

University of Groningen

Bridging the implementation gap

Goense, Pauline Brigitta

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:
2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Goense, P. B. (2016). *Bridging the implementation gap: A study on sustainable implementation of interventions in child and youth care organizations*. [Thesis fully internal (DIV), University of Groningen]. Rijksuniversiteit Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Bridging the implementation gap

A study on sustainable implementation of interventions
in child and youth care organizations

Pauline Brigitta Goense



Bridging the implementation gap

A study on sustainable implementation of interventions
in child and youth care organizations

Pauline Brigitta Goense

The research for this dissertation was conducted as part of the program of the Implementation and Effectiveness of Youth Care Services research group at the Faculty of Applied Social Sciences and Law, Amsterdam University of Applied Sciences. The studies presented in Chapters 4 and 5 were financially supported by a grant from Stichting Innovatie Alliantie – a fund for research conducted by Dutch universities of applied sciences (2012-14-18P).

Publication of this dissertation was financially supported by Groningen University.

© Pauline Brigitta Goense, Amsterdam 2016

ISBN: 978-90-367-8984-4

Printing: Ipskamp Printing

Cover design: Goedsnik Graphic Design, www.goedsnik.com

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the copyright holder.



rijksuniversiteit
 groningen



Hogeschool van Amsterdam



rijksuniversiteit
groningen

Bridging the implementation gap

A study on sustainable implementation of interventions
in child and youth care organizations

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 22 september 2016 om 12.45 uur

door

Pauline Brigitta Goense

geboren op 2 april 1986
te 's-Gravenhage

Promotor

Prof. dr. T.A. van Yperen

Copromotor

Dr. L. Boendermaker

Beoordelingscommissie

Prof. dr. E.J. Knorth

Prof. dr. S.A. Reijneveld

Prof. dr. G.J. Overbeek

Table of contents

Chapter 1	General introduction	7
Chapter 2	Implementation of Treatment Integrity Procedures: An Analysis of Outcome Studies of Youth Interventions Targeting Externalizing Behavioral Problems	15
Chapter 3	Making ‘What Works’ Work: A meta-analytic study of the effect of treatment integrity on outcomes of evidence- based interventions for juveniles with antisocial behavior	33
Chapter 4	Support systems for treatment integrity	57
Chapter 5	Measuring Treatment Integrity: Use of and experience with measurements in child and youth care organizations	71
Chapter 6	Stimulating quality of service delivery in mental health care settings for youth. Suggestions for integrating support systems	101
Chapter 7	General discussion	109
	References	125
	Nederlandse samenvatting	141
	Dankwoord	151
	About the author	155
	List of publications	157

CHAPTER 1

General introduction

1.1 Introduction

Every year, a substantial number of children, young people, and their families come into contact with child and youth care organizations. The vast majority of this group is the population of children and young people with externalizing behavior problems, along with their families (Parhar, Wormith, Derkzen, & Beauregard, 2008). Multiple psychosocial interventions targeting this population have been developed over the years. Some of these interventions have demonstrated efficacy for youths with these types of problems, including oppositional, aggressive, and/or delinquent behavior (for a review, see: Brestan & Eyberg, 1998; Burns, Hoagwood, & Mrazek, 1995; Carr, 2000; Kazdin, & Weisz, 1998; Ollendick & King, 2000). Unfortunately, not all interventions have proven to be effective, and some have even led to adverse outcomes (Dishion, McCord, & Poulin, 1999; Farrington & Welsh, 2006). Leaving children and young people with externalizing behavior problems underserved or unserved may have serious negative consequences for both these youngsters and society. Some of these children, young people, and their families are confronted with out-of-home placements and imprisonment, and the larger society is confronted with substantial monetary costs, specifically when the behavior of these youngsters turns into chronic delinquent behavior (Cohen, Piquero, & Jennings, 2010).

There is pressure, and organizations have the responsibility, to provide high-quality services while making efficient and effective use of limited financial resources. Child and youth care organizations and their financers in the Netherlands and elsewhere, increasingly emphasize the effectiveness of interventions for this population in order to maximize therapeutic gains and reduce the nation's youth mental health costs (Boendermaker, 2011; Southam-Gerow & Prinstein, 2014).

The evidence concerning psychosocial interventions for children and young people with externalizing behavior problems has amassed at an impressive pace in recent years (Southam-Gerow & Prinstein, 2014). Interventions that have been proven effective are now considered the vehicles through which the knowledge of “what works” can be applied in practice. Outcomes for children, young people, and their families, however, have not improved in line with these advances in knowledge.¹ This deficit is known as the “implementation gap,” that is, the difference between the knowledge of “what works” and the application of this knowledge in real-life practice (Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005).

¹ A meta-analysis of RCTs that had tested youth evidence-based interventions in more clinically representative contexts, pitting them against usual care, showed a mean effect size of 0.29 for these interventions (Weisz et al., 2013a).

The implementation gap raises the following questions about the application of interventions:

1. What does it mean to apply interventions – as vehicles of the knowledge of “what works” – and how is this operationalized in outcome studies?
2. Does the application of these interventions make a real difference to the end-users of the services?
3. What types of support for professionals can strengthen implementation processes?

The aim of this dissertation is to answer these three questions and to present knowledge of factors that contribute to bridging the implementation gap. In answering these questions, the focus is on whether professionals are delivering the interventions as intended. In doing so, I hope to contribute to making “what works” work for children and young people with externalizing behavior problems and their families in the Netherlands and abroad.

1.1.1 Interventions

Child and youth care organizations in the Netherlands typically provide a wide range of advice, guidance, help, and care to an equally wide range of target groups. One broadly applied approach in the sector is the solution-oriented approach. Central to this approach is stimulating and activating clients to work on solutions themselves. In addition, many professionals also deliver specific training programs. For example, they may provide clients with a training in parenting skills, such as the Positive Parenting Program (Sanders, 2008), or provide young people with a training to strengthen their social, emotional, and cognitive skills, such as Tools4U (Albrecht & Spanjaard, 2007). These training programs differ from approaches such as the solution-oriented approach because they target the specific problem behavior of a specific target group, whereas the broadly applied approaches are directed at all clients, regardless of their problem behavior. Elements of an approach are often also used in specific training programs. Signs of Safety, for example, is a program with a solution-oriented approach for families where the safety of the child is a problem (Turnell & Edwards, 1999).

All training programs and approaches that focus on the specific behavior of a specific target group fall within the definition of an “intervention” as used in this dissertation. More specifically, the term “intervention” in this dissertation refers to programs, projects, training methods, courses, treatment and counseling forms, and sanctions that target the reduction or compensation of a risk or problem in the development of a child or young person, or are aimed at making the risk or problem more bearable. An intervention is guided by a theoretical and practically applicable, goal-centered, and systematic approach, aimed at the child or young person, his or her caretaker(s), and/or his or her educational environment. The

length of an intervention and the frequency of client contact is defined in the intervention manual (van Yperen, 2010).

Although there is a growing consensus in the youth care field that interventions should be evidence-based, the exact definition of evidence-based interventions is a contentious matter (De Swart et al., 2012; Weisz, Jensen-Doss, & Hawley, 2006). Definitions range from interventions that receive qualitative, theoretical, and/or clinical support, to interventions that have clear empirical support provided by at least two randomized controlled trials (Veerman & van Yperen, 2007). In this dissertation, evidence-based interventions refer to interventions that, minimally, are theoretically based, well-documented,² protocolled, structured, and manualized, and have gained empirical support in experimental or quasi-experimental research (Weisz, Jensen-Doss, & Hawley, 2006). For these interventions there are indications for their effectiveness and they have the potential to be disseminated. These interventions can be seen as the vehicles through which the knowledge of “what works” for a target population with a specific problem can be applied in practice.

For child and youth care organizations, evidence-based interventions seem to be the way to provide justifiable, effective care (Southam-Gerow & Prinstein, 2014; Weisburd, 2003). One of the main difficulties with evidence-based interventions is the disappointing treatment outcomes outside the research setting. According to Gendreau, Goggin, and Smith (1999), even the most state-of-the-art intervention will not produce the desired outcomes in actual practice settings, unless the organization pays attention to the process of implementing the intervention.

1.1.2 Implementation of interventions

For the purpose of this dissertation, “implementation” is defined as a set of planned, intentional activities³ that are performed to put into practice interventions in real-world organizations. The goal of effective implementation is to benefit the end-users of services, namely children, youth, adults, families, and communities (Fixsen et al., 2005). Information about the implementation of interventions is needed to determine whether an intervention failed due to the

² “Well-documented” includes the documentation of clinical expertise and patient values with regard to the intervention, as evidence-based practice is the integration of best research evidence with clinical expertise and patient values (Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000, p 1).

³ As also pointed out by Fixsen et al. (2005), it is important to distinguish implementation-related “interventions” with community leaders, agency directors, supervisors, practitioners, policymakers, and funders, from the treatment and/or prevention programs that are commonly (and in this dissertation) defined as “interventions.” For clarification purposes, I use the term “implementation efforts” to mean efforts to incorporate a program or practice at the community, agency, or practitioner level.

failure of the intervention or components thereof, or due to its insufficient or inadequate application (Schoenwald et al., 2011). As Dobson and Cook (1980) stated decades ago, we have to avoid making a “type III” error, that is, evaluating an intervention that was described but not implemented. This means that it is necessary to discriminate between implementation outcomes (whether the intervention is implemented as intended) and effectiveness outcomes (whether the intervention is implemented as intended, and is/is not resulting in good outcomes) (Fixsen et al., 2005).

Many factors can hinder or facilitate the effective implementation of interventions. Factors related to the delivery of an intervention as intended are assumed to be important in this respect (Fixsen et al., 2005). “Delivery of interventions as intended” in its broad sense means delivery of the intervention with the intended content, duration, frequency, and scope. This is referred to as program integrity (Carroll et al., 2007). There is an increased awareness that what is delivered (the content) is in fact the intervention. In implementation literature, delivering the content of the intended intervention is referred to as treatment integrity (Perepletchikova, Treat, & Kazdin, 2007). Many authors share the opinion that treatment integrity must be measured to identify what the moderators are in the outcome effects. Measuring treatment integrity is essential in understanding what adaptations can be made to the intervention without sacrificing its effectiveness (Dane & Schneider, 1998; Durlak & DuPre, 2008; Moncher & Prinz, 1991; Perepletchikova & Kazdin, 2005; Sanetti, Gritter, & Dobey, 2011; Tennyson, 2009; Weissberg, Kumfer, & Seligman, 2003).

According to Fixen and Ogden (2014), implementation research is rapidly becoming an integral part of outcome studies of evidence-based interventions. Researchers also frequently conclude that low treatment integrity could be the cause of disappointing results (Schoenwald, Chapman, Sheidow, & Carter, 2009a; Sexton, & Turner, 2010; Tennyson, 2009). Despite this attention to treatment integrity, there has been no overview available of the operationalization of treatment integrity procedures for outcome studies of interventions that target juveniles with externalizing behavior problems. It was unclear how treatment integrity was operationalized in these studies, and unclear whether the operationalization was comprehensive enough to be able to judge delivery of the intervention content as intended. The first objective of the research underlying this dissertation was to examine the operationalization of treatment integrity procedures in this type of study. The resultant overview provides information about the adequacy of treatment integrity procedures that are implemented in primary studies. It also provides knowledge of how to interpret the association between treatment integrity and client outcomes found in these individual studies.

