experiences are present also in this thesis. I dedicate this to all teachers, who are committed to ensure that all children can learn and grow in a safe and bullying-free environment.

When I started working in the KiVa project in 2010, dark clouds started to gather over Finnish economy. Unfortunately, during the past consecutive five years, Finnish economy has either contracted or stagnated leading to the stressful situation in many families, including mine. Euripides, a tragedian of classical Athens, has wisely stated that friends show their love in times of trouble, not in happiness. My dear friends, Viivi Vepsä, Annina Kurkio and Virva Wahlstedt, thank you for being such dear friends to me. I am also deeply grateful to you for not forgetting me when I was buried under all the articles. Viivi, I always felt you were walking with me on this journey by believing in me even when the burden felt too heavy.

Finally, I want to thank my husband Timo and our four lovely children, Roosa, Emmi, Helmi and Elias, for your patience and endless understanding. During this process my two eldest became adults and two youngest teenagers. As a mother I may have had less than acceptable listening skills over these years. However, when I look at you all, I know that I got something in my life perfectly right even so. You have made my life very rich and beautiful.

In Kauniainen, in my back yard, listening to baby birds chirping

June 1, 2016

Anne Haataja

LIST OF ORIGINAL PUBLICATIONS

- I. Haataja, A., Voeten, M., Boulton, A.J., Ahtola, A., Poskiparta, E., & Salmivalli, C. (2014). The KiVa Antibullying Curriculum and Outcome: Does Fidelity Matter? *Journal of School Psychology, 52*, 479–483. DOI: 10.1016/j.jsp.2014.07.001
- II. Haataja, A., Ahtola, A., Poskiparta, E., & Salmivalli, C. (2015). A Process View of Implementing an Antibullying Program: How Teachers Differ and What Explains the Variation? *School Psychology Quarterly, 30*, 564–576. DOI: 10.1037/spq0000121
- III. Haataja, A., Sainio, M., Turtonen, M., & Salmivalli, C. (2015). Implementing the KiVa Antibullying Program: Recognition of Victimized Students. *Educational Psychology, 36*, 595–611. DOI: 10.1080/01443410.2015.1066758

The original publications have been reproduced in the Appendix of this thesis with the permission of the copyright holders. All rights reserved. Publication I © 2014 by Elsevier; Publication II © 2015 by American Psychological Association; Publication III © 2015 by Taylor & Francis.

1. INTRODUCTION

World Health Organization (WHO) has defined schools as health promoting when they constantly strengthen their capacity as a healthy setting for living, learning and working. Such schools implement policies and practices that respect the well-being and dignity of students. They also provide multiple opportunities for success, and acknowledge good efforts as well as personal achievements (WHO, 1993). The well-being and safety of students, however, is severely threatened when they experience bullying—repeated and intentional aggressive behavior committed by one or more children against a physically or socially less powerful peer (Olweus, 1993; Salmivalli, 2010; Smith & Brain, 2000). Unfortunately, despite of a declining trend in bully and victim rates (Molcho et al., 2009), it is a fairly common problem in schools: 10–20% of students are frequently bullied by their peers (Bradshaw, Sawyer, & O’Brennan, 2007; Fekkes, Pijpers, & Verloove-Vanhorick, 2005; Solberg & Olweus, 2003).

In a nationally representative Finnish School Health Promotion Study (SHP¹) in 2013, as many as 68% of students from middle school (aged 14–16 years) felt that school personnel had not taken actions against bullying. While such a pessimistic view of teacher intervention is alarming, there is relatively little systematic research on the teachers’ sustained actions against bullying, including both prevention and intervention. To this end, in this thesis I examine the fidelity of implementation of the KiVa antibullying program and its’ association with program outcomes. Then, I provide a process-view of curriculum implementation by identifying if and why variation between teachers occurred. Finally, I focus on the ability of school staff to reach out to victimized students.

¹ Available only in Finnish <http://www.thl.fi/fi/tutkimus-ja-asiantuntijatyo/vaestotutkimukset/kouluterveyskysely/tulokset/tulokset-aiheittain/tapaturmat-ja-vakivalta>

1.1 The KiVa Antibullying Program: Development and Implementation

In Finland the first steps toward nationwide bullying prevention and intervention policies were taken in 1998 and 2003 through the Finnish Basic Education legislation. Accordingly, Finnish schools were not only obligated to have a strategy or action plan against violence, bullying, and harassment, but to execute the plan, supervise its implementation and the adherence to it. At that time, there were no evidence-based programs available and schools were developing their own plans. In 2006, the Finnish Ministry of Education financed the development of a research-based antibullying program at the university of Turku. From the very beginning, the shared vision of politicians and researchers was to develop a program, which would be suitable for nationwide implementation in Finland. The program was entitled KiVa, which is an acronym from the Finnish words *Kiusaamista Vastaan* meaning against bullying. Additionally, the Finnish word “kiva” denotes “nice”. When empirical evidence of the KiVa program’s effectiveness (reported in Kärnä et al., 2011b) was found, the nationwide program diffusion started in 2009.

The KiVa program includes both universal and indicated actions. The *universal actions* such as the KiVa curriculum (lessons and online games) are directed to all students and focus mainly on bullying prevention. There are three different developmentally appropriate curriculum units—Unit 1 (for children 7–9 years old), Unit 2 (for children 9–12 years old), and Unit 3 (for adolescents 13–16 years old). In elementary schools, the curricula in Unit 1 and Unit 2 include 20 hours of student lessons (10 lessons lasting for 90min each), which are scheduled so that lessons are systemically carried out every month over the course of a school year. In middle schools (Unit 3), the contents are organized within four themes to be implemented as a series of lessons or during theme days. The topics of each unit cover a variety of issues related to group

interaction and group pressure, the mechanisms of bullying, and especially, what students can do together in order to counteract bullying and support their victimized peers. The *indicated actions* are to be used to tackle bullying when it has emerged. These actions consist of series of discussions between the adults in the school, the perpetrators of bullying and the targeted students. In each school implementing KiVa, there is a KiVa team whose members are responsible for organizing the discussions.

The KiVa antibullying program was found to be effective in reducing bullying and victimization both in a randomized controlled trial conducted 2007–2009 (Kärnä et al., 2011b, 2013) and during the broad roll-out in Finnish schools (Kärnä et al., 2011a). However, relatively little is known about the teachers' adherence to the program and how well the school personnel have recognized systematic bullying. Therefore, the aims of the present thesis are: (a) to examine whether stronger effects of the KiVa program can be achieved with improved fidelity, that is, a higher degree of adherence to the curriculum as well as higher quality of implementation related to the curriculum content; (b) to provide insight into the preconditions of success of evidence-based antibullying programs more generally—for instance, which factors are related to program sustainability across a school year; and (c) to examine why teachers have difficulties to reach out victims of bullying.

1.2 Evaluating Fidelity to Antibullying Programs

Researchers have long asserted that the effectiveness of prevention programs depends on implementation fidelity, which is a degree to which teachers and other program providers implement programs *as intended* by the program developers (Dane & Schneider, 1998; Durlak & DuPre, 2008; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Ryan & Smith, 2009). In a review of the social and emotional interventions, Durlak, Weissberg, Dymnicki,

Taylor, and Schellinger (2011) found that about a half of the studies had not paid attention to implementation fidelity. In 35% of the studies high levels of fidelity was achieved whereas in the remaining studies (22%) implementation problems were reported, and the positive effects remained minimal. Overall, a level of 100% fidelity to program content is rarely reached (Dane & Schneider, 1998; Durlak & DuPre, 2008; Dusenbury, Brannigan, Falco, & Hansen, 2003), regardless of assessment methods, targeted students, or types of programs.

Very limited attention has been paid to implementation fidelity when the effectiveness of antibullying programs has been evaluated (for meta-analysis, see Ttofi & Farrington, 2010; Vreeman & Carroll, 2007). Only few studies have used implementation information or tested the association between fidelity and program outcomes; the findings have been mixed (Hirschstein, Edstrom, Frey, Snell, & MacKenzie, 2007; Low, Ryzin, Brown, Smith, & Haggerty, 2014; Olweus & Kallestad, 2010).

1.3 Measuring Fidelity: Adherence and Implementation Quality

When implementation fidelity has been measured, it often focuses on two dimensions: adherence to program content being implemented at one time (how much was done), and the quality of implementation (how well the content was delivered). In many school-based studies, the analysis of implementation data has focused on the *adherence degree* to program contents delivered to targeted individuals (e.g., children) or groups (e.g., classrooms). It has been operationalized, for instance, as the number of lessons delivered, percentage of learning tasks covered, or amount of time program delivery lasted as reported by teachers (Elliott & Mihalic, 2004; Jones, Brown, & Lawrence Aber, 2011; Ennett et al., 2011; Dane & Schneider, 1998; Durlak & DuPre, 2008; Dusenbury et al., 2003; Cross, Hall, Hamilton, Pintabona, & Erceg, 2004).

Another aspect of fidelity is the *quality of implementation* or delivery competence. Several studies have used classroom observations (naturalistic or video recordings) to rate aspects of implementation quality, such as program-related instruction skills, the degree of contents taught correctly (Goncy, Sutherland, Farrell, Sullivan, & Doyle, 2015; Hansen, Pankratz, & Bishop, 2014; Kam, Greenberg, & Walls, 2003; Lillehoj, Griffin, & Spoth, 2004; Melde, Esbensen, & Tusinski, 2006), the level of student participation and engagement, or teachers' sensitivity to students' responses (Hahn, Noland, Rayens, & Christie, 2002; Hirschstein et al., 2007; Melde et al., 2006; Mihalic, Fagan, & Argamaso, 2008; Pettigrew et al., 2013; Resnicow et al., 1998; Tobler & Stratton, 1997). In practice, however, classroom observations require a great deal of resources, especially when the research design includes a large number of schools. Besides observations, teachers' own positive perceptions of the program and preparedness can be used to indicate implementation quality (Dane & Schneider, 1998). In bullying research teacher-reported program knowledge has been found to be linked with higher degree of implementation adherence (Kallestad & Olweus, 2003).

1.4 Looking at Fidelity as a Process

Whilst teachers have a central role in antibullying interventions, there is likely to be variation across teachers in program delivery (Ahtola, Haataja, Kärnä, Poskiparta, & Salmivalli, 2013; Kallestad & Olweus, 2003). Kallestad and Olweus (2003), for instance, found that the use of antibullying curriculum can be less-than-ideal, showing low frequency of some program elements. During a long-running program, teachers' lesson adherence might vary across time.

According to Concerns-Based Adoption Model (CBAM; G. E. Hall, Loucks, Rutherford, & Newlove, 1975) the level to which teachers use a given innovation develops through a set of stages before they become familiar and

confident with the material. The progress in the delivery of a program requires that teachers' concerns regarding themselves as program users, implementation tasks, and program benefits are acknowledged (Hall, 2013). Some teachers, however, do not adhere to a new program at all or implement it with reduced quality over time (Hall et al., 1975; Hall, 2013). With regard to studies on bullying prevention, there has been no systematic investigation on the implementation process, meaning how often or how well the prescribed content has been delivered *over time*. Besides implementation information, teachers can provide judgments about the implementation process in general, for example, by evaluating the clarity of teacher manuals and student engagement. The viewpoints of teachers are useful for understanding better the findings of quantitative data, enhancing the use of the program, or even updating the content. Later in this thesis, I will turn to user feedback, interpreting it and quoting the teachers themselves (unpublished study).

1.5 Teacher Competence and Curriculum Implementation

Teaching skills contain knowledge of *subject matter* (i.e., knowledge of the content to be taught) and *lesson structure* (i.e., knowledge required to construct and deliver a lesson) (Leinhardt & Greeno, 1986; Shulman, 1986). Furthermore, motivating students to participate, organizing co-operation, and dealing with student (mis)behavior are essential aspects of teachers' competence. For instance, effective classroom management can increase pro-social behavior and decrease aggression among peers (Bergsmann, Van De Schoot, Schober, Finsterwald, & Spiel, 2013; Luckner & Pianta, 2011), which both are at the center of enhancing a positive learning environment. With respect to bullying prevention, recent findings from classroom observations have shown that student-oriented instructional strategies (i.e., encouragement and interest towards students) are associated with higher levels of student responsiveness

(i.e., active participation and following the rules) during program sessions (Goncy et al., 2015).

In this thesis, I consider teacher competence and willingness to implement to be premises for an active and committed lesson delivery. Both can be improved by teachers themselves through time devoted for planning which, in turn, builds confidence to use different learning techniques with students and fidelity to the curriculum in its entirety. These aspects are used in Study I and Study II, which focus on teacher actions in their classrooms. Specifically, in Study II we² tested the influence of the teachers' beliefs in the effectiveness of program, perceptions of support from the principal, and preparatory training on implementation process. In addition, I have employed qualitative research to assess the elements that were present in teachers' experiences regarding the delivery of the KiVa program (unpublished study).

1.6 Recognition of Victimization: A Challenge for Intervention

Challenging behavior, as teachers view it, includes students' noncompliance and disruptive behavior as well as bullying (Snell, Berlin, Voorhees, Stanton-Chapman, & Hadden, 2012). Many teachers, however, can be unaware of how extensive a problem bullying is (Bradshaw et al., 2007; Craig, Henderson, & Murphy, 2000). Recently, Espelage, Polanin and Low (2014) carried out a school-level study of teacher and school staff perceptions about school environment. They found only a modest correlation ($r = .52$ to $.55$) between staff perceptions of bullying as a problem and student-reports of peer victimization and bully perpetration. Detecting relational and indirect bullying, as compared with physical bullying, can be especially challenging (Boulton,

² When using the pronoun "we" I refer to the authors contributing to the original publications included in this thesis.

1997; Craig et al., 2000; Hazler, Miller, Carney, & Green, 2001; Mishna, Scarcello, Pepler, & Wiener, 2005).

Each school implementing KiVa has a team (a group of 3 to 4 adults) dealing with identified cases of bullying. During the evaluation phase, the work of the KiVa teams was effective, regarding newly emerged cases of bullying. Namely, a large majority of the victimized students (98%), who had been recognized at school, and whose case had been tackled by the KiVa teams (utilizing the series of group discussions) felt that their situation had improved significantly or bullying had stopped completely (Garandeau, Poskiparta, & Salmivalli, 2014). However, in the light of prior research showing weak association between student and teacher agreement on victimization and bullying (Wienke Totura, Green, Karver, & Gesten, 2009), it is seems that not all victims are recognized by the school personnel. Considering the relatively high prevalence of self-reported victims in schools and the low number of cases handled by KiVa teams during the school year as shown by Garendeau and colleagues (2014), the disparity becomes even more evident. In Study III, we examined how well teachers in schools using the KiVa program for the first time were able to identify victims of school bullying. Specifically, we were interested in knowing what factors were associated with teacher recognition of chronically victimized students (or lack of it).

.

2. AIMS OF THE THESIS

The aim of this thesis was to examine teachers' implementation of the KiVa antibullying program.

The specific research questions were as follows:

1. Does teacher-reported implementation fidelity, meaning a higher degree of lesson adherence, delivery time and lesson preparation, have a positive impact on the program's effectiveness, that is, reductions in classroom-level bullying and victimization? (Study I)
2. How do teachers adhere to the curriculum content over time? (Study II)
3. Which factors influence the implementation of the KiVa antibullying curriculum over a school year? (Study II)
4. Which factors make the identification of long-term victims challenging for school staff? (Study III)
5. How teachers have experienced implementation of the KiVa program? (Unpublished study)

3. METHODS

3.1 Data Collection

Each of the three studies utilizes data from a different sample of participants (students or their teachers) who all were studied or worked in intervention schools in the randomized controlled trials (RCT) of the KiVa antibullying program during 2007–2009. This selection of intervention schools was done through a random stratified sampling procedure. Altogether 275 schools (of all 3,418 comprehensive schools) volunteered to participate in the study. Among these schools, a total of 77 elementary schools started to implement KiVa; 39 schools providing education for students in Grades 1 through 6 (Studies I and II), and 38 schools providing education for students in Grades 7 through 9 (included in Study III).

Implementation data on teachers' adherence to the lesson content were collected from 439 teachers working in 77 elementary schools. The homeroom teachers were asked to fill in the lesson booklet immediately after each given lesson over the course of the intervention year, starting from mid-August to the end of May. This information on fidelity to the curriculum was used in Study I and Study II. Another source of teacher data was a web-based questionnaire (filled in the beginning of the school year) mapping out several teacher characteristics. In Study II, the sample of teachers consisted of those responding on both lesson booklets as well as the teacher questionnaire. A third type of teacher-reported data was related to the work done in schools' by KiVa teams in order to stop bullying. Teachers documented all cases of bullying, which came to their attention. In Study III, which focused on recognizing long-term victims of bullying, the data from those schools were used who had documentation on

indicated actions (e.g., all cases of bullying which were tackled in the KiVa teams of the intervention schools).

In Study I and Study III, teacher and student data were used. All students with parental consent for participation received personal passwords to log in to the internet-based questionnaire. There were three waves of measurement: the pre-test in May (T1), one in December-January (T2), and the final post-test in May (T3; one year after the baseline). In Study I, the change score of student-reported victimization and bullying between the final post-test (T3) and the pre-test (T1) was used as the outcomes (i.e., pre-scores were subtracted from the post-scores to get the difference score which indicated the magnitude of effects). In Study III, the stability of student-reported victimization was measured twice at T1 and T2 with 6 months apart (Kärnä et al., 2011b, 2013). The term bullying was defined for the students similarly to the revised Olweus' Bully/Victim Questionnaire (OBVQ; Olweus, 1996); this definition emphasizes the repetitive nature of bullying and the power imbalance between the victim and the bully.

3.1.1 Participants in Study I

The total number of students in the 77 intervention schools was 8452. In the main analysis in Study I (see Table 1), we used data from 7413 students (49% girls and 51% boys) and 417 teachers from 76 schools. The 1039 (11%) students were excluded because they did not participate in either the pre-test or the post-test, or participated only in the pre-test but were not in the schools during the intervention year. Another reason for excluding the students from the final sample was lack of parental consent (6%).

Of the 439 teachers in the intervention schools, 22 were excluded from the analyses because of the missing information of their students. None of these teachers co-operated during the implementation study. From the sample of 417 teachers included in the analyses, 332 returned the lesson booklets, from which

all implementation variables were derived and used in the analyses. We had 85 teachers without information on implementation (20% of the teachers). To deal with the missing value patterns and to investigate the relationship between degree of implementation and the effects on program outcomes, we used full information maximum likelihood (FIML).

3.1.2 Participants in Study II

In Study II, we used a sample of 282 teachers (65% from the total sample of 439) from 69 schools (see Table 1). These teachers had filled out a web-based survey at the pre-test (containing information on attitudes and beliefs related to bullying), and they had returned lesson booklets including information of curriculum content delivered over a school year (outcome). The sample consisted of 78% females. The majority of teachers (78%) were a permanent appointment, and 87% of them taught in regular education classrooms. The average extent of teaching was 14.7 years ($SD = 9.1$; range 0 to 36 years).

3.1.3 Participants in Study III

In Study III, we used data from the grade cohorts 3–6 (elementary school) and 8–9 (middle school) in order to examine the recognition for stable victimization. We focused on a sample of 348 long-term victims (from 76 schools) who perceived themselves victimized 2–3 times a month before the program began in May (T1) and continued to feel so after the five months of KiVa intervention in December/January (T2). The majority of long-term victims were boys (60.3%), and 33.6% of them were in middle school. They represented 3.8% of all students in the sample.

Table 1
Study samples from the KiVa antibullying program

	<i>Study I</i>	<i>Study II</i>	<i>Study III</i>
Grade cohorts	1–6	1–6	3–6 and 8–9
Schools, <i>N</i>	77	77	116
Teachers, <i>N</i>	439	439	–
Students, <i>N</i>	8 452	–	9 428
Active parental consent	94%	–	92%
Response rate	79% (T1); 90% (T3)	64 %	93% (T1); 89% (T2)
Unit(s) of analysis	Student and classroom	Classroom	Student and school
Teachers and/or schools in main analyses, <i>n</i>	417 teachers; 76 schools	282 teachers; 69 schools	76 schools
Students in main analyses, <i>n</i>	7 413	–	348
Boys, %	51%	–	66%
Age, <i>M</i>	10 years (T1)	–	12 years (T2)

3.2 Measures

The measures that reflected teachers' implementation adherence were obtained from lesson booklets. All lesson-specific activities for each of the 10 lessons were listed in the booklets. The teachers marked which activities they had implemented. Thus, teachers estimated the time they had spent (in minutes) for preparing and delivering each lesson, and the proportion of students being active during a lesson. These measures were in many cases averaged across many items (Study I and II). If there were missing data on all measures asked in the booklets, the lesson was considered not delivered.