Previous research has produced somewhat inconsistent findings on the association between treatment integrity and client outcomes. A meta-analysis should provide insight into the overall effect of treatment integrity. Previous meta-analyses have suggested that delivering an intervention with a high level of integrity is associated with positive client outcomes (see Lipsey, 2009; Tennyson, 2009). However, these meta-analyses did not take into account the quality of treatment integrity procedures of the included studies. The validity of treatment integrity procedures probably has consequences for the interpretation of findings. The second objective of the research underlying this dissertation was to meta-analytically examine, in a multilevel model, the effect of treatment integrity on client outcomes. The focus was on evidence-based interventions for juveniles exhibiting antisocial behavior. Only studies that, to a certain level, adequately implemented treatment integrity procedures were included. This inclusion criterion enabled the possibility to draw firmer conclusions on the moderating effect of level of treatment integrity on client outcomes compared to previous meta-analyses on this subject.

1.1.3 Stimulating quality of delivery of interventions

Research suggests that providing professionals with frequent and targeted support is an effective way to establish and maintain treatment integrity (Kerby, 2006; Mikolajczak, Stals, Fleuren, Wilde & Paulussen, 2009; Schoenwald et al., 2009b). Most evidence-based interventions therefore incorporate specific demands concerning the support for the professionals who carry out the interventions. The support systems of these evidence-based interventions, however, differ from each other. There was no specific knowledge of what the content of the support system should be, or of the standard minimum rules for effective support. As Beidas and Kendall (2010) conclude in their review on the training of professionals in using evidence-based interventions, which is often referred to as evidence-based practice (EBP): “Despite the importance of EBP, we know less than preferred regarding how to best train therapists in EBP” (p. 26). The third objective of the present research was to extend the knowledge of how best to support professionals in establishing and maintaining treatment integrity in planned interventions.

Various instruments are used to measure levels of treatment integrity in outcome studies of evidence-based interventions (Schoenwald & Garland, 2013). But as Schoenwald and Garland (2013, p. 154) conclude in their review of treatment adherence measurement methods, “there is a gap that warrants bridging between adherence measurement methods devised for use primarily as independent variable checks in efficacy studies and those that can be used in diverse practice contexts.” Little is known about the feasibility of the use of treatment integrity

measurements in child and youth care organizations as part of quality assurance procedures, or as a tool to provide performance feedback to therapists. Details about the resources required for the implementation of integrity measurement methods are also rarely reported (Schoenwald & Garland, 2013). The fourth objective of this research was to ascertain whether and, if so, how treatment integrity measurements are used within child and youth care organizations. That knowledge provides information about the conditions that seem necessary to successfully implement this type of measurement in organizations.

Merely providing knowledge of the ideal conditions of support for professionals in establishing and maintaining treatment integrity in planned interventions is not sufficient to change practice. It will not bridge the gap between the ideal conditions and the actual conditions within child and youth care organizations. One of the major difficulties with the provision of support systems to professionals is that child and youth care organizations have limited time and capability to provide such systems. The last objective of this research was to devise a potential way to organize support systems for professionals that take into account these organizations' capacities and incapacities.

1.2 Structure of this dissertation

As a first step toward a better understanding of the implementation gap, it is necessary to understand how “delivery as intended” is operationalized in this context. Chapter 2 describes the systematic review of the operationalization of treatment integrity procedures in outcome studies of interventions that target juveniles with externalizing behavior problems. The moderating effect of level of treatment integrity on the reduction of youth antisocial behavior after an intervention is meta-analytically examined in a three-level model, which is described in Chapter 3.

Most interventions are provided by youth care professionals. They have an important role in the delivery of the intervention as it is intended. Chapter 4 describes the systematic review on effective support for youth care professionals in order to enable them to deliver the intended intervention with treatment integrity. Essential elements of support systems for professionals are discussed in this chapter.

Instruments that have the potential to be used to support professionals are treatment integrity instruments. These instruments provide information about the delivery of interventions as intended. Chapter 5 describes the qualitative study of the experiences and use of treatment integrity instruments of 12 interventions for children and young people with externalizing behavior problems provided in the Netherlands. The conditions under which these instruments can

be successfully implemented in child and youth care organizations are discussed in this chapter.

The question that remained was how child and youth care organizations can organize support for professionals in an effective and efficient way to secure quality of delivery. Chapter 6 presents a potential way to integrate support systems for professionals around overlapping factors of interventions. Lastly, in Chapter 7, the findings of this dissertation are summarized and limitations and practical implications are discussed. In addition, recommendations for future research and concluding remarks are made. An overview of the research questions, objectives, and corresponding chapters is presented in Table 1.1.

Table 1.1

Overview of research questions, objectives, and corresponding chapters

Research question	Objective	Corresponding chapter
1. What does it mean to apply interventions – as vehicles of the knowledge of “what works” – and how is this operationalized in outcome studies?	To examine the adequacy of the implementation of treatment integrity procedures in outcome studies of interventions targeting externalizing behavior problems of youth.	Chapter 2
2. Does the application of these interventions make a real difference to the end-users of the services?	To meta-analytically examine the moderating effect of level of treatment integrity on the reduction of youth antisocial behavior after an intervention.	Chapter 3
3. What types of support for professionals can strengthen implementation processes?	To examine the essential ingredients of support for youth care professionals to enable them to deliver the intended intervention with treatment integrity.	Chapter 4
	To examine the experiences and use of treatment integrity instruments within child and youth care organizations.	Chapter 5
	To devise a potential way to integrate support systems for professionals around overlapping factors of interventions.	Chapter 6

CHAPTER 2

Implementation of Treatment Integrity Procedures: An Analysis of Outcome Studies of Youth Interventions Targeting Externalizing Behavioral Problems

Published as:

Goense, P., Boendermaker, L., van Yperen, T., Stams, G.J., & van Laar, J. (2014). Implementation of Treatment Integrity Procedures: An Analysis of Outcome Studies of Youth Interventions Targeting Externalizing Behavioral Problems. *Zeitschrift für Psychologie*, 222(1), 12–21.

Abstract

This systematic review evaluates the implementation of treatment integrity procedures in outcome studies of youth interventions targeting behavioural problems. The Implementation of Treatment Integrity Procedures Scale (ITIPS), developed by Perepletchikova, Treat and Kazdin (2007), was adapted (ITIPS-A) and used to evaluate 32 outcome studies of evidence-based interventions for youth with externalising behaviour problems. Integrity measures were found to be still rare in these studies. Of the studies that take integrity into account, 80 percent approaches adequacy in implementing procedures for treatment integrity. The ITIPS-A is recommended as an instrument to guide development of integrity instruments and the implementation of treatment integrity procedures in youth care.

Keywords

Treatment integrity, fidelity, adherence, competence, treatment outcome, implementation

2.1 Introduction

The most complicated, expensive and burdensome research designs are used to test the causal relations between specific interventions and client outcomes. In this quest to prove effectiveness of interventions, the point of assuring that *what* is studied in an outcome study in fact is the intervention – well implemented and used with high integrity - does not get much attention. This is a serious omission, as without this assurance no conclusions may be drawn on the relation between an intervention and client outcomes. This means that for many interventions there would still be no proof of their effectiveness.

If there is one population that is exposed to interventions, it is the population of children and young people with externalising behaviour problems (Parhar, Wormith, Derkzen, & Beauregard, 2008). For years, these youngsters have undergone all sorts of interventions, which turned out to be ineffective and in some cases even led to adverse outcomes (Dishion, McCord, & Poulin, 1999; Farrington & Welsh, 2006).

Providing only ‘evidence-based’ interventions seems to be the answer to providing justifiable, effective care (Weisburd, 2003). However, if the research that led to the conclusion that an intervention is ‘evidence-based’ did not take treatment integrity into account, there is a chance that such intervention might not produce the desired effect (Perepletchikova & Kazdin, 2005).

Program integrity refers to the delivery of the intervention as it is intended, including its content, duration, frequency and scope (Carroll et al., 2007). Some outcome studies of evidence-based interventions for children and young people with externalising behaviour problems take the integrity into account. Barnoski (2004), for instance, measured treatment integrity in his outcome study on ‘Functional Family Therapy’. Treatment integrity of Functional Family Therapy was measured by asking a staff person and some consultants to recall the therapists’ competence in the delivery of the intervention. Although treatment integrity was taken into account, it is highly questionable whether this kind of measurement is valid and comprehensive enough to assure the delivery of the intervention as intended. This example raises the question how treatment integrity measures are implemented in other outcome studies.

Perepletchikova, Treat and Kazdin (2007) have examined the implementation of treatment integrity in adult and child psychotherapy outcome studies published between 2000-2004 in high impact journals. They found that only 3,5% of the 147 articles met criteria for adequate implementation of treatment integrity procedures (Perepletchikova et al., 2007). The present study is the first to review the implementation of treatment integrity in outcome studies of evidence-based interventions for youth with externalising behaviour problems.

2.1.1 Aspects of treatment integrity

Perepletchikova and colleagues (2007) selected an extensive body of studies in order to examine the adequacy of treatment integrity procedures in psychotherapy research. According to Perepletchikova and colleagues (2007), treatment integrity encompasses three aspects: 1) therapist adherence, 2) therapist competence, and 3) treatment differentiation. Therapist adherence is the degree to which the therapist utilizes prescribed procedures and avoids proscribed (or prohibited) procedures. Therapist competence refers to the level of therapist (technical) skills and the judgment in delivering the components of the treatment (Barber et al., 2006; Barber, Scharpless, Klostermann, & McCarthy, 2007a; Barber, Triffelman, & Marmar, 2007b; Perepletchikova et al., 2007). Treatment differentiation is the degree to which the treatment differs from other treatments along critical dimensions (Perepletchikova et al., 2007; Waltz, Addis, Koerner, & Jacobson, 1993).

Treatment differentiation is never measured in treatment integrity research because adherence to the manual is considered to preserve intervention purity (Perepletchikova et al., 2007; Waltz et al., 1993). Therapist adherence and competence constitute a much more complicated relation than the relation between adherence and differentiation in the sense that competence presupposes adherence, but adherence does not presuppose competence (McGlinchey & Dobson, 2003).

2.1.2 Treatment integrity procedures

Following Perepletchikova, Treat and Kazdin (2007), there are four domains of treatment integrity that outcome studies on treatment integrity have to take into account: the establishment, assessment, evaluation, and reporting of treatment integrity. Procedures for establishing treatment integrity encompass the provision of a manual, training of the therapists in the intervention and supervision of these therapists. The purpose of a manual is to specify the treatment and strategies for its implementation, therewith reducing the variability in treatment implementation. A distinction is made between providing a general manual and a specific manual. A manual is general when it is written at a high level of abstraction. A manual is specific when it is detailed and explicit, and treatment components are operationally defined (Perepletchikova, 2006).

Training of therapists is necessary for a faithful rendition of the treatment. Training procedures can be indirect or direct. Indirect procedures include didactic instructions and written materials about the intervention. Direct procedures include opportunities for practice and role-play. Including these opportunities are said to make it less likely that therapists deviate from the treatment protocol (Perepletchikova, 2006). To assure the consistency and accuracy of the imple-

mentation of the treatment, therapists should be supervised. The procedures to establish treatment integrity enable therapists to deliver an intervention as intended (Schoenwald, Garland, Southam-Gerow, Chorpita, & Chapman, 2011). It can be seen as a *sine qua non* for therapist adherence and competence.