Another source of data for teacher-reported measures was the internet-based questionnaire that was filled in the beginning of the school year (Study II). Items and scales were developed for mapping out teachers' self-efficacy for classroom management, the support from the school's principal for antibullying work, the teachers' beliefs in the effectiveness of the program, and participation to pre-implementation training.

Student reports related to victimization and bullying were collected across varying time points (Study I and III). However, the studies differed to some degree on how the measures were created and whether the measure was based on self- or peer reports (explained in the description of each measure). For instance, in Study I we used latent variables whereas in Study III multiple-item scales were used.

3.2.1 Teacher-reported measures

Lesson adherence (Study I). The first measure, designed to assess the total degree of lesson adherence to the KiVa curriculum, was calculated as the proportion of tasks delivered for each lesson. These proportions were averaged over the ten lessons. The average proportion of curriculum tasks completed ranged from 3% to 100% with a mean of 68% ($SD = 20$).

Duration of lessons (Study I). The number of minutes spent for teaching the lesson content was averaged across the lessons a teacher reported to have delivered. The duration of lessons ranged from 25 to 180 minutes, with a mean of 79 minutes ($SD = 19$).

Lesson Preparation (Study I). The time spent in preparing the lessons was calculated by averaging the reported numbers of minutes across the lessons delivered by a teacher. The time devoted to preparing a lesson ranged from 8 to 98 minutes, with a mean of 29 minutes ($SD = 16$).

Preparation hours (Study II). The total time that teachers spent preparing the lessons was calculated by summing the teacher-reported preparation time (as given in minutes) across the ten double lessons. The total preparation time ranged from 0 to 838 minutes. The estimates were rescaled and divided by 60 minutes, reflecting the amount of hours used for planning. The average score for planning was 4.11 hours ($SD = 2.38$)

Student engagement (Study II). After delivering each lesson, teachers were asked to rate the student engagement during the lesson: “Estimate the proportion of the students who participated enthusiastically in this lesson”. The answers were given on a four-point ordinal scale ranging from 1 (0–25%), 2 (25–50%), 3 (50–75%) to 4 (75–100%). The average score of student engagement across lessons was 3.38 ($SD = .60$).

Participation in training (Study II). Teachers were asked whether they had participated in pre-implementation training (no=0, yes=1). The majority of the teachers responding (63.3%) had participated. Responses were missing from 18 teachers (6.3%).

Self-efficacy for classroom management (Study II). Teachers were asked to rate how well they function in different classroom situations (6 questions) such as “To what extent are you able to calm down a disruptive and noisy students?”, “To what extent are you able to motivate students to behave according to the common rules of the class?”. The nine-point Likert scale (1 = *not at all*, 9 = *very well*) had an internal consistency of .85. The average score on classroom management was 6.13 ($SD = .94$).

Principal support (Study II). Teachers were asked to rate the principal support for antibullying work as assessed by five statements such as “The principal supports the antibullying work in our school”, “The principal ensures that there are enough resources (such as time) for antibullying work”. The five-point

Likert-scale (1 = *completely disagree*, 5 = *completely agree*) had an internal consistency of .89 and the average score was 3.42 ($SD = .58$).

Belief in the effectiveness of the program (Study II). Prior to three statements related to teacher belief in the effectiveness of the program, teachers were reminded of the main contents of the KiVa program (as they had not been implementing it yet). They were then asked to evaluate the extent to which they believed that the program will have an influence on the occurrence of bullying, on the well-being of the victims, and on the students' overall satisfaction of the school. The five-point Likert scale (1 = *very little*, 5 = *very much*) had an internal consistency of .86, with an average score of 2.77 ($SD = .63$).

School's commitment: Proportion of lesson booklets returned (Study III). We used the delivered documentation of antibullying curricula as an indicator of school-level commitment to the KiVa intervention as whole. This variable was used to predict recognition at the school-level. Lesson booklets that were returned from each school were totaled and divided by the number of expected booklets, resulting in a score ranging from 0.0 to 1.0. In our sample, the average proportion of returned booklets was .79 ($SD = .24$).

3.2.2 *Student-reported measures*

Self-reported victimization (Study I). At pre-test (T1) and post-test (T3) assessments four items representing typical forms of victimization (verbal, exclusion, physical, and manipulative) were used as indicators of latent variables for victimization. Specific questions of experienced victimization were presented on separate pages and were seen one by one "How often have you been bullied at school in the last two months in this way?" All four items were responded on a 5-point scale (0 = *not at all*, 4 = *several times a week*). The ordinal coefficient alpha (Gadernann, Guhn, & Zumbo, 2012) for the four victimization items at pre-test was .87 and at post-test the coefficient was .88.

Self-reported bullying (Study I). At pre-test (T1) and post-test (T3) assessments four items representing typical forms of bullying others were used as indicators of bullying behavior; “Have you been bullying others in this way in the last two months?” The specific questions of bullying were presented on separate pages so that four forms (verbal, exclusion, physical and manipulative) were seen one by one. All four items were responded on a 5-point scale (0 = *not at all*, 4 = *several times a week*). At pre-test the ordinal alpha coefficient was .86; at post-test the coefficient was .88.

Self-reported direct victimization (Study III). At pre-test (T1) assessment four items representing direct victimization were used; “I was called mean names, was made fun of or teased in a hurtful way”; “I was hit, kicked, or shoved”, “I was stolen money or things from or my things were broken”, “I was threatened or forced to do things I didn’t want to do”. Students were prompted by asking: “Have you been bullied at school during the past couple of months in this way?” They responded in 5-point scale (0 = *not at all*, 4 = *several times a week*). Cronbach’s alpha was .69.

Self-reported indirect victimization (Study III). At pre-test (T1) assessment two items measured indirect victimization: “Other students ignored me completely or excluded me from things or from their group of friends”, “Other students tried to make others dislike me by spreading lies about me.” Students were prompted by asking: “Have you been bullied at school during the past couple of months in this way?” They responded in 5-point scale (0 = *not at all*, 4 = *several times a week*). Cronbach’s alpha was .66.

Peer-reported victimization (Study III) at pre-test (T1) was obtained through peer nominations. Students were asked to nominate an unlimited number of classmates who they perceived as being bullied in the following ways: “S/he gets shoved and hit”, “S/he is called names and made fun of”, and “Rumors are

spread about her/him.” For each student in the sample, the received numbers were totaled and divided by the number of classmates responding which resulted in a score ranging from .00 to 1.00 for each individual student on each item. The proportion scores were averaged across three items. In the present sample, Cronbach’s alpha was .80 for the victimization scale.

Bullying others (Study III) was measured by asking students a question at T2 (Olweus, 1986): “How often have you bullied another student during the past couple of months?” They answered on a 5-point scale (0 = *not at all*, 4 = *several times a week*).

Telling an adult about the victimization (Study III). At time 2 students were asked if they had told anyone about the victimization, and if so, who. The alternatives were “the teacher”, “another adult at school”, “mom, dad, or guardian”, “sibling”, “a friend” and “someone else”. We considered the first three options indicating telling an adult, and created a dichotomous variable of victims’ responses (0 = *has not told an adult*, 1 = *has told*).

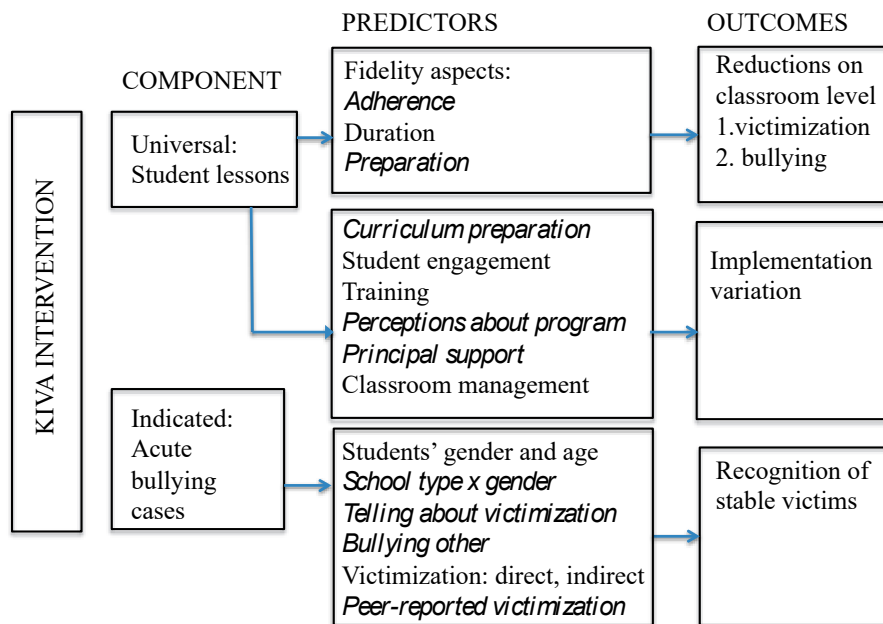


Figure 1. An evaluation map of Studies I–III regarding implementing the KiVa. All significant paths are in italics.

3.3 Statistical Analyses

The objective of Study I was to examine the effects of varying fidelity degree on bullying problems on the classroom-level. The outcome constructs—victimization and bullying—were assessed with four observed variables before the intervention (T1), and again nine months after the intervention (T3). The four forms of bullying and victimization (ordinal variables but treated as continuous) were considered as indicators of bullying/victimization, and used as a latent construct instead of a single item to indicate the change from T1 to T3. It should be noted that the data was also highly skewed. Categorical specifications and testing for measurement invariance were attempted. In

victimization models the findings were robust across categorical and continuous models. However, due to estimation problems with categorical specifications in bullying items the final models were treated as continuous. A longitudinal multilevel structural equation modeling (MSEM) approach (Muthén & Asparouhov, 2011) and a latent difference score model between T1 and T3 (McArdle & Nesselroade, 1994; McArdle, 2001) was used to investigate how the relationship between changes in victimization and bullying were associated with teacher fidelity to lessons. In the models the change score was regressed on the pretest scores (McArdle, 2009; Selig & Preacher, 2009) instead letting it merely be correlated (making no difference for the results at classroom level). This selection implied that initial status predicted change. These models were run using Mplus 7.0 (Muthén & Muthén, 1998–2012).

In Study II, the objective was to examine the variation in the degree of implementation fidelity over time and to predict why some teachers showed higher levels of lesson adherence than others. Teachers were classified into latent (unobserved) classes based on their similar implementation pattern over time. This was done by a factor mixture modeling (FMM) approach which probabilistically assign each individual into subgroups as being represented by a categorical latent variable (Lubke & Muthen, 2005; Lubke & Neale, 2008). An exploratory factor structure (Asparouhov & Muthen, 2009) was applied to the models. In order to predict the probability of a categorical group membership, multinomial logistic regression models were performed using maximum likelihood estimation in Mplus 7.11 (Muthén & Muthén, 1998–2012). Although FMM can be less than optimal in modeling the change as a function of time, FMM was chosen instead of growth mixture modeling because there was no information available about the interval spacing of 10 lessons that might have been irregular between teachers. In fact, in some booklets there were notes indicating that the lessons were not necessary held once a month in all cases.

Moreover, the groupings of the consistent vs. inconsistent implementers could be examined in both analytical methods.

In Study III, the objective was to examine the recognition of long-term victims in schools beginning to implement the KiVa antibullying program. Long-term victimization as an outcome variable was created by selecting students who had been targets of persistent bullying (at least two times a month) across two assessment points; that is, before summer break in May, and again after 4–5 months of implementing the KiVa program in December–January. Multilevel logistic regression analyses were performed with individual features at the within-level and school features at the between-level. Analyses were run using Mplus 7.11 with MLR estimator (Muthén & Muthén, 1998–2012).

4. OVERVIEW OF THE EMPIRICAL STUDIES

STUDY I

Haataja, A., Voeten, M., Boulton, A.J., Ahtola, A., Poskiparta, E., & Salmivalli, C. (2014). The KiVa Antibullying Curriculum and Outcome: Does Fidelity Matter? *Journal of School Psychology, 52*, 479–483. DOI: 10.1016/j.jsp.2014.07.001

The aim of the study was to examine implementation fidelity to the KiVa curriculum using the data from the randomized controlled trial (RCT). Prior to this evaluation study, however, very limited attention had been paid to teachers' implementation fidelity to antibullying programs. With a large sample of 7,413 students (7–12 years) from 417 classrooms within 76 elementary schools, we tested whether the degree to which teachers had adhered to the KiVa curriculum was related to effectiveness of the program, that is, reducing bullying problems in classrooms. Results of multilevel structural equation modeling revealed that after nine months of implementation, lesson adherence and preparation time (but not duration of lessons) were associated with reductions in victimization at the classroom level. No statistically significant effects, however, were found for classroom-level bullying. Overall, our results support the existing literature showing that effectiveness of the program is affected by implementation fidelity.

STUDY II

Haataja, A., Ahtola, A., Poskiparta, E., and Salmivalli, C. (2015). A Process View on Implementing an Antibullying Curriculum: How Teachers Differ and What Explains the Variation. *School Psychology Quarterly*, 30, 564–576.

DOI:10.1037/spq0000121

The aim of this study was to examine if and why variation between teachers occurred during their first year implementation of the KiVa curriculum. The sample consisted of 282 elementary school teachers. The results from factor mixture modeling indicated that there were 3 different types of teacher adherence regarding sustained curriculum implementation. For most of the teachers (55%; group high), implementation adherence was high at the beginning, and remained so over time (except for last lessons). In the second group (26%; moderate), teachers displayed moderate adherence; they utilized approximately a half of the lesson material consistently. Teachers in the third group (19%; surrenders) started high but their curriculum implementation steadily declined. Results in multinomial logistic regression revealed that support for the antibullying work from school principal predicted sustained implementation for groups high and moderate rather than the surrender group. Moreover, implementation at high level throughout school year was predicted by lesson preparation. Teacher beliefs in the effectiveness of the program were positively associated with starting at higher levels of fidelity (high and surrenders). Implementation training, student engagement and classroom management skills were unrelated different degrees of adherence to the KiVa lessons. The findings of the study displayed that both individual and interpersonal factors (including the necessary support from school principal) facilitated the implementation process.

STUDY III

Haataja, A., Sainio, M., Turtonen, M., & Salmivalli, C. (2015). Implementing the KiVa Antibullying Program: Recognition of Stable Victims. *Educational Psychology*, 36, 595–611. DOI:10.1080/01443410.2015.1066758

In this study, we examined recognition of long-term victims in schools that were beginning to implement the KiVa antibullying program. We used a sample of 348 victims in 76 schools that reported victimization at the pre-test and still at wave 2, after five months of program implementation. School personnel were able to recognize and help only 24% of these long-term victims. Multilevel logistic regression analyses revealed that male victims were recognized more often than female victims in elementary school, but the gender of the victim was not linked with recognition in middle school. Telling an adult about victimization as well as higher peer-reported victimization increased the likelihood of recognition by adults, whereas the victimized student's own bullying behavior towards other peers decreased it. However, 80% of the frequently victimized withheld disclosing their harmful experiences during several months. Also, girl victims, middle school students had a higher risk to remain unnoticed, of which school personnel should be aware. This study informed schools in their quest for improving a confidential reporting system of bullying and understanding the risk factors that may make adult support less likely.

5. TEACHERS' IMPLEMENTATION EXPERIENCES

As shown in Study I and Study II (utilizing quantitative data), teachers displayed practically and statistically significant differences in their adherence to the KiVa antibullying curriculum. Thus, I expected that their subjective experiences regarding implementing the program might vary too. In the research literature related to antibullying interventions, however, teacher perspective and mixed method design (i.e., using both quantitative and qualitative data) has been overlooked. My aim of using qualitative data, that is teachers' open feedback about the program (i.e., their written comments of program characteristics and its' implementation) was a) to examine how teachers felt about the program, b) to explore the themes that were present in their implementation experiences, and c) to examine if implementation concerns/problems that teachers expressed were associated with their implementation fidelity.

A total of 295 elementary school teachers (from among the 439) responded to the internet-based questionnaire in May, after the first year of implementation of KiVa. The questionnaire started with demographic questions including questions on sex, experience in teaching, job role in school as well as implementing the program (i.e., being a team member and/or delivering the lessons). Teachers were also asked to evaluate several implementation related issues such as which lessons of all 10 they had delivered, the number of discussions they had been involved with, to what extent they believed in program effectiveness and in which grade they had delivered the lessons. After this quantitative section in the questionnaire, teachers had an opportunity to write about their implementation experiences.

In all, 221 teachers (50% of all targeted teachers) had written feedback at the end of survey. Majority (61%) had a class in Grades 4 through 6. On average,

they had 14.4 years experience in teaching and had implemented 8.6 KiVa lessons. Nearly half (49.5%) reported that they had participated in the school network meetings where they had met other teachers from different school and one person from the KiVa research project during a school year. With respect the grades given to program (on a 5-point grading scale ranging from 0 = *fair*, 1 = *acceptable*, 2 = *good*, 3 = *very good*, 4 = *excellent*), program was evaluated as (very) good with an average of 2.56 ($SD = 0.81$).

I used thematic analysis (Braun & Clarke, 2006) to identify the themes within open statements that were given by 221 teachers. First, I coded the statements at the word or phrase level in order to analyze whether teachers had (dis)liked the program. As a result, four different types of user feedback were deduced: Compliments (35%), complaints (24%), a mixture of compliments and complains (30%), and neutral statements without any (un)favorable tone (11%). I found the themes repeated (Braun & Clarke, 2006) by re-reading through the entire data set systematically, and giving equal attention to each statement in teachers' feedback. It was relatively common that one statement contained more than one issue. For instance, while evaluating the feasibility of the program material, teachers described their own situation. In analyses of this kind, three key themes were identified in the data, and they were related to 1) program instrumentality, 2) perceived benefits and rewards, and 3) external factors including organization context and research situation. In order to see how the themes were distributed across the feedback categories, I used contingency table (crosstabulation) analysis. The results are presented in Figure 2. Those providing compliments only had expressed the benefits more likely (52% vs. 24%, 15.4%, 7.7%) than the teachers providing other types of feedback ($\chi^2(3) = 12.79, p = .01$). They also had mentioned external factors less likely in their comments (17.5% vs. 35%, 32.5%, 17.5%) than the others ($\chi^2(3) = 17.61, p = .001$).

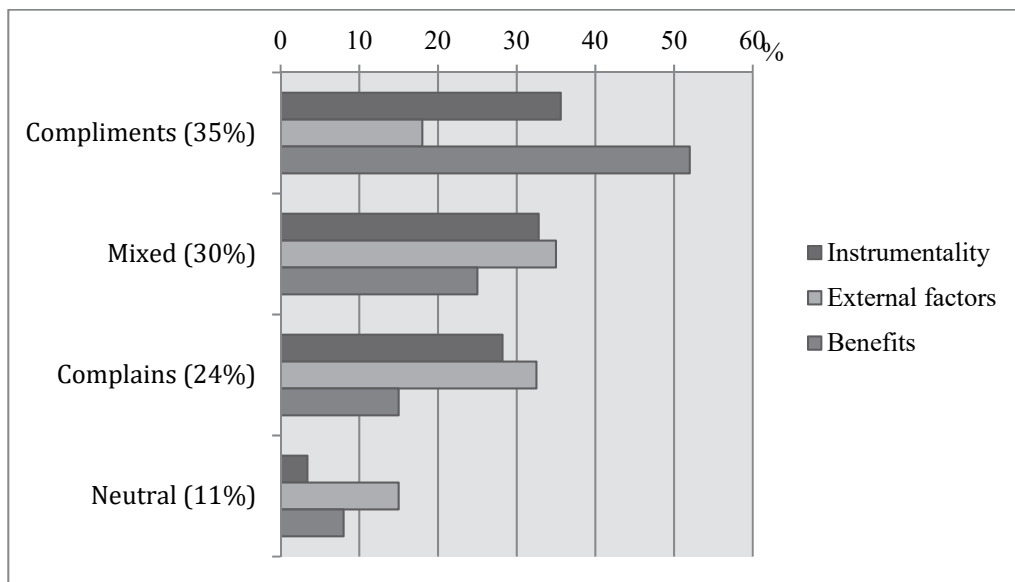


Figure 2 Crosstabulation of four qualitatively different teacher feedback (%) and three specific implementation themes teachers ($n = 221$) included in their written statements.