Procedures for assessment of treatment integrity relate to the method used to assess adherence and/or competence, and the validity and reliability of the instruments that are used to measure adherence and/or competence. A distinction is made between direct and indirect instruments to assess treatment integrity. Direct instruments are used to directly observe treatment delivery, such as a videotape of the session. Indirect instruments are used by therapists to rate their own adherence and/or competence levels, by subjects to rate what was done by a therapist (by means of interviews or questionnaires), or can consist of collection of products such as written assignments made by the therapist. Indirect instruments are sensitive for biases and distortions because they are subject to the tendency to provide socially desirable answers and subjective recollections. These distortions can affect the accuracy of the reported adherence. In order to measure integrity accurately, research should therefore not primarily rely on indirect ratings. It is rather recommended to use indirect ratings only to supplement observational data gathered by direct instruments (Perepletchikova et al., 2007).

Procedures for the evaluation of treatment integrity involve the accuracy of the representation of the obtained integrity data, the training of the raters, the assessment of interrater reliability and control over the reactivity of therapists on the measures taken, referred to as measurement reactivity (Perepletchikova et al., 2007). The last domain of the implementation of treatment integrity in outcome studies is the reporting of the findings. Procedures involve the reporting of numerical data of treatment integrity levels and reporting information of overall, component or/and session integrity.

The main goal of this systematic review is to evaluate the adequacy of the implementation of treatment integrity procedures in outcome studies of evidence-based youth interventions for externalising behaviour problems. We have formulated three questions;

1. Are treatment integrity procedures overall implemented adequately in outcome studies on externalising behavioural problems in youth?
2. Are the four domains of treatment integrity procedures implemented adequately in outcome studies on externalising behavioural problems in youth?
3. To what extent do researchers implement the procedures that relate to the four domains of treatment integrity in outcome studies on externalising behavioural problems in youth?

2.3 Method

2.3.1 Literature search procedures

To identify studies for this review we searched the following databases: Academic Search Premier, Cochrane Database of Systematic Reviews (CDSR), Cochrane Controlled Trials Register (CENTRAL), Database of Abstract of Reviews of Effects (DARE), ERIC, MEDLINE, NARCIS , Picarta, PsychINFO, Sciencedirect, and Web of Science.

The search terms used are: *therapist adherence* OR *therapist competence* OR *integrity* OR *fidelity* AND *outcome* AND (*juvenile, youth, adolescents, youngsters, children*). All databases but Web of Science have been searched on abstract level. No restriction has been made on publication date. This search resulted in a total of 686 obtainable studies.

2.3.2 Inclusion criteria

Included studies:

- evaluate the effects of an evidence-based intervention for children and young people with externalising behavioural problems.
- are about interventions for children and young people falling in the age of 0-23.
- target (and assess) externalising behaviour problems, including delinquency, disruptive behaviour, bullying, drug (ab)use, school dropout, temper tantrums, aggressive behaviour, conduct disorder or oppositional defiant disorder.
- take treatment integrity/fidelity into account.
- are primary studies.
- are published in English, Dutch or German language.

2.3.3 Exclusion criteria

Excluded studies:

- have a purpose other than the evaluation of the effects of an evidence-based intervention for children and young people aged 0-23 with externalising behavioural problems, including examination of mediators or moderators of therapeutic processes, risk factors, cost effectiveness of the intervention, barriers to treatment implementation, characteristics of treatment sample and treatment setting.

- evaluated interventions that are not delivered by treatment agents (e.g. bibliotherapy, computerized or mail-based therapies, self-help therapies)
- evaluated pharmacological interventions only.

2.3.4 Regarding the definition of ‘evidence-based’ interventions

In the field of mental health for children and adolescents there is a growing consensus that provided interventions should be evidence-based. Despite this consensus, the exact definition of evidence-based interventions can be regarded as a contentious matter (De Swart et al., 2012; Weisz, Jensen-Doss, & Hawley, 2006). Qualifications of evidence-based stretch from the perspective that interventions receive qualitative, theoretical, and/or clinical support, to the perspective that evidence-based constitutes clear empirical support provided by at least two randomized controlled trials (Veerman & van Yperen, 2007). In this review evidence-based interventions refer to interventions that at least: *are theoretically based, well-documented, protocolled and structured, contain a manual, and have gained empirical support in (quasi-)experimental research.* Evidence-based is thus considered in a broad sense. Interventions that can be considered promising –in that there are indications for their effectiveness- are included. Whenever it was not clear from the article that the described intervention suffices this description, additional sources, like the internet and manuals, were used to gather information on this inclusion criterion.

2.3.5 Study selection

A three step decision-making process was used for the selection of studies. At first the study titles were evaluated. Studies that obviously did not meet the inclusion and exclusion criteria of this review were rejected ($N=441$). Second, the study abstracts were screened. Studies that did not meet the criteria of this study were rejected ($N=121$). The third step then, involved a complete analysis of the study. When studies did not meet the criteria of this review they were rejected ($N=98$).

The search procedure was carried out by the first author using a scale with three categories: 1) obviously within all inclusion criteria, 2) doubtful, 3) obviously not within inclusion criteria. The doubtful cases ($N=26$) were screened by the last author. Whenever that led to a categorization in category 3 the study got rejected, studies categorized by the second reader in either category 1 or 2 were discussed until agreement on the inclusion or exclusion of the study was reached.

The search resulted in a total of 26 *articles* covering 32 outcome *studies* of 11 different evidence-based *interventions* for children and young people with exter-

nalising behaviour problems. Some studies examined the same evidence-based interventions under different circumstances⁴, when taking these into account a total of 13 interventions was included. Two studies (Liddle, Dakof, Henderson, & Rowe, 2011; Toffalo, 2000) used a definition of treatment adherence other than the degree of utilization of specified procedures by the therapist (e.g., following the manual verbatim, performing all prescribed task and activities). These studies have been excluded for further examination. From here on we therefore refer to 24 articles⁵ covering 30 outcome studies of 11 (with different conditions 13) evidence-based interventions. Almost half of the studies, 45 percent, pertained to Multisystem Therapy (MST); the other interventions were other multi-systemic or intensive family interventions such as Functional Family Therapy (FFT) and Multidimensional Family Therapy (MDFT).

2.3.6 Measure

In an effort to build a coherent literature base on the implementation of treatment integrity procedures, it is necessary to use a common language to define and measure treatment integrity. The Implementation of Treatment Integrity Procedures Scale (ITIPS) has proven to allow for systematic, reliable coding of integrity procedures in outcome studies (Perepletchikova et al., 2007). The ITIPS enables coding of TI procedures as well as evaluation of these procedures based on multiple recommendations in the implementation literature. For a discussion of these recommendations in more detail see Perepletchikova, Treat and Kazdin (2007).

We used a modified version of the Implementation of Treatment Integrity Procedures Scale (ITIPS) (Perepletchikova et al., 2007). The ITIPS originally consists of 22 items, covering the domains of establishment, assessment, evaluation and reporting of treatment integrity in outcome studies (Perepletchikova, 2006). Each item is rated on a 4-point scale.

In the adapted version (ITIPS-A) item 2 (definition of competence) was adapted in rating procedure. In the ITIPS-A studies are continued to be rated even when competence is not measured or another definition of competence (those of the authors) is given. In the original ITIPS this situation would lead to a rating of 1 (lowest possible rating) on every other item, regardless of the reported assessment and implementation procedures. The authors chose to continue rating to be able to collect information on the remaining items. Item 15 (training of raters) score 4 and item 16 (assessment interrater reliability) scores 3 and 4 in the ITIPS-A, were extended with the option 'indirect instrument'. When studies

⁴ In two MST studies MST therapists were provided with extended supervision and counseling to determine the effects on treatment integrity.

⁵ All studies included in this systematic review are indicated with a star in the list of references.

use indirect instruments it is mere logic that there is no training of raters and no assessment of interrater reliability. The ITIPS does not account for this and forces to score these studies with a 1 on the 4-point scale. The type of instruments studies use for adherence and/or competence, however, is already scored in items 8 and 9 of the ITIPS.

2.3.7 Data evaluation procedures

As outcome studies do not always provide full detailed information on intervention specific items of the ITIPS-A, additional sources were used to gather information on the specificity of the manual, the training protocol for therapists, the supervision protocol for therapists and the validity and reliability of the measurement instruments.

Following the procedure of Perepletchikova, Treat and Kazdin (2007), the implementation of integrity procedures in the outcome study was classified as inadequate, approaching adequacy, and adequate. A classification was given for the total score on the ITIP-A as well as for the 4 domains of the ITIPS-A. Table 1 provides an overview of the classification and its range of scores.

Table 1

Classification levels and range of scores on the domains of the ITIPS-A

	Inadequate (IA)	Approaching adequacy (AA)	Adequate (AD)
Establishment	6-12	13-18	19-24
Assessment	7-14	15-20	21-28
Evaluation	5-10	11-15	16-20
Reporting	4-8	9-12	13-16
Total score	22-44	45-66	>66

2.3.8 Internal consistency of the ITIPS-A

The 22-item ITIPS-A demonstrated sufficient internal consistency for three domains of integrity (.66 for establishing; .65 for assessing; .64 for evaluating), but showed marginal inconsistency for the domain of reporting treatment integrity (.55).

2.3.9 Rater training and interrater agreement

In two sessions the first author trained a master student in children's studies how to apply the coding criteria. To determine interrater agreement, the first

author independently recoded all studies coded by the master student ($N=9$, 28%). Interrater reliability was estimated using Cohen's Kappa. An interrater agreement of 0.633 was obtained. After determination of interrater reliability the author and student discussed all differently scored items until consensus was reached. These ultimate scores were used in the data file.

2.4 Results

2.4.1 Research Question 1

1. *Are treatment integrity procedures overall implemented adequately in outcome studies on youth externalising behavioural problems?*

Three studies (10%) adequately implemented treatment integrity procedures (see table 2). A total of 24 studies (80%) approached adequacy in the implementation of treatment integrity procedures. The remaining three studies (10%) implemented the procedures inadequately.

Table 2

Adequacy levels of the total implementation of treatment integrity procedures in outcome studies

Variable	Total Treatment Integrity		
	IA	AA	AD
Overall (N)	3	24	3
Overall (%)	10	80	10
Mean Score	40	56.63	79.33
SD	4.36	4.51	0.58
Min-Max	35-43	45-63	79-80
Range	22-44	45-66	>66

Note. IA= inadequate; AA= approaching adequacy; AD= adequate

Total studies $N=30$

2.4.2 Research Question 2

2. *Are the four domains of treatment integrity procedures implemented adequately in outcome studies on youth externalising behavioural problems?*

Table 3 shows the adequacy levels of the studies per domain. Procedures for establishing treatment integrity were implemented inadequately in 2 studies (6.7%), approached adequacy in 6 studies (20%), and were adequate in 22 studies (73.3%). Procedures for assessing treatment integrity were implemented inadequately in 7 studies (23.3%). In 20 studies (66.7%) the assessment procedures approached adequacy and in 3 studies (10%) the assessment procedures were adequate.