Next, I shall provide a summary of the qualitative findings. I have chosen teacher quotations because they represent common views within the themes. The background information of the teachers (if available) regarding their participation on the KiVa training, their degree of adherence to lessons (high, moderate, surrenders) and the years of teaching experience are presented within parentheses. There were no statistically significant differences in the feedback quality (positive, negative, mixed, neutral) between the three adherence groups ($\chi^2(6) = 9.06, p = .17$).

5.1 Program Instrumentality

As many as 80% of teachers had commented the instrumentality of the program. Many of them had also provided their views about how well the program fits in the general schedule. If teacher responses contained issues such as ease of teaching, clarity of instructions, time requirements, and planning for lessons,

then the responses were classified under program instrumentality. More than one third of the teachers, who complimented the program, mentioned the program material as a completely ready-made package, or as easy to use:

The materials were well-made, implementing an important topic was made easy. (Untrained teacher, no other information available)

Many comments also implied that teachers found the lessons (relatively) easy to prepare:

The material has been easy to use. Simple guidelines, and the lessons have not required excessive preparation. Students have responded well and are looking forward to new lessons. (Untrained teacher, moderate, 3 years of experience)

While most of the compliments regarding instrumentality appeared to be about lessons, some teachers pointed to the indicated actions:

The method in solving bullying incidents seems to work well. Keeping a record of the incidents is good for all students. The same method in all classrooms and in all incidents of bullying within the school. (Trained teacher, moderate, 25 years of experience)

Written statements revealed that one third of the teachers were not completely satisfied with the program material. On one hand, the teachers' dissatisfaction seemed to be related with the length of the curriculum, or having limited classroom time available for the program:

There were a lot of good things in the package. In order for to follow and implement the program, the topics should be condensed and the layout should be clearer. Now, the manual is cumbersome to use. (Trained teacher, moderate, 7 years of experience)

On the other hand, some teachers were pleased with only one of the KiVa components, that is, universal or indicated, but not with both:

On the whole, the lessons are excellent. The discussion model and the use of the team in solving bullying incidents are clumsy and contrived. (Trained teacher, surrenders, 5 years of experience)

The teachers, who were critical of program instrumentality and reported only problems with program implementation, had in many cases perceived the KiVa as time consuming or too intensive in length. The expressions such as “too wide” and “labor intensive” reflected this concern of organizing the lessons:

Too many things/it begins to repeat itself/it takes too much time from lessons. (Trained teacher, no other information available)

In the critical feedback, it was typical that teachers expressed difficulties in organizing a lesson and using the material in larger classrooms:

The material and the lessons are not suitable for teaching large groups in a big school. The lessons were fully loaded with material that was difficult to edit without losing important parts of the whole. (Trained teacher, high, 17 years of experience)

Teachers’ neutral feedback (as I coded it) was very brief so that it was not possible to make any quality judgments:

Things were already familiar, there was nothing new. (Untrained teacher, high, 32 years of experience)

Alternatively, neutral feedback reflected their own situation rather than contained any comments about program quality or implementation challenges:

I was not able to deliver the final lessons in the spring due to rush. Could the program include a tentative schedule, which would help carrying out the lessons? (Untrained teacher, surrenders, 13 years of experience)

5.2 Perceived Benefits and Rewards

Another theme, mentioned in 30% of the responses (65 teachers out of 221), consisted of perceptions of program benefits to their students or to the teachers themselves as the following comment well illustrates:

A problem has always occurred and will always occur. Now we CAN take actions. When we SPEAK LOUDLY about bullying, the young victims of bullying also feel respected. The solutions offered are well received at schools. The KiVa project has absolutely been the best thing for my students during my 30 years of career. MANY THANKS!!!! (Untrained teacher, high, 27 years of experience)

Positive changes in pro-social skills of the students were mentioned:

The results of the program are clearly visible: a supportive and considerate atmosphere can be observed, no more lonely students, everybody together, less quarreling, students have learned to negotiate with each other. (No teacher information available)

If teachers had mentioned that students liked activities or were responsive towards the content, it was considered an expression of rewards:

The KiVa lessons have raised a great deal of discussion in the classroom and the ideas and principles of the lessons have often come to light in everyday practices. (Untrained teacher, high, 16 years of experience)

Thus, there were some positive statements regarding a method in KiVa for handling bullying incidents:

The KiVa procedure has clarified the tackling of bullying incidents and the lessons that have been regularly given have helped students to understand what bullying is and why it is wrong. (Trained teacher, high, 20 years of experience)

There were also some teachers (13 teachers out of 65) who were either skeptical of the effectiveness of the program or had a group of unenthusiastic students:

It is laborious and difficult to know whether it works. (Trained teacher, high 23 years of experience)

For some reason, it seems that many students have some kind of dislike for KiVa lessons. Often, when I am about to have a KiVa lesson, the students are very disruptive and noisy. (Trained teacher, moderate, 5 years of experience)

5.3 External Factors: Context and Situation

A third theme deduced from the teachers' comments, altogether in 36% of the responses, was related to external factors such as technical problems, classroom size, teacher absence and participation on research:

It is difficult to have the lessons due to a lack of time, and booking our school's computer lab is a bit complicated at the moment. There are not many times available for playing the KiVa game or answering the surveys. (Untrained teacher, high, 7 years of experience)

In the mixed feedback, teachers expressed a positive attitude towards KiVa but had difficulties in finding time for lessons or team discussions:

You get carried away with everyday life in schools (=hectic, difficult to find the time for discussions), more shared guidelines for the KiVa-teams, it is a good and useful project! (Untrained teacher, moderate, 5 years of experience)

While in most of the cases external factors were problematic, a handful of teachers (4%) mentioned teacher training and network meetings providing support:

There was a variety of material to be chosen from, training days and collaboration (network and team) were important, the principal should be

committed for everything to work. (Trained teacher, high, 20 years of experience)

Some statements were so brief so that it was unreasonable to make any qualitatively meaningful conclusions about implementation of the program:

Commitment to the whole staff and training too. (Trained teacher, high, 15 years of experience)

5.4 The Link Between Implementation Problems Expressed and Fidelity

According to the Concerns-Based Adoption Model (CBAM; Hall, Loucks, Rutherford, & Newlove, 1975) the growth in using the new strategies/innovation requires that the concerns teachers express during the process are adequately addressed. Next I will focus on the implementation problems in order to examine if the proportion of problems expressed was linked with the curriculum fidelity. Examples of teacher complains/problems (such as finding time for lessons, planning for lessons, dealing with larger classrooms and/or unenthusiastic students, organizing facilities for student survey; as presented in previous section) were summed up. The amount of problems reported ranged from 0 to 3. On contrast to the expectations that surrenders would have reported more problems, it turned out that in the group of moderate teachers reported more problems ($M = .84, SD = .67$) than teachers in the high ($M = .55, SD = .62$) and surrender groups ($M = .52, SD = .65, F(2) = 3.20, p = .04$).

5.5 Conclusion

Teachers were motivated by immediate facts such as an ease of program implementation (instrumentality) and program benefits to their students that were not restricted to reductions of bullying—teachers observed more pro-

social behavior among their students. Altogether, for more than two thirds of teachers who provided feedback, implementing the KiVa program was (fairly) easy. Overall, it may be that flexibility (i.e., freedom to choose instructions from relatively wide array of activities) is less important to some teachers. Major challenges seemed to be related to finding time and using the instructional material (lesson plans).

It is somewhat surprising that these challenges were lowest in the surrender group. It can be they did not regard bullying prevention as professionally relevant (or important) as the quantitative findings of Study II suggested. It needs to be noted that the results of teacher feedback are limited to 50% of teachers who were responsible for implementing KiVa. Therefore the response proportions reported here might be somewhat different if all the teachers had provided a short feedback and responded the online questionnaire after the pilot. Thus, in the future survey design could be improved by asking pertinent and intelligible open-ended questions to identify the “why”: barriers and facilitators to sustained implementation, and the “how”: program components in need of improvement.

Ideally, this kind of feedback loop provides insights about how the program could be expected to fit in local conditions. Thus, user feedback provides tools for training and providing support to teachers, or even reviewing and updating the program. For instance, people who provide pre-implementation can understand better the processes and differences between teachers and schools implementing the KiVa.

6. DISCUSSION

In this thesis, I have investigated what it takes to implement a curriculum for bullying prevention and how different degrees of teachers' curriculum adherence predict subsequent changes in bullying behavior in their classes. Also, I have reviewed teachers' feedback regarding implementing the KiVa program. Altogether, adherence to curriculum content was close to 70 %, which can be considered satisfactory (Study I). Higher adherence and better implementation quality had positive effects on the success of the program: the reductions in victimization were larger in classes where teachers displayed better fidelity. As shown in Study II, implementation (even in short run of nine months) is an ongoing process where a varying degree of adherence to the curriculum content is evident.

Based on teachers' feedback about the program (unpublished study) many teachers perceived the program as user-friendly, and observed interest among their students during the lessons. In addition to positive views, some teachers reported that implementing KiVa was demanding. They made critical remarks on scheduling and integrating antibullying lessons into the general schedule as well as on organizing student assessment for research purposes.

Besides preventing bullying, another key aspect of the KiVa program is intervening systemically with ongoing bullying that comes to attention. Discussions have been perceived successful by the targeted children (Garandeau, Poskiparta, et al., 2014). However, not all victimized students receive school support. I found that there was a group of students being systematically bullied over a longer period (Study III)—only about one fourth of them were recognized by school personnel. In the following, I will discuss practical implications of the findings and their relation to prior research.

6.1 Effectiveness of Implementation: Importance of Fidelity Examined

A lot can be learnt from evaluating fidelity, the ultimate goal of which is to identify how well a program is implemented in the field (i.e., classrooms, schools) within a certain time frame. Previous research on school-based prevention programs displays a positive relationship between implementation fidelity and outcome (e.g., Durlak & DuPre, 2008; Dusenbury et al., 2003). The findings of Study I are in line with this evidence because the impact of the KiVa antibullying program (i.e., a reduced degree of victimization) was affected by improved fidelity—adherence and lesson preparation, but not by duration of the lesson. For another outcome, reductions of bullying, higher rates of fidelity had no additional effect. As discussed in Study I, the initial proportion of numbers victims was higher than the proportion of bullies, which did not leave much room for change. The behavior of bullies is not always easy to affect, especially if they were powerful and popular (Garandeau, Lee, & Salmivalli, 2014; Rodkin, Farmer, Pearl, & Van Acker, 2006; Vaillancourt & Hymel, 2006). Breaking the imbalance of social power may need more frequency and intensity of the elements that focus on enhancing equality and respect in relationships and evoking a shared norm for not tolerating bullying of any kind. Another interpretation might be that the victims have adopted more effective strategies to cope with bullying. If true, these are all desirable outcomes teachers should be aware of, even when preventing all bullying is beyond the control of the teacher.

Both quantity and quality of implementation can be assessed via teacher reports, as in the present study. Although lesson preparation is less than optimal measure of implementation quality, it probably indicates teachers' motivation to anti-bullying work. Given the cost and effort involved, it would be important to

examine how much bullying and victimization reduce as a result of program high-quality implementation over several years.

When teachers aim to meet the objectives of the lessons (i.e., to increase students' awareness about the influence of the peer group on bullying) their attempts to change the peer dynamics should be understood as a long-term process, which cannot be affected by covering sporadic lessons with a limited amount of learning experiences. Importantly, teachers need to be well prepared for delivering the lessons. Otherwise—with the limited understanding of each topic and low awareness of the variety of activities and material—the expected influence on bystander behaviors and group norms may not be reached through lessons.

6.2 Maintaining Fidelity: The Process and Teacher Experiences Examined

What can be measured, can be supported and improved. In Study II, three groups of teachers with distinct types of implementation profiles were identified (high, moderate and surrenders). Among the high and moderate groups (80% of the teachers), there was commitment and consistency over a period of nine months. However, 20% of the teachers (surrenders) displayed less-than-adequate fidelity to a half of the lessons. The three groups based on teachers' curriculum adherence are in line with recent research on teacher differences in their fidelity to a HIV prevention program (Wang et al., 2015). Based on our findings, it seems that the support strategies for ensuring high fidelity need to address both individual factors such as competence and motivation to implement (e.g., Wang et al., 2015), as well as intrapersonal factors such as principal support for allocating resources on antibullying curriculum delivery (Ahtola et al., 2013).

In regard to preparatory training included in the program, it probably enhances the ability to implement the program according to its' goals, but sustained implementation is not necessarily premised on the training. From the same study, neither students' engagement nor classroom management skills predicted sustained implementation. These findings may be due to measurement limitations. Previous classroom observations have shown a positive association between engagement and teacher actions (Goncy et al., 2015; Pettigrew et al., 2013) Also, in teacher feedback (unpublished study) it was quite common for teachers to describe the responsiveness of their students. This implies a transaction between student engagement and teacher actions against bullying, at least for some teachers.

Regarding a lack of sustained input, it is possible that the group of surrenders consisted of less motivated teachers and/or those who had a poor understanding of the group perspective to bullying than those teachers who were consistent (high and moderate groups). Unfortunately, there was no data to validate these speculations regarding personal norms/knowledge. As teachers' written feedback revealed, surrenders did not report more implementation difficulties (i.e., time constraints) than two other groups. The initial beliefs for program effectiveness supported lesson delivery at the early stage of implementation. However, such beliefs might not be enough. As the findings displayed, active and consistent implementation was predict by the support from principal (high and moderate groups) as well as preparing lessons well (high). Since teachers are likely to encounter obstacles of various kinds, personal involvement in lesson planning will make a difference in their efficacy to implement. Confidence and skills will surely improve if teachers have a "learning by doing" attitude that can be supported by collegial interaction regarding antibullying work.

As proposed in the model of Concern-Based Adoption by Hall and his team (1975) teachers express different concerns and feelings that point to the current stage of their efficacy and involvement. Such concerns regarding task (i.e., organizing and restructuring the curriculum) and impact (i.e., whether students or teachers will benefit from the program) were also more or less present in teachers' open-ended comments regarding the KiVa program (unpublished). The previously identified challenges such as finding time for the KiVa lessons within curriculum (Study II and unpublished) are not necessarily going to be a problem in the future if remedial actions are taken.

Eventually all innovations, also the effective ones such as KiVa, need to be updated. As attention shifts from adoption to sustainable implementation, program instrumentality should be prioritized. The program has components that have high importance but also high degree of freedom to choose from a list of activities. Because time (especially in the spring) seems to be a critical factor in implementation, it may be essential to slightly narrow the curriculum and evaluate the proportion of components in each lesson as well the curriculum arrangement. A logical starting point for updating the parts of program would be to use research to guide which components can be improved (or even omitted) and how the modifications can be expected to operate. For instance, regarding implementation strategies teachers may be most useful to provide factual information but peer-led teaching and role modeling can be most useful for providing situation-based and personalized information.

6.3 Recognition of Victimized Students: The Challenge Examined

In preventing and intervening in bullying, including recognizing all sorts of bullying acts, various levels of school system need to be addressed (Espelage & Swearer, 2003; Rodkin & Hodges, 2003). When school staff is committed to bullying prevention, students can experience less peer victimization, aggression

and bully perpetration (Espelage et al., 2014). If prevention at some level fails and bullying takes place, intervening strategies are needed in order to reduce the numbers of victims as well as the negative effects of bullying on all participants: victims, bullies (e.g., Gini, 2008; Sainio et al., 2013), and even students witnessing it (e.g., Nishina & Juvonen, 2005).

When teachers are aware of bullying, they are likely to intervene (e.g., Fekkes et al., 2005; Novick & Isaacs, 2010). However, there are only low to moderate correlations between informants of peer victimization (self-, peer-, or teacher-reporting) (Bouman et al., 2013; Graham & Juvonen, 1998; Ladd & Kochenderfer-Ladd, 2002; Wienke Totura et al., 2009). Unfortunately, our findings showed that adult identification of victims was low, and it was not higher in the case of students who were harassed directly (physically and verbally) than when students were targeted by relational aggression.

It has been noted that students do not necessarily perceive teacher responses to bullying as effective (Fekkes et al., 2005). As illustrated in the recent School Health Promoting Study (2013; with over 99 000 respondents) almost 70% of Finnish youth perceive that school personnel have not intervened. These findings display that a lot needs to be done in order to enhance students' trust on adult intervention. The basic point is to understand that structured and well-implemented antibullying principles in schools can be expected to give a clear and strict signal that adults do not tolerate bullying of any kind. Importantly, when KiVa is used well (including both universal and indicated actions), it should make discussing the bullying problem easier so that victims and anyone witnessing bullying can no longer stay silent.

6.4 Strengths

The longitudinal data with a large number of teachers and students offered valuable opportunities to evaluate multiple aspects of program fidelity and its' effects on program outcomes (Study I). Second, in Study II a person-centered approach was used in order to shed a light to individual differences regarding adherence to bullying prevention. Surprisingly, previous bullying research has paid little attention to teachers' perspective and their program adherence over time in spite of the fact that they are the key agents in preventing bullying and intervening in it. Third, I utilized teachers' experiences (unpublished study) on program use in order to get a holistic view on implementation process. Fourth, the nested structure of student and teacher data was taken into account in the analyses: In Studies I and III by two-level analyses and in Study II by correcting the standard errors for clustering. Fifth, the use of latent variables for victimization and bullying across time (Study I) provided a clear advantage over observed variables (including measurement error) that are commonly used in bullying research. Finally, the findings have many practical implications for both evaluation and implementation of antibullying programs.

6.5 Limitations and Directions for Future

Although teacher-reports measures were collected systemically across the implementation period (instead of collecting information after the trial), teachers might have rated their implementation degree somewhat higher than it actually was. Thus, teachers may not have a set of standards for evaluating whether lesson activities were correctly and clearly delivered. For instance, teachers might have started with a new lesson without carefully wrapping up things learnt last month, or delivered a set of lessons more frequently than they were instructed. For improving accuracy of self-reporting, a more detailed

format (i.e., including options to rate possible changes teachers make) should be considered. Thus, supplementary observations or interviews (if possible) on instructional quality would bring a richer knowledge on how implementation fidelity is achieved and maintained.

Another critical issue is limited attention to students' attendance to lessons. Consequently, I was not able to evaluate program impacts at the individual level (Study I). It is also possible that lesson attendance and student perceptions may moderate the link between higher adherence/better quality and program outcomes such as reduced victimization (and many others, including increased pro-social behavior in classroom). In fact, recent longitudinal findings by Saarento et al. (2014) showed that students in KiVa schools, in comparison to students in control schools, evaluated their teachers becoming more disapproving of bullying over time, and this collective perception of teacher attitudes was a significant predictor for reductions in in the perpetration of bullying both at the student as well as the classroom-level. It may be that students' perceptions of teacher commitment against bullying are even stronger in the classrooms where a teacher displays greater fidelity to curriculum implementation. More research is needed on the relationships between teacher-level variables and characteristics of individual students/classrooms. Also, students' perceptions about the program content (i.e., whether they like the program, how useful they consider program topics personally) have been very little (if ever) used when the effects of antibullying programs have been evaluated.

A process evaluation addressing teachers' commitment (or lack of it) to bullying prevention (in this case, a systematic delivery of curriculum) was limited to the individual perspective of teachers (Study II). Implementing a school program as extensive as KiVa requires a multilayered approach, which is why school-level factors (e.g., staff commitment to prevent bullying) are

important considerations in future studies. We found that 11 to 25% of the variance regarding the predictors of teacher commitment was accounted by the school context (Study II). In the examination of school-level antibullying actions, Kallestad and Olweus (2003) found that open communication between teachers about teacher-student relationships and school attention to bullying were significant predictors of implementation of Olweus Bullying Prevention Program.

Delivering the KiVa curriculum was examined only in elementary grades. As shown in a recent meta-analysis of 19 antibullying programs (Yeager, Fong, Lee, & Espelage, 2015), including the studies of KiVa for Grades 1 to 9 (Kärnä et al., 2011b, 2011a, 2013), the success rates of the programs for multiple age groups seem to vary between elementary and middle schools. In general, program effects are much smaller, or without any statistically significant declines in bullying among middle school students (Yeager et al., 2015). Also, an evaluation of KiVa during its nationwide implementation (Kärnä et al., 2011a) showed a positive correlation between number of lessons delivered and outcomes obtained in all grade levels of elementary schools, but only in one grade level in middle schools. While, it has been proposed several explanations for the weaker effects such as developmental stage and ineffective program methods (Yeager et al., 2015), the degree of program fidelity over time can, to some extent, explain variation in the intervention results.