Procedures for evaluating treatment integrity were implemented inadequately in 7 studies (23.3%), approached adequacy in 19 studies (63.3%), and were adequate in 4 studies (13.3%). Procedures for reporting treatment integrity were implemented inadequately in 10 studies (33.3%). A little over half of the studies ($N=17$, 56.7%) approached adequacy in the domain of reporting treatment integrity. There were 3 studies (10%) that implemented procedures for reporting of treatment integrity at an adequate level.

2.4.3 Research Question 3

3. *To what extent do researchers implement the procedures that relate to the four domains of treatment integrity in outcome studies on youth externalising behavioural problems?*

Establishing treatment integrity

In all outcome studies a manual of the intervention was provided to the therapists. Almost all studies ($N= 26$, 86.7%) provided a specific manual in which treatment components are operationally defined. A general manual, which is a manual written at a high level of abstraction, was provided in 4 (13.3%) studies.

Table 3
Adequacy levels of the implementation of treatment integrity procedures in outcome studies per domain

Variable	Establishing			Assessing			Evaluating			Reporting		
	IA	AA	AD	IA	AA	AD	IA	AA	AD	IA	AA	AD
Overall (N)	2	6	22	7	20	3	7	19	4	10	17	3
Overall (%)	6.7	20	73.3	23.3	66.7	10	23.3	63.3	13.3	33.3	56.7	10
Mean Score	12	16.67	19.95	12	16.25	25	6.71	13.37	16.25	7.70	10.76	14.67
SD	0.00	1.03	1.68	2.24	0.55	1.73	1.98	1.12	0.50	0.95	0.83	0.58
Min-Max	12-12	16-18	19-24	7-13	16-18	24-27	5-10	11-15	16-17	5-8	9-12	14-15
Range	6-12	13-18	19-24	7-14	15-20	21-28	5-10	11-15	16-20	4-8	9-12	13-16

Note. IA= inadequate; AA= approaching adequacy; AD= adequate
 Total studies N=30

Table 4

Scores of the outcome studies on procedures for establishing treatment integrity

	Training of therapists		Supervision of therapists	
	Yes	No	Yes	No
%	86.7	6.7	80	13.3
N	26	2	24	4

Table 4 provides information on the training and supervision of therapists. Almost all therapists who were included in the studies were trained in the intervention (86.7%). In 2 studies the therapists who had to deliver the intervention were not trained in the intervention (Bruns, Suter, Force, & Burchard, 2005; Effland, Walton, & McIntyre, 2011). In 2 studies (Huey, Henggeler, Brondino, & Pickrel, 2000; Robbins et al., 2011) authors only mentioned that therapists were trained, but no other information was provided and no information could be found on general training procedures for the specific interventions used. One study (Walker, Golly, McLane, & Kimmich, 2005) used indirect training strategies only. The remaining 25 studies used indirect and direct training strategies to train the therapists in the intervention.

Supervision of therapists was not provided in 4 studies (Bruns et al., 2005; Effland et al., 2011; Holth, Torsheim, Sheidow, Ogden, & Henggeler, 2011). Two studies (Huey et al., 2000; Walker et al., 2005) only mentioned that therapists were supervised, but no other information was provided and no information could be found on general supervision procedures for the specific interventions used. All other studies ($N=24$) had ongoing supervision for the therapists during treatment in which they discussed cases and/or provided opportunities for practice and feedback. Closer examination of the studies that did not provide training and supervision for the therapists showed that these were studies in which the manual was general.

Assessment of treatment integrity

In one article (Stambaugh et al., 2007), adherence and/or competence was not assessed at all. Only the studies that assessed treatment integrity ($N=29$) were taken into account in the calculations. Table 5 shows that 3 studies assessed treatment integrity in terms of adherence *and* competence (Hogue et al., 2008; Robbins et al., 2011), 2 studies assessed treatment integrity only as competence (Eames et al., 2010; Eames et al., 2009), and the remaining 24 studies assessed treatment integrity solely as therapist adherence. In the 29 studies assessing treatment integrity, a total of 34 treatment integrity measurements were made.

Seventeen studies (50%) apply indirect methods⁶ only for assessing adherence. Another 13 (38,2%) do the same with direct methods. Only two studies (Glisson et al., 2010; Robbins et al., 2011) utilize both direct and indirect methods for their adherence and/or competence ratings.

Table 5

Scores of the outcome studies on procedures for assessment of treatment integrity

	Assessment of treatment integrity		Valid instruments		Reliable instruments		
	Adherence	Competence	A&C	Adherence	Competence	Adherence	Competence
%	82.8	6.9	10.3	77.8	100	77.8	60
N	24	2	3	21	5	21	3

Note. Percentages are based on the total of 29 studies assessing treatment integrity

Measuring adherence

A non-validated measuring method was used in 6 studies of the 27 studies assessing therapist adherence. This means that three-quarters ($N=21$, 77.8%) of the studies used validated methods to measure adherence. A non-reliable measuring method was also used in 6 studies of the 27 studies assessing therapist adherence. This means that three-quarters ($N=21$, 77.8%) of the studies used reliable methods to measure adherence. Closer examination of the studies shows that the 6 studies using non-validated methods were the same as the 6 studies using non-reliable methods.

Measuring competence

All studies ($N=5$) assessing competence used valid methods. Two studies (Hogue et al., 2008) assessing competence used non-reliable methods.

Evaluation of treatment integrity

From all the studies that assessed treatment integrity ($N=29$), all but 4 reported about the accuracy of the representation of the obtained integrity data. This means that 25 (86%) studies did provide information on this subject. Two studies (Bruns et al., 2005; Walker et al., 2005) collected treatment integrity data across one

⁶ Indirect methods refer to: therapists self-reports of procedures and activities implemented during sessions, debriefing subjects (by interview or questionnaire) on what was done by a therapist and collection of permanent products (e.g. planning notes, homework assignments) (Perepletkikova, 2006).

condition. In all other studies $N=23$, 79%) treatment integrity data was obtained under three or more conditions.

In the 29 studies assessing treatment integrity, a total of 34 treatment integrity measurements were used. Table 6 shows in how many cases the raters were trained to rate treatment integrity, whether interrater reliability was calculated for the measurements and if there was a control for measure reactivity. In 6 studies raters received no training to apply the measurement instruments (17.6%). For 19 studies (56%) training of raters was irrelevant, because they made use of an indirect method for assessing adherence and/or competence. In three studies (8.8%) the raters were trained in rating treatment integrity.

Table 6

Scores of the outcome studies on procedures of representation of treatment integrity

	Training raters			Assessment interrater reliability			Controlled for measure reactivity	
	Yes	No	Indirect	Yes	No	Indirect	Yes	No
%	8.8	17.6	56	29	15	56	50	50
N	3	6	19	10	5	19	16	16

Table 6 also shows the percentages of studies in which interrater reliability was assessed on the ratings of treatment integrity. Because 19 studies (56%) used an indirect method for assessing adherence and/or competence, these studies did not assess interrater reliability. Of the remaining studies, there were 5 studies (15%) that did not assess interrater reliability and 10 studies (29%) that did assess interrater reliability. In 16 (50%) measurements of treatment integrity, measurement reactivity was controlled for. In the remaining 50% of the measurements, studies did not mention that they controlled for measurement reactivity and no such thing could be indicated in the study description (see Table 6).

Reporting of treatment integrity

Table 7 shows the procedures for reporting treatment integrity. There were in total 27 studies assessing therapist adherence, 9 of these studies did not provide numerical data for adherence levels. Three of the studies that did provide numerical data provided data that was not informative of adherence levels. Informative data on adherence levels was provided in 15 studies assessing adherence. One (Eames et al., 2010) of the 5 studies assessing competence did not provide numerical data

on competence levels. All the others studies did provide numerical data, and that data was informative of competence levels.

Table 7

Provision of numerical data of treatment adherence and/or competence in the outcome studies

	Adherence			Competence	
	No	Not informative	Informative	No	Informative
%	33	11	56	20	80
N	9	3	15	1	4

2.4 Discussion

The results show that outcome studies of evidence-based interventions for children and young people with externalising behaviour problems that assess treatment integrity are (still) rare. Our search resulted in only 24 articles covering 29 studies that actually assessed treatment integrity. Of these studies a stunning 45 percent was about the same intervention -MST- mostly executed by the same researchers. This indicates that not only these kinds of studies are rare, but also are limited in scope, making it even more difficult to generalize findings.

Although it is generally recognized in the literature that treatment integrity is to be conceptualized as therapist adherence and competence, almost none of the studies addressed both aspects. Competence is an absolute outlier in treatment integrity measurements. An explanation might be that competence is a more difficult construct to measure, since measuring competence requires a judgment of behaviour in terms of quality. However, without measuring competence it is still unknown which indicators of competence may have compromised treatment progress and had an impact on the intervention outcome.

Although most studies did not address competence, almost all studies (80%) did approach adequacy in implementing treatment integrity procedures. The adequacy levels of implementation of treatment integrity procedures obtained with the ITIPS-A can only be used for descriptive purposes. It gives a mere overall evaluation of the implementation of procedures on treatment integrity in a study. For instance, it is possible to score 'adequate' on the total treatment integrity, while using a non-validated and non-reliable measure for adherence. Caution also has to be made with the interpretation of the scores on the domains of treatment integrity, since the internal consistency of the domains was only sufficient

to marginal. This indicates that the procedures in the domains on the ITIPS-A do not cover the domains well.

Procedures for establishing treatment integrity were implemented adequately in most studies. All studies reported on an intervention where a manual was provided to the therapists. This is congruent to the definition we hold for an evidence-based intervention. Still not all studies provided training and supervision for the therapists; these were studies in which the manual was general. An explanation could be that a specific manual is more accessible for formulating a training and supervision protocol. One can also reason that the more specific the manual, the easier it is to develop an instrument for measuring integrity. In a specific manual the elements of the intervention are clearly formulated and an instrument can be developed to measure these.

Almost three-quarters of the studies used a valid and reliable instrument to measure treatment integrity. However, since most studies were about the same intervention, Multi System Therapy, these studies all used the same valid and reliable instrument. Validity and reliability data of the instruments of the other interventions was not available in many cases. This makes interpretation of the data gathered with these instruments highly questionable. Moreover, most assessments of adherence and/or competence (56%) used indirect instruments. As stated before, indirect instruments have serious limitations because they are subject to the tendency to provide socially desirable answers and subjective recollections, which can cause biases and distortions in the adherence ratings. Ratings with indirect instruments should therefore be supplemented with ratings from direct instruments (Perepletchikova et al., 2007). Half of the ratings (50%) in the studies included in this review were based on indirect instruments only, which limit their ability to measure integrity accurately.

Evaluation procedures that are related to the use of the instrument, such as training of the raters and assessing interrater reliability, were not applicable in most studies because indirect instruments were used. When it comes to the reporting of data on treatment integrity measurements it was surprising to find that many studies did not report informative data on the integrity measures. It seems some authors do recognize the need to assess integrity, but then give priority to outcome information.

Although many authors share the opinion that measuring treatment integrity is not getting as much attention as it should, and have been stimulating the use of these measures, our findings suggest that measuring treatment integrity is still a forgotten issue in outcome studies of evidence-based interventions for youth with externalising behaviour problems. The lack of studies assessing treatment integrity adequately undermines the confidence we can have in statements made about the relationship between treatment integrity and intervention outcomes.