I have examined teachers' actions in bullying prevention in the context of their first time implementation of the KiVa antibullying program during the effectiveness trial. With respect to KiVa, its' widespread adoption in Finnish schools started in the fall of 2009. During the first and second years of the national rollout lesson fidelity reduced from the fidelity rates obtained in the RCT (Salmivalli, Poskiparta, Ahtola, & Haataja, 2013). Over the years, schools

(90% of all Finnish schools are registered as KiVa users) display a great deal of variation in fidelity to antibullying curriculum in both elementary and middle schools. Overall, it seems that the degree of implementation fidelity becomes lower the longer KiVa is implemented: The first half of the full curriculum is delivered while the latter half may even be left out (Sainio, 2014). Indeed, focusing on both implementation fidelity and school/classroom environment will be critical for understanding why a) the magnitude of effects can range between subgroups, and b) whether the effects are achieved/maintained over time.

Given that the tendency of schools is to adhere less KiVa curriculum over time, this can be critical in the future—not only for preventing and reducing bullying effectively, but also for achieving (more) safety in the school environment. Naturally, schools and teachers, in good will, may have created their own materials, or shifted from theory-driven and research-based methods to others. However, the use of non-evidence-based practices is somewhat deceptive (see, Ennett et al., 2011). Sustained inputs of KiVa are necessary if communities, policy makers and all practitioners who work in schools want to rely on the best available evidence against bullying.

Importantly, the implementation of the KiVa antibullying program benefits to a wider group of students (not only the ones being harassed by their peers) as the positive side effects can be observed in school liking and academic motivation (Salmivalli, Garandeau, & Veenstra, 2012) as well as in reducing internalizing symptoms among students (Williford et al., 2012). Thus, the KiVa program fits well with the underlying values of the Finnish education system supporting equality and human rights. The reform of core curriculum for basic education in 2016–2017 takes steps to improve learning environments that encourage interaction, cooperation, and joint responsibility—elements that also characterize the KiVa curriculum. The challenge of implementation remains,

but if schools can deal with the time pressure, work out a system to communicate and collaborate on a regular basis, and to see the value of evidence-based programs, they are already moving in the right direction for improving well-being for all children.

7. REFERENCE

- Ahtola, A., Haataja, A., Kärnä, A., Poskiparta, E., & Salmivalli, C. (2013). Implementation of anti-bullying lessons in primary classrooms: How important is head teacher support? *Educational Research*, *55*, 376–392. <http://doi.org/10.1080/00131881.2013.844941>
- Asparouhov, T., & Muthen, B. O. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397–438. <http://doi.org/10.1080/10705510903008204>
- Bergsmann, E. M., Van De Schoot, R., Schober, B., Finsterwald, M., & Spiel, C. (2013). The effect of classroom structure on verbal and physical aggression among peers: A short-term longitudinal study. *Journal of School Psychology*, *51*, 159–174. <http://doi.org/10.1016/j.jsp.2012.10.003>
- Boulton, M. J. (1997). Teachers' views on bullying: Definitions, attitudes and ability to cope. *British Journal of Educational Psychology*, *67*, 223–233. <http://doi.org/10.1111/j.2044-8279.1997.tb01239.x>
- Bouman, T., van der Meulen, M., Goossens, F. A., Olthof, T., Vermande, M. M., & Aleva, E. A. (2013). Peer and self-reports of victimization and bullying: Their differential association with internalizing problems and social adjustment. *Journal of School Psychology*, *51*, 759–774. <http://doi.org/10.1016/j.jsp.2013.04.001>
- Bradshaw, C. P., Sawyer, A. L., & O'Brennan, L. M. (2007). Bullying and peer victimization at school: Perceptual differences between students and school staff. *School Psychology Review*, *36*, 361–382.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. <http://doi.org/10.1191/1478088706qp063oa>
- Craig, W. M., Henderson, K., & Murphy, J. G. (2000). Prospective teachers' attitudes toward bullying and victimization. *School Psychology International*, *21*, 5–21. <http://doi.org/10.1177/0143034300211001>
- Cross, D., Hall, M., Hamilton, G., Pintabona, Y., & Erceg, E. (2004). Australia: The Friendly Schools project. In P. K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools. How successful can interventions be?* (pp. 187–210). Cambridge: Cambridge University Press.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*, 23–45. [http://doi.org/10.1016/S0272-7358\(97\)00043-3](http://doi.org/10.1016/S0272-7358(97)00043-3)
- 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. *American Journal of Community Psychology*, *41*, 327–350. <http://doi.org/10.1007/s10464-008-9165-0>
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*, 405–432. <http://doi.org/10.1111/j.1467-8624.2010.01564.x>
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, *18*, 237–256. <http://doi.org/10.1093/her/18.2.237>
- Elliott, D., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science*, *5*, 47–53. <http://doi.org/10.1023/B:PREV.0000013981.28071.52>
- Ennett, S. T., Haws, S. W., Ringwalt, C. L., Vincus, A. A., Hanley, S., Bowling, J. M., & Rohrbach, L. A. (2011). Evidence-based practice in school substance use prevention: Fidelity of implementation under real-world conditions. *Health Education Research*, *26*, 361–371. <http://doi.org/10.1093/her/cyr013>
- Espelage, D. L., Polanin, J. R., & Low, S. K. (2014). Teacher and staff perceptions of school environment as predictors of student aggression,

- victimization, and willingness to intervene in bullying situations. *School Psychology Quarterly*, *29*, 287–305.
- Espelage, D. L., & Swearer, S. M. (2003). Research on school bullying and victimization: What have we learned and where do we go from here? *School Psychology Review*, *32*(3), 365–383.
- Fekkes, M., Pijpers, F. I. M., & Verloove-Vanhorick, S. P. (2005). Bullying: Who does what, when and where? Involvement of children, teachers and parents in bullying behavior. *Health Education Research*, *20*, 81–91. <http://doi.org/10.1093/her/cyg100>
- Forgatch, M. S., Patterson, G. R., & DeGarmo, D. S. (2005). Evaluating fidelity: Predictive validity for a measure of competent adherence to the oregon model of parent management training. *Behavior Therapy*, *36*(1), 3–13. [http://doi.org/10.1016/S0005-7894\(05\)80049-8](http://doi.org/10.1016/S0005-7894(05)80049-8)
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, *17*(3), 1–13.
- Garandau, C. F., Lee, I. A., & Salmivalli, C. (2014). Differential effects of the KiVa anti-bullying program on popular and unpopular bullies. *Journal of Applied Developmental Psychology*, *35*, 44–50. <http://doi.org/10.1016/j.appdev.2013.10.004>
- Garandau, C. F., Poskiparta, E., & Salmivalli, C. (2014). Tackling acute cases of school bullying in the KiVa Anti-Bullying Program: A comparison of two approaches. *Journal of Abnormal Child Psychology*, *42*, 981–991. <http://doi.org/10.1007/s10802-014-9861-1>
- Gini, G. (2008). Associations between bullying behaviour, psychosomatic complaints, emotional and behavioural problems. *Journal of Paediatrics and Child Health*, *44*, 492–497. <http://doi.org/10.1111/j.1440-1754.2007.01155.x>
- Goncy, E. A., Sutherland, K. S., Farrell, A. D., Sullivan, T. N., & Doyle, S. T. (2015). Measuring teacher implementation in delivery of a bullying prevention program: the impact of instructional and procedural adherence and competence on student responsiveness. *Prevention Science*, *16*, 440–450. <http://doi.org/10.1007/s11121-014-0508-9>
- Graham, S., & Juvonen, J. (1998). Self-blame and peer victimization in middle school: An attributional analysis. *Developmental Psychology*, *34*, 587–599. <http://doi.org/10.1037//0012-1649.34.3.587>
- Hahn, E. J., Noland, M. P., Rayens, M. K., & Christie, D. M. (2002). Efficacy of training and fidelity of implementation of the Life Skills Training program. *Journal of School Health*, *72*, 282–287. <http://doi.org/10.1111/j.1746-1561.2002.tb01333.x>
- Hall, G. E. (2013). Evaluating change processes: Assessing extent of implementation (constructs, methods and implications). *Journal of Educational Administration*, *51*, 264–289. <http://doi.org/10.1108/09578231311311474>
- Hall, G. E., Loucks, S. F., Rutherford, W. L., & Newlove, B. W. (1975). Levels of use of the innovation: A framework for analyzing innovation adoption. *Journal of Teacher Education*, *26*, 52–56. <http://doi.org/10.1177/002248717502600114>
- Hansen, W. B., Pankratz, M. M., & Bishop, D. C. (2014). Differences in Observers' and Teachers' Fidelity Assessments. *The Journal of Primary Prevention*, *35*(5), 297–308. <http://doi.org/10.1007/s10935-014-0351-6>
- Hazler, R. J., Miller, D. L., Carney, J. V., & Green, S. (2001). Adult recognition of school bullying situations. *Educational Research*, *43*, 133–146. <http://doi.org/10.1080/00131880110051137>
- Hirschstein, M., Edstrom, L., Frey, K., Snell, J., & MacKenzie, E. (2007). Walking the talk in bullying prevention: Teacher implementation variables related to initial impact of the Steps to Respect program. *School Psychology Review*, *36*, 3–21.
- Jones, S. M., Brown, J. L., & Lawrence Aber, J. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development*, *82*, 533–554. <http://doi.org/10.1111/j.1467-8624.2010.01560.x>
- Kallestad, J. H., & Olweus, D. (2003). Predicting teachers' and schools' implementation of the Olweus bullying intervention program: A