As we stated before, we hold the opinion that without adequate integrity *measurements*, the actual delivery of the intervention remains unknown and no statements can be made about the relationship between treatment integrity and outcomes.

2.4.1 Limitations and future directions

The ITIPS-A gives a clear view of procedures for implementing treatment integrity. The ITIPS-A has the potential to be a useful guide in developing integrity instruments and the use of these instruments in practice and research. The internal consistency of the domains of the ITIPS-A, however, was sufficient to marginal. This indicates that not all the procedures in the domains on the ITIPS-A cover the domains well. More research and practical use of the ITIPS-A is necessary to extend the procedures in these domains.

Specific search terms were used to find studies in the different databases. Studies that do take integrity/fidelity into account but have not used these words as key words thereby might have fallen out of reach of this review. The focus of this review is on evidence-based youth interventions targeting externalising behavioural problems in youth. The total scope of youth interventions is much broader than externalising behavioural problems, and many children and youth services also provide interventions that do not fall within the range of the definition used for evidence-based or promising interventions in this review. Future research could evaluate whether implementing procedures for treatment integrity differs between different types of interventions. The measurement of treatment integrity on a greater scale will not only make it possible to compare research on this topic in a more comprehensive way, it also ultimately can lead to more power in defining the relationship between treatment integrity and treatment outcome than has been done so far. As a first step in this direction the authors are now performing a meta-analysis with the studies of this review that did adequately implement integrity procedures to have a closer look at the relationship found in these studies.

CHAPTER 3

Making ‘What Works’ Work: A meta-analytic study of the effect of treatment integrity on outcomes of evidence-based interventions for juveniles with antisocial behavior

Provisionally accepted as:

Goense, P.B., Assink, M., Stams, G.J.J.M., Boendermaker, L., & Hoeve, M. (2016). *Making ‘What Works’ Work: A meta-analytic study of the effect of treatment integrity on outcomes of evidence-based interventions for juveniles with antisocial behavior.*

Abstract

This study meta-analytically examined the effect of treatment integrity on client outcomes of evidence-based interventions for juveniles with antisocial behavior. A total of 17 studies, from which 91 effect sizes could be retrieved, were included in the present 3-level meta-analysis. All included studies, to a certain level, adequately implemented procedures to establish, assess, evaluate and report the level of treatment integrity. A moderator analysis revealed that a medium-to-large effect of evidence-based interventions was found when the level of treatment integrity was high ($d = 0.633, p < 0.001$), whereas no significant effect was found when integrity was low ($d = 0.143, ns$). Treatment integrity was significantly associated with effect size even when adjusted for other significant moderators, indicating the specific contribution of high levels of treatment integrity to positive client outcomes. This implies that delivering interventions with high treatment integrity to youth with antisocial behavior is vital.

Highlights

- The moderating effect of treatment integrity was examined meta-analytically
- Studies were included if adequate treatment integrity procedures were applied
- Medium-to-large significant effects were found for high treatment integrity
- Small and non-significant effects were found for low treatment integrity
- Delivering interventions with high treatment integrity should be stimulated

Keywords

Treatment integrity; adherence; competence; client outcomes; evidence-based interventions; meta-analysis.

3.1 Introduction

It takes about seven years to develop and implement an evidence-based intervention in a community setting, around 17,000 dollar to provide it to a single juvenile, and on average, juveniles in youth care are exposed to an intervention for a 12-month period (Aos, Miller, & Drake, 2006; Kalidien, de Heer- de Lange, & van Rosmalen, 2010). Without assuring the proper delivery of interventions, there is a chance that interventions might not produce the desired effects and leave many youths with significant problems underserved or unserved (Fulda, Lykens, Bae, & Singh, 2009; Kataoka, Zhang, & Wells, 2002; McLeod, Southam-Gerow, Tully, Rodriguez, & Smith, 2013b; Perepletchikova & Kazdin, 2005), which can have serious negative consequences for both these youngsters and their social environment. The community can be confronted with criminal (re)offenses, which impose substantial psychological costs (e.g., victimization) and financial costs on society (e.g., the expenses of imprisonment are on average 700 dollar a person a day), especially when this behavior turns into persistent delinquent behavior (Algemene Rekenkamer, 2012; Cohen, Piquero, & Jennings, 2010). For that reason, it is important to effectively prevent or decrease juvenile antisocial behavior. This meta-analysis is the first to examine the effect of treatment integrity (i.e., delivery of the intervention as intended) on the effectiveness of evidence-based interventions for juveniles with antisocial behavior, while taking the operationalization of treatment integrity into account.

3.1.1 Treatment Integrity and Client Outcomes

There is a growing number of intervention studies examining the effect of treatment integrity on client outcomes. These studies have found mixed effects. Several studies showed that higher levels of treatment integrity were associated with greater reduction of adolescent's antisocial behavior, whereas other studies did not find such an association. Interestingly, one study examining the effects of individual drug counseling in adult patients, found support for a curvilinear relation between treatment integrity and outcomes, with both low and high levels of integrity showing worse outcomes, and intermediate levels showing the best outcomes (Barber et al., 2006). Barber et al. (2006) argued that very high levels of treatment integrity might reflect a lack of flexibility on the part of the therapist in responding to the client's needs, whereas very low levels of treatment integrity might reflect an inability to translate a therapeutic model or theory into practice as prescribed, which may lead to unsatisfying outcomes. In addition to this explanation, Weisz, Ugueto, Cheron, and Herren (2013b) have pointed out that community clinic youths have high rates of comorbidity, which may require a shift of focus during treatment in order to be able to target the most

pressing problems, resulting in intermediate levels of treatment integrity. In their review of research on the influence of implementation on program outcomes in prevention and health promotion programs for children and adolescents, Durlak and DuPre (2008) found the maximum level of treatment integrity in outcome studies to be around 80%, and they estimated that positive client outcomes can be expected when levels of treatment integrity are over 60%.

Given that previous research has revealed somewhat inconsistent findings on the association between treatment integrity and client outcomes, a meta-analysis could provide insight into the overall effect of treatment integrity. Previous meta-analyses on the effects of interventions for juveniles with antisocial behavior have suggested that delivering an intervention with high integrity is associated with positive client outcomes. Based on 548 independent study samples, Lipsey (2009) demonstrated that higher quality implementation of interventions targeting juvenile delinquency, such as surveillance, deterrence, discipline, restorative programs, counseling, and skill building programs, was associated with a reduction in recidivism of offending juveniles. Based on 30 independent study samples, Tennyson (2009) concluded that individual, family, group, or multisystemic therapies, as well as correctional programs, parent training, interventions focusing on peer influences, or restitution programs that were delivered with the highest level of treatment integrity produced the greatest reduction in recidivism of juvenile offenders. Thus, based on this research it can be concluded that higher levels of treatment integrity are related to more positive outcomes, which is specifically true for the reduction of recidivism. However, these previous meta-analyses did not take the quality of treatment integrity procedures of the included studies into account, while the validity of treatment integrity measurement likely has consequences for the interpretation of findings.

3.1.2 Measurements of Treatment Integrity

Measuring treatment implementation is needed to determine whether an intervention failed due to the failure of the intervention or its components, or due to the insufficient or inadequate application of the intervention (Schoenwald et al., 2011). Treatment integrity encompasses two aspects: 1) therapist adherence and 2) therapist competence (Perepletchikova, Treat, & Kazdin, 2007; see for a thorough discussion, Goense, Boendermaker, & van Yperen, 2016b, Goense, Boendermaker, van Yperen, Stams, & van Laar, 2014). Therapist adherence can be described as the degree to which the therapist delivers the prescribed components of a specific intervention (i.e., the delivery of an intervention is consistent with the intervention manual). Therapist competence refers to the level of the therapist's technical skills and judgment (timing and appropriateness) in delivering the components of the intervention (Barber et al., 2006; Barber,

Sharpless, Klostermann, & McCarthy, 2007; Barber, Triffleman, & Marmar, 2007; Perepletchikova et al., 2007). As for therapist competence, McLeod et al. (2013b) divided competence into a) technical competence, pertaining to specific components of the intervention, such as the delivery of behavioral cognitive elements in interventions for youth with aggression problems and b) common competence, pertaining to common (non-specific) elements of treatment (e.g., alliance and creating positive expectancies).

Therapists might also be experienced in delivering particular treatment methods acquired in previous therapeutic work that are not part of the specific intervention under study (McLeod et al., 2013b). The degree to which the therapists deliver these other treatment methods and consequently deviate from the planned intervention is referred to as treatment differentiation (Kazdin, 1994). Some researchers have suggested that measuring treatment differentiation is not necessary, because the assessment of treatment adherence is considered to preserve intervention purity (e.g., Perepletchikova et al., 2007; Waltz, Addis, Koerner, & Jacobson, 1993). However, McLeod et al. (2013) argued that without measuring treatment differentiation, examining additional treatment methods that may decrease or increase treatment effects is not possible.

The meta-analyses from Lipsey (2009) and Tennyson (2009) on the effects of interventions for juveniles with antisocial behavior examined if treatment integrity increased treatment efficacy. Lipsey (2009) considered level of involvement of the researcher in treatment implementation as a proxy for the extent to which attention was given to implementing the intervention as intended. Tennyson (2009) examined whether a specific treatment was manualized, if training was provided to practitioners, therapists received supervision, and/or were engaged in adherence checks. Tennyson (2009) grouped these four measures together as a novel means of assessing treatment integrity. Considering the construct of treatment integrity, it is highly questionable whether the operationalization of treatment integrity used by Tennyson (2009) and Lipsey (2009) was valid and comprehensive enough to assess the delivery of the intervention as intended in the studies that were included. These meta-analyses (Lipsey 2009; Tennyson 2009) did not actually measure delivery of the intervention in terms of adherence and/or competence, and therefore the assessment of treatment integrity was compromised because of construct underrepresentation.

It can be argued that meta-analyses on this topic have not operationalized treatment integrity in such a way that delivery as intended can be determined in the primary studies that were included. Therefore, a new meta-analytic study that focuses on studies that have incorporated an adequate (sufficiently comprehensive) operationalization of treatment integrity procedures, is needed. With the upcoming focus on treatment integrity, and the growing resources to measure

this construct, the demands on the measurement and reporting of treatment integrity in clinical trials are increasing (Fixsen & Ogden, 2014). This enables to conduct a meta-analytic study that takes an adequate (sufficiently comprehensive) operationalization of treatment integrity procedures in the primary studies into account.

3.1.3 The Present Meta-analysis

The aim of this study is to determine the impact of treatment integrity on client outcomes of evidence-based interventions for juveniles with antisocial behavior. This study differs from previous meta-analyses of Lipsey (2009) and Tennyson (2009) in how treatment integrity procedures are operationalized. In the present meta-analysis, a more valid and comprehensive operationalization of treatment integrity procedures has been used as a selection criterion for the primary studies that were to be included. This operationalization enables an assessment of the degree to which interventions are delivered as intended in the primary studies. In the present meta-analysis we examined whether treatment integrity is a moderator of the reduction of client antisocial behavior after an intervention. In addition to treatment integrity, other study characteristics possibly moderate the reduction of client antisocial behavior after an intervention, including intervention (e.g., intervention type, intervention duration, intervention modality) and methodological characteristics (e.g., study design and follow up time). We subsequently examined these characteristics as moderators. Finally, we examined the unique contribution of several moderating variables in a multivariate (multiple moderator) model.