- multilevel study. *Prevention & Treatment*, 6, 3–21. <http://doi.org/10.1037/1522-3736.6.1.621a>
- Kam, C.-M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. Promoting Alternative Thinking Skills Curriculum. *Prevention Science*, 4, 55–63.
- Kärnä, A., Voeten, M., Little, T. D., Alanen, E., Poskiparta, E., & Salmivalli, C. (2011a). Going to Scale: A nonrandomized nationwide trial of the KiVa antibullying program for grades 1-9. *Journal of Consulting and Clinical Psychology*, 79, 796–805. <http://doi.org/10.1037/a0025740>
- Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Alanen, E., & Salmivalli, C. (2013). Effectiveness of the KiVa anti-bullying program: Grades 1-3 and 7-9. *Journal of Educational Psychology*, 105, 535–551. <http://doi.org/10.1037/a0030417>
- Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011b). A large-scale evaluation of the KiVa antibullying program: Grades 4-6. *Child Development*, 82, 311–330. <http://doi.org/10.1111/j.1467-8624.2010.01557.x>
- Ladd, G. W., & Kochenderfer-Ladd, B. (2002). Identifying victims of peer aggression from early to middle childhood: Analysis of cross-informant data for concordance, estimation of relational adjustment, prevalence of victimization, and characteristics of identified victims. *Psychological Assessment*, 14, 74–96. <http://doi.org/10.1037/1040-3590.14.1.74>
- Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology*, 78, 75–79. <http://doi.org/10.1037/0022-0663.78.2.75>
- Lillehoj, C., Griffin, K., & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education & Behavior*, 31, 242–257. <http://doi.org/10.1177/1090198103260514>
- Low, S., Ryzin, M. J. V., Brown, E. C., Smith, B. H., & Haggerty, K. P. (2014). Engagement matters: Lessons from assessing classroom implementation of Steps to Respect: A bullying prevention program over a one-year period. *Prevention Science*, 15, 165–176. <http://doi.org/10.1007/s11121-012-0359-1>
- Lubke, G. H., & Muthen, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods March 2005*, 10, 21–39.
- Lubke, G. H., & Neale, M. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models. *Multivariate Behavioral Research*, 43, 592–620. <http://doi.org/10.1080/00273170802490673>
- Luckner, A. E., & Pianta, R. C. (2011). Teacher–student interactions in fifth grade classrooms: Relations with children’s peer behavior. *Journal of Applied Developmental Psychology*, 32, 257–266. <http://doi.org/10.1016/j.appdev.2011.02.010>
- McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analyses. In R. Cudeck, S. Du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 342–380). Lincolnwood, IL: Scientific Software International, Inc.
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. In *Annual Review of Psychology* (Vol. 60, pp. 577–605). Palo Alto: Annual Reviews.
- McArdle, J. J., & Nesselroade, J. R. (1994). Structuring data to study development and change. In S. H. Cohen & H. . Reese (Eds.), *Life-span developmental psychology* (pp. 223–267). Hillsdale, NJ: Erlbaum.
- Melde, C., Esbensen, F.-A., & Tusinski, K. (2006). Addressing program fidelity using onsite observations and program provider descriptions of program delivery. *Evaluation Review*, 30, 714–740. <http://doi.org/10.1177/0193841X06293412>
- Mihalic, S., Fagan, A. A., & Argamaso, S. (2008). Implementing the LifeSkills Training drug prevention program: Factors related to implementation fidelity. *Implementation Science*, 3:5. <http://doi.org/10.1186/1748-5908-3-5>
- Mishna, F., Scarcello, I., Pepler, D., & Wiener, J. (2005). Teachers’ understanding of bullying. *Canadian Journal of Education / Revue Canadienne de L’éducation*, 28, 718–738. <http://doi.org/10.2307/4126452>

- Molcho, M., Craig, W., Due, P., Pickett, W., Harel-Fisch, Y., Overpeck, M., & Group, the H. B. W. (2009). Cross-national time trends in bullying behaviour 1994–2006: findings from Europe and North America. *International Journal of Public Health, 54*(2), 225–234. <http://doi.org/10.1007/s00038-009-5414-8>
- Muthén, B. O., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. J. Hox & J. . Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 15–40). New York, NY: Taylor & Francis.
- Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén.
- Nishina, A., & Juvonen, J. (2005). Daily Reports of Witnessing and Experiencing Peer Harassment in Middle School. *Child Development, 76*(2), 435–450. <http://doi.org/10.1111/j.1467-8624.2005.00855.x>
- Novick, R. M., & Isaacs, J. (2010). Telling is compelling: The impact of student reports of bullying on teacher intervention. *Educational Psychology, 30*, 283–296. <http://doi.org/10.1080/01443410903573123>
- Olweus, D. (1986). *The Olweus Bully/Victim Questionnaire*. Mimeo. Bergen, Norway: University of Bergen.
- Olweus, D. (1993). *Bullying at school: What we know and what we can do*. Wiley-Blackwell.
- Olweus, D. (1996). *The Revised Olweus Bully/Victim Questionnaire*. Mimeo. Bergen, Norway: Research Center for Health Promotion (HEMIL Center), University of Bergen.
- Olweus, D., & Kallestad, J. H. (2010). The Olweus Bullying Prevention Program: Effects of classroom components at different grade levels. In K. Österman (Ed.), *Indirect and direct aggression* (pp. 115–131). New York: Peter Lang Publisher.
- Pettigrew, J., Miller-Day, M., Shin, Y., Hecht, M. L., Krieger, J. L., & Graham, J. W. (2013). Describing teacher–student interactions: A qualitative assessment of teacher implementation of the 7th grade keepin' it REAL substance use intervention. *American Journal of Community Psychology, 51*, 43–56. <http://doi.org/10.1007/s10464-012-9539-1>
- Resnicow, K., Davis, M., Smith, M., Lazarus-Yaroch, A., Baranowski, T., Baranowski, J., ... Wang, D. (1998). How best to measure implementation of school health curricula: A comparison of three measures. *Health Education Research, 13*, 239–250. <http://doi.org/10.1093/her/13.2.239>
- Rodkin, P. C., Farmer, T. W., Pearl, R., & Van Acker, R. (2006). They're cool: Social status and peer group supports for aggressive boys and girls. *Social Development, 15*, 175–204. <http://doi.org/10.1111/j.1467-9507.2006.00336.x>
- Rodkin, P. C., & Hodges, E. V. E. (2003). Bullies and victims in the peer ecology: Four questions for psychological and school professionals. *School Psychology Review, 32*, 384–400.
- Ryan, W., & Smith, J. D. (2009). Antibullying programs in schools: How effective are evaluation practices? *Prevention Science, 10*, 248–259. <http://doi.org/10.1007/s11121-009-0128-y>
- Saarento, S., Boulton, A. J., & Salmivalli, C. (2014). Reducing bullying and victimization: Student- and classroom-level mechanisms of change. *Journal of Abnormal Child Psychology, 43*, 61–76. <http://doi.org/10.1007/s10802-013-9841-x>
- Sainio, M. (2014). KiVa Koulu® -ohjelma suomalaisissa peruskouluissa: Miten ohjelmaa toteutetaan tai miksi sitä ei käytetä? Retrieved from <http://www.kivakoulu.fi/assets/files/KiVa-tutkimusraportti.pdf>
- Sainio, M., Veenstra, R., Little, T. D., Karna, A., Ronkko, M., & Salmivalli, C. (2013). Being bullied by same-versus other-sex peers: Does it matter for adolescent victims? *Journal of Clinical Child and Adolescent Psychology, 42*, 454–466. <http://doi.org/10.1080/15374416.2013.769172>
- Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior, 15*(2), 112–120. <http://doi.org/10.1016/j.avb.2009.08.007>
- Salmivalli, C., Garandeau, C. F., & Veenstra, R. (2012). KiVa anti-bullying program: Implications for school adjustment. In G. W. Ladd & A. M. Ryan (Eds.), *Peer Relationships and Adjustment at School* (pp. 279–303). NC: Information Age Publishing.

- Salmivalli, C., Poskiparta, E., Ahtola, A., & Haataja, A. (2013). The implementation and effectiveness of the KiVa antibullying program in Finland. *European Psychologist, 18*, 79–88. <http://doi.org/10.1027/1016-9040/a000140>
- Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development, 6*, 144–164. <http://doi.org/10.1080/15427600902911247>
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.
- Smith, P. K., & Brain, P. (2000). Bullying in schools: Lessons from two decades of research. *Aggressive Behavior, 26*, 1–9. [http://doi.org/10.1002/\(SICI\)1098-2337\(2000\)26:1<::AID-AB1>3.0.CO;2-7](http://doi.org/10.1002/(SICI)1098-2337(2000)26:1<::AID-AB1>3.0.CO;2-7)
- Smith, P. K., Smith, C., Osborn, R., & Samara, M. (2008). A content analysis of school anti-bullying policies: Progress and limitations. *Educational Psychology in Practice, 24*(1), 1–12. <http://doi.org/10.1080/02667360701661165>
- Snell, M. E., Berlin, R. A., Voorhees, M. D., Stanton-Chapman, T. L., & Hadden, S. (2012). A survey of preschool staff concerning problem behavior and its prevention in Head Start classrooms. *Journal of Positive Behavior Interventions, 14*, 98–107. <http://doi.org/10.1177/1098300711416818>
- Tobler, N. S., & Stratton, H., H. (1997). Effectiveness of school-based drug prevention programs: A meta-analysis of the research. *The Journal of Primary Prevention, 18*, 71–128. <http://doi.org/10.1023/A:1024630205999>
- Ttofi, M. M., & Farrington, D. P. (2010). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of Experimental Criminology, 7*, 27–56. <http://doi.org/10.1007/s11292-010-9109-1>
- Vaillancourt, T., & Hymel, S. (2006). Aggression and social status: The moderating roles of sex and peer-valued characteristics. *Aggressive Behavior, 32*, 396–408. <http://doi.org/10.1002/ab.20138>
- Vreeman, R., & Carroll, A. (2007). A systematic review of school-based interventions to prevent bullying. *Archives of Pediatrics & Adolescent Medicine, 161*, 78–88. <http://doi.org/10.1001/archpedi.161.1.78>
- Wang, B., Stanton, B., Deveaux, L., Poirier, M., Lunn, S., Koci, V., ... Rolle, G. (2015). Factors influencing implementation dose and fidelity thereof and related student outcomes of an evidence-based national HIV prevention program. *Implementation Science, 10*, 44. <http://doi.org/10.1186/s13012-015-0236-y>
- WHO. (1993). *The European Network of Health Promoting Schools, Resource Manual*. World Health Organization.
- Wienke Totura, C. M., Green, A. E., Karver, M. S., & Gesten, E. L. (2009). Multiple informants in the assessment of psychological, behavioral, and academic correlates of bullying and victimization in middle school. *Journal of Adolescence, 32*, 193–211. <http://doi.org/10.1016/j.adolescence.2008.04.005>
- Williford, A., Boulton, A., Noland, B., Little, T. D., Kärnä, A., & Salmivalli, C. (2012). Effects of the KiVa anti-bullying program on adolescents' depression, anxiety, and perception of peers. *Journal of Abnormal Child Psychology, 40*, 289–300. <http://doi.org/10.1007/s10802-011-9551-1>
- Yeager, D. S., Fong, C. J., Lee, H. Y., & Espelage, D. L. (2015). Declines in efficacy of anti-bullying programs among older adolescents: Theory and a three-level meta-analysis. *Journal of Applied Developmental Psychology, 37*, 36–51. <http://doi.org/10.1016/j.appdev.2014.11.005>