3.2 Methods

3.2.1 Inclusion Criteria

To be included in the current meta-analysis, studies had to evaluate the effects of an evidence-based intervention targeting juveniles with antisocial behavior. We included studies on the basis of four criteria. First, studies had to examine the effectiveness of evidence-based interventions. In this meta-analysis, evidence-based interventions refer to interventions that at least are theoretically based, well documented, protocolled and structured, described in a manual, and have gained empirical support in (quasi-) experimental research (Weisz, Jensen-Doss, & Hawley, 2006). Second, studies had to focus on juvenile participants. Participants could be male or female children, adolescents and/or young adults with various ethnic backgrounds up to 23 years of age. Third, out-

come measures reported in primary studies had to include antisocial behavior. The interventions examined in the primary studies had to target (and assess) antisocial behavior, operationalized either as delinquency, disruptive behavior, bullying, drug (ab)use, school dropout, temper tantrums, aggressive behavior, conduct disorder, or oppositional defiant disorder.

Finally, studies had to adequately implement procedures to establish, assess, evaluate, and report the level of treatment integrity. The Implementation of Treatment Integrity Procedures Scale – Adapted (ITIPS-A) (Goense et al., 2014) was used to determine whether a study implemented treatment integrity *procedures* stringently enough for inclusion in the present meta-analysis. The ITIPS-A is an adapted version of the ITIPS, which was developed by Perepletchikova (2006). The ITIPS-A consists of 22 items, covering the domains of establishment (provision of manual, training, and supervision of therapists), assessment (methods used, and validity and reliability of instruments), evaluation (accuracy of data, training of raters, interrater reliability, and measurement reactivity) and reporting (numerical data on integrity) of treatment integrity in outcome studies. Each item is rated on a 4-point scale. The ratings on the scale are based on multiple recommendations in the implementation literature (Goense et al., 2014). For a discussion, see Perepletchikova et al. (2007). In the study by Goense et al. (2014), the ITIPS-A has shown disputable internal consistency for the domains establishing (.66), assessing (.65), and evaluating (.64), treatment integrity, and marginal consistency for reporting treatment integrity (.55). However, the items on the domains measure broad constructs (establishing, assessing, and evaluating of treatment integrity), and as a result lower Cronbach's alpha have been found than one would expect for scales measuring more narrow constructs (Peters, 2014). For that reason we consider the values to be acceptable for this particular instrument.

Using the ITIPS-A, the quality of the implementation of integrity procedures (establishment, assessment, evaluation and reporting) in outcome studies can be classified as adequate (score > 66), approaching adequacy (score ≥ 45 and ≤ 66), and inadequate (score ≤ 44). Only studies that could either be classified as approaching adequacy or adequate based on the total score on the ITIPS-A, were included in this study. Interrater reliability of the ITIPS-A was assessed in this study for 26.7% of the cases and was estimated using Cohen's Kappa. An interrater agreement of .734 was obtained, indicating that the agreement was sufficient (Landis & Koch, 1977).

3.2.2 Search Strategy

To identify relevant studies, the following databases were inspected: Academic Search Premier, Cochrane Controlled Trials Register (CENTRAL), ERIC, MEDLINE, NARCIS, Picarta, PsychINFO, Scencedirect, and Web of Science. Searches were conducted and studies were included up to November 2015. The databases were searched using the search string: (therapist adherence OR therapist competence OR integrity OR fidelity) AND (outcome) AND (juvenile OR youth OR adolescents OR youngsters OR children). This search yielded in 1,272 unique and obtainable reports. The final sample of studies consisted of 17 independent (non-overlapping) outcome studies that were described in 14 articles. See Appendix A for a flow chart of the search results.

3.2.3 Coding of studies

The aim of this study was to examine whether treatment integrity is a moderator of the reduction of client antisocial behavior after an intervention. In the primary studies, various instruments were used to assess levels of treatment integrity. In one study, a dichotomous checklist was used to assess treatment integrity, all other studies used Likert scales. We collected background information and indication criteria of scores of all instruments on the webpages of the interventions' agencies, in relevant articles and (when available) intervention manuals. We used these indication criteria to interpret the integrity level scores reported in the primary studies (low, intermediate or high). We collapsed the categories low and intermediate integrity to preserve adequate statistical power and coded in two levels, low integrity versus high integrity.

Besides classifying levels of treatment integrity, we coded study and intervention characteristics that are possible moderators of the effects of interventions. We coded the following intervention characteristics: intervention type (Multisystem Therapy (MST), Wrap-Around (Wrap), Functional Family Therapy (FFT), Motivational Interviewing (MI), and remaining interventions), intervention modality ((multi)-systemic versus individual) and intervention duration (in months). As for study characteristics, we coded: study design (experimental versus non-experimental), follow-up time (in months), and outcome measure (substance abuse, delinquency, externalizing behavior problems, other behavior problems). See Appendix B and C for an overview of included primary studies and several characteristics of interventions examined in these studies.

To assess interrater agreement, the first author (a PhD-candidate in Behavioral and Social Sciences with a Master's degree in Criminal Law and in Forensic Child and Youth Care Sciences) and a second coder (a graduate student in Child Development Studies) independently coded 15 of the 17 studies (88.2%). Percentages

of agreement were calculated to assess inter-rater reliability. The percentage agreement ranged from good (at least 70% agreement) for the variables integrity level in study (72.6%), calculated effect size (79%), score on ITIPS-A (89%), study design (91.8%), intervention duration (93.2%), outcome measure (97.3%), to perfect (100% agreement) for the variables follow-up time and intervention modality. Disagreements were solved by discussion of the first and second author.

3.2.4 Data analysis

For all studies, Cohen's d was computed as the common effect size. In most instances, Cohen's d was calculated on the basis of mean scores and standard deviations. In some cases the calculation of Cohen's d was based on reported test statistics, such as F , p or t values, or on differences between percentages. The reported values were transformed into Cohen's d using an effect size computation program based on the formulas of Lipsey and Wilson (2001). If available, we calculated effect sizes for follow up assessments in which follow up time was defined as the period after end of treatment.

Due to the small number of available primary studies in which treatment integrity procedures were adequately operationalized and which met all other inclusion criteria, effect sizes of both experimental and non-experimental studies were included in order to increase the statistical power of the analyses. Because attention is warranted when effect sizes are extracted from studies with different study designs (see Lipsey & Wilson, 2001), we examined whether study design moderated the overall effect and the potential effect of treatment integrity on effect size. We computed standardized mean gain effect sizes for non-experimental studies and standardized mean difference effect sizes for experimental studies, using formulas of Lipsey and Wilson (2001).

We used a multilevel random effects model in calculating the overall effect and in performing moderator-analyses, in order to explain heterogeneity of effect sizes (Hox, 2002). We used a three-level meta-analytic model to analyze the data, modelling three sources of variance: sampling variance of the observed effect sizes (Level 1), variance between effect sizes from the same study (Level 2), and variance between studies (Level 3) (Cheung, 2014; Houben, Van den Noortgate, & Kuppens, 2015; Van den Noortgate, López-López, Marin-Martinez, & Sánchez-Meca, 2013, 2014). Using this three-level approach to meta-analysis makes it possible to use all available effect sizes because effect sizes extracted from the same study (i.e., dependent effect sizes) can be modelled, so that all information in the studies can be preserved and maximum statistical power can be achieved (see also, Assink, van der Put, Hoeve, de Vries, Stams, & Oort, 2015).

The model was used to obtain an overall estimate of the effect of evidence-based interventions on antisocial behavior. If variation between effect sizes from the

same study or variation between studies was significant, we extended the model by including the moderator variables to determine whether this variation can be explained by characteristics of studies or effect sizes. The conventions formulated by Cohen (1988) were used to interpret the magnitude of the effect sizes, with effect sizes of $d \geq .20$ considered as small effects, $d \geq .50$ as medium and $d \geq .80$ as large effects.

We used the function “`rma.mv`” of the `metafor` package (Viechtbauer, 2010) in the R environment (version 3.2.2; R Core Team, 2015) for the statistical analyses. We followed other scholars (Assink et al., 2015; Houben et al., 2015; Weisz et al., 2013a) in their procedures for analyzing data using a three-level meta-analytic model. To determine whether the variance between effect sizes from the same study (Level 2), and the variance between studies (Level 3) was significant, two separate one-tailed log-likelihood-ratio-tests were performed in which the deviance of the full model was compared to the deviance of a model excluding one of the variance parameters. It is preferable to conduct these tests one-tailed, because variance components can only deviate from 0 in the positive direction. For these tests, the null hypothesis (H_0) was that a variance component is equal to 0 and the alternative hypothesis (H_1) was that a variance component is larger than 0. Conducting these tests two-tailed is considered too conservative (Wibbelenk & Assink, 2015). All model parameters were estimated using the restricted maximum likelihood estimation method. We conducted all other tests two-tailed because we did not have hypotheses about the relations between the potential moderators and the client outcomes. We considered p -values $< .05$ as statistically significant.

3.2.5 Publication Bias

A common phenomenon, referred to as the ‘file drawer problem’, is that studies with no significant findings are less likely to be published than studies with significant findings (Rosenthal, 1995), which may lead to publication bias. Because only published studies were included in this meta-analysis, publication bias was examined in three different ways. First, the fail-safe number was calculated. If the fail-safe number exceeds the critical value obtained with Rosenthal’s (1994) formula of $5 * k + 10$ (k is the number of studies included in a multilevel meta-analysis), there is no indication of publication bias. Second, we inspected the distribution of effect sizes, which should be shaped as a (symmetric) funnel if no publication bias is present (Sutton, Duval, Tweedie, Abrams, & Jones, 2000). Third, funnel plot asymmetry was tested by regressing the standard normal deviate, defined as the effect size divided by its standard error, against the estimate’s precision (i.e., the reciprocal of the standard error), which largely depends on sample size (Egger, Smith, Schneider, & Minder, 1997).

3.3 Results

3.3.1 Descriptive Statistics, Central Tendency and Variability, and Assessment of Missing Data

The present study included 14 articles describing $k = 17$ studies with a total of 91 effect sizes (see Appendix B). In total seven different intervention types were described in the studies: Multisystem Therapy (MST) ($k = 7$), Wrap-Around ($k = 3$), Functional Family Therapy (FFT) ($k = 2$), and Motivational Interviewing (MI) ($k = 2$). Further, the remaining interventions category of the moderator intervention type consisted of three different interventions: Tip-Sheet ($k = 1$), Multidimensional Family Therapy (MDFT) ($k = 1$), and Cognitive Behavioral Therapy (CBT) ($k = 1$). These interventions target antisocial behavior of juveniles, including delinquency, disruptive behavior, bullying, drug (ab)use, school dropout, temper tantrums, aggressive behavior, conduct disorder, or oppositional defiant disorder. The participants in these interventions were almost always mixed (male and female $n = 16$, 94.1%, unknown $n = 1$, 5.9%), up to 23 years of age, and (in case of system interventions) their families. More than half ($n = 4$, 57.1%) of the interventions were (multi)system focused, whereas CBT, MI and Tip-Sheet focus on individual participants. Treatment duration was between 1 and 15 months and the average treatment duration was 4.9 months.

There were $k = 12$ studies, covering 45 effect sizes, with a non-experimental design and $k = 5$ studies, covering 46 effect sizes, with an experimental design (see Appendix C). There were $k = 6$ studies that had a follow-up measurement. Follow-up time ranged between 3.7 and 24 months. Each study assessed between one and four outcome measures. Almost half of the effect sizes ($n = 39$, 42.9%) were based on the outcome measure substance abuse, delinquency was the second most examined outcome measure ($n = 24$, 26.4%), followed by externalizing behavior ($n = 19$, 20.9%) and other behavior ($n = 9$, 9.9%). The number of effect sizes in each study ranged between 1 and 20. Total ITIPS-A scores ranged between 50 and 79. Most studies ($k = 13$) were scored as *approaching adequacy* using the ITIPS-A. There were $k = 6$ (35.3%) studies in which the intervention was implemented with high levels of treatment integrity and these studies covered 13 (14.3%) effect sizes. In the remaining $k = 11$ (64.7%) studies covering 78 (85.7%) effect sizes, the interventions were implemented with low levels of treatment integrity.

The overall mean effect size was $d = 0.300$, $p = .005$, indicating that evidence-based interventions for juveniles with antisocial behavior, significantly reduced antisocial behavior (see Table 1). The fail-safe number was 583, which is above the critical value of 95 [$5 * 17 + 10$], indicating that publication bias

was unlikely. Possible publication bias was also examined by testing funnel plot asymmetry. The standard normal deviate was regressed against the estimate's precision. As the intercept did not significantly deviate from zero ($t = 0.912, p = .364$), there was no indication of funnel plot asymmetry and therefore no indication of publication bias.

The results of the likelihood-ratio tests revealed significant variance between effect sizes from the same study (i.e., level 2 variance) and significant variance between studies (i.e., level 3 variance; see Table 1). This means that, although overall studies showed a positive effect on antisocial behavior, this finding was not consistent across all effect sizes and studies. This indicates that intervention effects were likely to be influenced by study or intervention characteristics, which justifies the investigation of potential moderators.

3.3.2 Moderator Analysis

Moderator analyses were conducted for treatment integrity, study design, follow-up time, outcome measure used, intervention type, intervention modality, and intervention duration in order to determine characteristics of effect sizes or studies that can explain level 2 or level 3 variance. The results of the moderator analyses are presented in Table 1.

Table 1
Results for the overall mean effect size and moderators

Moderator variables	# Studies	# ES	Mean <i>d</i> (95% CI)	β (95% CI)	F(df1, df2)	Level 2 variance	Level 3 variance
Overall	17	91	0.300 (0.091, 0.510)**			0.077***	0.151***
<i>Integrity</i>					F(1, 89) = 6.367*	0.077***	0.101***
Low (RC)	11	78	0.143 (-0.071, 0.358)				
High	6	13	0.633 (0.312, 0.954)***	0.490 (0.104, 0.876)*			
<i>Study Design</i>					F(1, 89) = 6.898*	0.079***	0.091***
Non-experimental (RC)	12	45	0.457 (0.245, 0.669)***				
Experimental	5	46	-0.021 (-0.314, 0.272)	-0.478 (-0.840, -0.116)*			
<i>Follow up time</i>	6	39	0.289 (0.111, 0.466)**	-0.014 (-0.036, 0.008)	F (1, 37) = 1.594	0.128***	0.018***
<i>Outcome measure</i>					F (3, 87) = 1.302	0.076***	0.150***
Substance abuse (RC)	6	39	0.258 (-0.022, 0.538)+				
Delinquency	6	24	0.238 (-0.026, 0.503)+	-0.020 (-0.312, 0.272)			
Externalizing behavior problems	7	18	0.450 (0.183, 0.717)**	0.192 (-0.122, 0.506)			

Table 1 (Continued)

Moderator variables	# Studies	# ES	Mean <i>d</i> (95% CI)	β (95% CI)	F(df1, df2)	Level 2 variance	Level 3 variance
Other behavior problems	5	10	0.206 (-0.134, 0.545)	-0.053 (-0.428, 0.323)			
<i>Intervention Type</i>					$F(4, 86) = 3.888^{**}$	0.074 ^{***}	0.079 ^{***}
MST (RC)	7	50	0.280 (0.040, 0.520)*				
Wrap	3	7	-0.293(-0.700, 0.114)	-0.573 (-1.045, -0.100)*			
FFT	2	2	1.013 (0.421, 1.605)**	0.733 (0.095, 1.372)*			
MI	2	4	0.610 (0.068, 1.152)*	0.3300 (-0.263, 0.923)			
Remaining	3	28	0.360 (0.013, 0.708)*	0.0805 (-0.342, 0.503)			
<i>Intervention modality</i>					$F(1, 89) = 0.566$	0.076 ^{***}	0.159 ^{***}
(Multi)Systemic (RC)	13	71	0.256 (0.012, 0.500)*				
Individual	4	20	0.448 (0.005, 0.891)*	0.192 (-0.314, 0.697)			
Intervention Duration	14	86	0.290 (0.141, 0.440) ^{***}	-0.088 (-0.141, -0.035) ^{**}	$F(1, 84) = 11.048^{**}$	0.078 ^{***}	0.049 ^{***}

Note. # studies= number of studies; # ES =number of effect sizes; mean *d*=mean effect size (d); CI= confidence interval; β = estimated regression coefficient; F(df1, df2) = Omnibus test of all regression coefficients in the model; Level 2 variance = variance between effect sizes from the same study; Level 3 variance= variance between studies; MST = Multisystem Therapy; Wrap = Wrap around; FFT = Functional Family Therapy; MI = Motivational Interviewing.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Treatment integrity

A significant moderating effect was found for treatment integrity, $F(1, 89) = 6.367$, $p = 0.0134$ (see Table 1). Significant medium-to-large effects of evidence-based interventions were found when the level of treatment integrity was high ($d = 0.633$, $p < 0.001$), whereas small/marginal effects were found when integrity was low ($d = 0.143$, *ns*). The results of the likelihood-ratio tests showed that there was still significant variance between effect sizes from the same study (i.e., level 2 variance) and significant variance between studies (i.e., level 3 variance) when the meta-analytic model was extended with the moderator level of treatment integrity (see Table 1).

Moderating effects of study and intervention characteristics

A significant moderating effect was found for study design, intervention type, and intervention duration (Table 1). Medium significant effects of evidence-based interventions were found when the study had a non-experimental design ($d = 0.457$), and a non-significant small and negative effect was found when the design was experimental ($d = -0.021$). As for the intervention type, the largest significant effects were found for FFT ($d = 1.013$) and MI ($d = 0.610$), whereas a non-significant small and negative effect was found for the intervention Wrap Around ($d = -0.293$). MST and the remaining category yielded significant small to medium effects (MST, $d = 0.280$, Remaining, $d = 0.360$). We also found that effects of interventions on the reduction of client antisocial behavior problems decreased when intervention duration increased ($\beta = -0.088$). The results of the likelihood-ratio tests showed that there was still significant variance between effect sizes from the same study (i.e., level 2 variance) and significant variance between studies (i.e., level 3 variance) when the meta-analytic model was extended with either of above mentioned moderators (see Table 1). Follow-up time, outcome measure used, and intervention modality were nonsignificant.

3.3.3 Multivariate model

A multivariate analysis was conducted to examine the effect of level of treatment integrity over and above the effect of other significant moderators. In addition to treatment integrity, we included the moderators intervention type and intervention duration (Table 2). Study design was not included because of overlap with the moderator treatment integrity (all experimental studies had a low level of treatment integrity) and high multicollinearity ($VIF > 2$). We found that studies with high treatment integrity yielded significantly larger effect sizes, even after

adjusting for the effect of intervention characteristics ($\beta = 0.572$, $p = 0.005$). Studies examining FFT also showed significantly larger effect sizes ($\beta = 0.619$, $p = 0.019$) and studies with interventions in the remaining category showed a trend indicating somewhat larger effect sizes ($\beta = 0.229$, $p = 0.064$).

Table 2

Results for the Multiple Moderator Model

Moderator variables	β (SE)	95% CI	t-statistic	p-value
Intercept	0.094 (0.076)	-0.058; 0.246	1.235	0.220
<i>Treatment Duration</i>	-0.039 (0.032)	-0.103; 0.025	-1.205	0.232
Treatment Integrity				
<i>High Integrity</i>	0.572 (0.199)	0.177; 0.968	2.880	0.005 **
Intervention Type				
<i>Intervention Wrap</i>	-0.337 (0.275)	-0.884; 0.210	-1.226	0.224
<i>Intervention FFT</i>	0.619 (0.258)	0.106; 1.132	2.401	0.019 *
<i>Intervention MI</i>	-0.167 (0.267)	-0.699; 0.365	-0.626	0.533
<i>Intervention Remaining</i>	0.229 (0.122)	-0.014; 0.472	1.879	0.064 +
Omnibus test	$F(6, 79) = 6.143^{***}$			
Variance level 2 ^a	0.074			
Variance level 3 ^b	0.014			
# ES	86			

Note. CI = confidence interval; # ES = number of effect sizes; MST = Multisystem Therapy; Wrap = Wrap around; FFT = Functional Family Therapy; MI = Motivational Interviewing.

^a Variance between the effect sizes from the same study.

^b Variance between studies.

* $p < 0.05$

** $p < 0.01$

** $p < 0.001$

3.4 Discussion

The present meta-analytic study integrated previous findings on the link between evidence-based interventions and client outcomes, and examined the moderating effect of treatment integrity on client outcomes. This meta-analysis differs from

previous meta-analyses because only outcome studies that, to a certain level, have adequately operationalized treatment integrity procedures were included. By adding this inclusion criterion, we ensured that the level of treatment integrity was validly assessed, making it possible to draw firmer conclusions on the association between treatment integrity and client outcomes than previous meta-analyses.

Treatment integrity was found to be a significant moderator of the reduction of client antisocial behavior, most often assessed as substance abuse or delinquency, after an intervention. We found a significant difference between the effect sizes of studies with a high level of treatment integrity and studies with a low level of treatment integrity. Significant medium-to-large effects of evidence-based interventions were found when the level of treatment integrity was high ($d = 0.633$) and non-significant small/marginal effects were found when the integrity was low ($d = 0.143$). Other significant moderators were study design, intervention type and intervention duration. We performed a multivariate analysis to examine the effect of level of treatment integrity on client outcomes, adjusting for intervention characteristics (intervention type and intervention duration). The results showed that the association between high levels of treatment integrity and positive client outcomes remained the same when controlled for the other significant moderators, indicating the specific contribution of high levels of treatment integrity to client outcomes over and above of the effect of intervention characteristics. These results confirm previous research revealing that high levels of treatment integrity are positively associated with positive client outcomes (Lipsey, 2009; Tennyson, 2009). This means that delivering interventions with high treatment integrity is critical and should be stimulated. Scholars have suggested that an effective way to establish and maintain treatment integrity of planned interventions is frequent and targeted support of professionals (see Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005; Garland & Schoenwald, 2013; Goense, Boendermaker, & van Yperen, 2015a). There should be room for performance feedback and professionals' interactions with their clients should be reviewed and discussed. The results of this study underline the need to incorporate support systems for professionals in clinical practice.

Notable is the small amount of studies in which the intervention was implemented with high levels of treatment integrity. In only a third ($k = 6, 35.3\%$) the treatment integrity level was high. This may indicate that delivering an intervention with a high level of integrity is a difficult task. As stated by Weisz et al. (2013b), in the actual youth ecosystem, in which so many real-world factors are at play, the clients' needs can shift midcourse treatment, requiring the professional to shift focus. This requires the professional to attend to the clients' needs without drifting from the intended intervention. It also requires that the intervention is

flexible in use. It needs further examination whether an (in)ability of an intervention and/or professional to adapt to changing real-world factors during treatment is a moderator affecting the reduction of client antisocial behavior.

3.4.1 Limitations and future directions

The inclusion criteria that each study had to implement treatment integrity procedures stringently enough to be included in the present meta-analysis reduced the variability and the range of studies available to test the moderating effect of treatment integrity on the reduction of client antisocial behavior after an intervention. The mere fact that only 17 studies were included in this meta-analysis indicates that an adequate operationalization of treatment integrity procedures in primary studies is still rare. With an overrepresentation of MST, which covered almost half ($k = 7$, 41.2%) of the studies included in this meta-analysis, an adequate operationalization of treatment integrity in primary studies is not only rare, but also limited in scope. In recent years there has also been a shift in combining evidence-based interventions with the understanding that ‘one size does not fit all’ (Stewart, Felleman, & Arger, 2015). Studies examining the effects of these combined interventions commonly do not take treatment integrity into account or focus on treatment integrity of only a specific part of the intervention (i.e., Stanger, Ryan, Scherer, Norton, & Budney, 2015), probably because the instrumentation to adequately measure integrity of these combined interventions are not (yet) available. For that reason, these type of studies have been excluded from this meta-analysis. However, having used a strict criterion for inclusion of studies does give more certainty that the measurements that were made in the studies were comprehensive enough to validly assess levels of treatment integrity of the available interventions.

The small number of included studies limited the possibility of examining possible curvilinear relations between different levels of treatment integrity and intervention outcomes, because we had to collapse categories (low and intermediate integrity) to preserve adequate statistical power. The small number of studies also limited the possibility of distinguishing between studies with an experimental design and with a pre-post measurement design, and for that reason we combined these effect sizes in the same meta-analysis. We conducted post-hoc moderator analysis of treatment integrity in a subset of studies with a pre-post design (experimental studies all reported low treatment integrity). Results, though attenuated by reduced power, were essentially unchanged (mean $d = 0.577$ for high treatment integrity, mean $d = 0.337$ for low treatment integrity, $F(1,43) = 2.8$, $p = .10$).

In the present study, we could not differentiate between levels of adherence and competence, because most studies included in this meta-analysis did not

differentiate between adherence and competence. Also, the concept of treatment differentiation (the degree to which the therapists delivers other treatment methods, and consequently deviate from the planned intervention) was not taken into account because most studies up to November 2015 have not incorporated treatment differentiation in their measurements, possibly because instruments for distinguishing differentiation are not commonly available. Because treatment differentiation has not been taken into account, it is not certain whether the positive clinical outcomes were a consequence of additive or alternative treatment interventions. As stated by McLeod et al. (2013b), without measuring treatment differentiation, it is not possible to determine whether or not (and which) additional therapeutic procedures decrease or increase treatment effects.

It is recommended that future research distinguishes between the different aspects of treatment integrity (therapist adherence, therapist technical competence, therapist common competence and treatment differentiation) in order to fully understand how these different aspects of treatment integrity interact and have an impact on client outcomes. The identification of relevant components of an intervention provides crucial information on the content of training and support of professionals responsible for the delivery of the intervention in real world settings. Because real world settings for juveniles with antisocial behavior are commonly dictated by a more judicial (punitive) framework, the challenge is to examine if the identified components can be implemented with high integrity in these settings. In secure residential settings, for example, a punitive measure such as time out can be imposed due to (aggression) incidents, which can hamper the continuity and delivery of treatment (De Valk et al., 2015; Knotter, Wissink, Moonen, Stams, & Jansen, 2013; Van der Helm, Boekee, Stams, & Van der Laan, 2011). The challenge is making 'what works' work. This not only requires more research on effective components of interventions targeting clinical populations experiencing problems in multiple domains, but also research on how to make these interventions work under clinically representative conditions.

Acknowledgment

We thank Jose van Laar for coding and helping with analyzing the data.

Declaration of Conflicting Interests

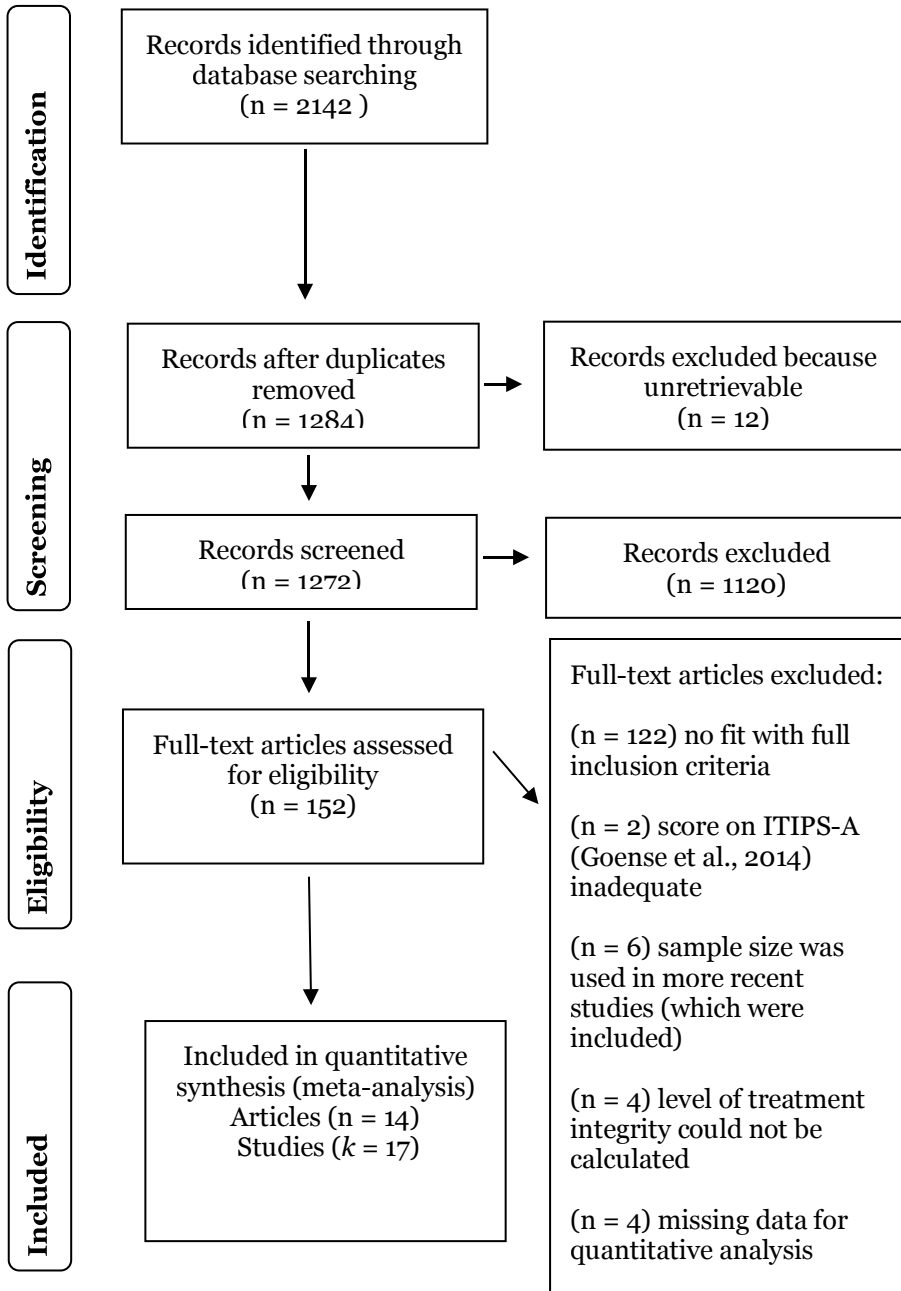
The authors declare no potential conflicts of interest with respect to the research, authorship, or publication of this article.

Funding

No funding was provided for this research.

Appendix A

PRISMA flow diagram of search procedures



Appendix B

Characteristics of interventions included in the meta-analysis

Authors	Publication year	Intervention type	Treatment modality	Treatment duration (months) ^a
Asscher, Dekovic, Manders, van der Laan, Prins, & Arum, & Dutch MST Cost-Effectiveness Study Group (2014)	2014	MST	(multi)system	5.7
Bruns, Suter, Force & Buchard	2005	Wrap around	(multi)system	6
Effland, Walton & McIntyre	2011	Wrap around	(multi)system	NA
Graham, Carr, Rooney, Sexton & Satterfield	2014	FFT	(multi)system	4.5
Graham, Carr, Rooney, Sexton & Satterfield	2014	FFT	(multi)system	4.5
Henggeler, Melton, Brondino, Scherer & Hanley	1997	MST	(multi)system	4
Henggeler, Pickrel & Brondino	1999	MST	(multi)system	4.3
Huey, Henggeler, Brondino & Pickrel	2000	MST (diffusion)	(multi)system	NA
Huey, Henggeler, Brondino & Pickrel	2000	MST (CDA)	(multi)system	NA
Liddle, Dakof, Henderson, & Rowe	2011	MDFT	(multi)system	5
Liddle, Dakof, Henderson, & Rowe	2011	CBT	Individual	5
Little, Hudson & Wilks	2002	Tip Sheet	Individual	1
Schoenwald, Sheidow, Letourneau & Liao	2003	MST	(multi)system	4.8
Stambaugh, Mustillo, Burns, Stephens, Baxter, Edwards & Dekraai	2007	Wrap around	(multi)system	15
Stewart, Felleman, & Arger	2015	MI	Individual	1
Smith, Ureche, Davis, & Walters	2015	MI	Individual	3
Sundell, Hansson, Löfholm, Olsson, Gustle & Kadesjö	2008	MST	(multi)system	4.8

Note. MST = Multisystem Therapy; Wrap = Wrap around; FFT = Functional Family Therapy; MI = Motivational Interviewing.

^aNA = not available

Appendix C

Study characteristics of studies included in the meta-analysis

Authors	Publication year	Research design ^a	Follow-up (months) ^b	Outcome measure	<i>k</i> (# effect sizes)	Total Integrity score ITIPS-A ^c	Integrity level
Asscher, Dekovic, Manders, van der Laan, Prins, & Arum, & Dutch MST Cost-Effectiveness Study Group (2014)	2014	Experimental	6 (10) & 24 (4)	Delinquency, externalizing behavior	20	50	Low
Bruns, Suter, Force & Buchard	2005	Non-experimental	No follow up	Other behavioral	4	50	High
Effland, Walton & McIntyre	2011	Non-experimental	No follow up	Externalizing behavior	1	52	Low
Graham, Carr, Rooney, Sexton & Satterfield	2014	Non-experimental	No follow up	Externalizing behavior	1	67	High
Graham, Carr, Rooney, Sexton & Satterfield	2014	Non-experimental	No follow up	Externalizing behavior	1	67	Low
Henggeler, Melton, Brondino, Scherer & Hanley	1997	Experimental	19 (2)	Delinquency & other behavior	5	59	Low
Henggeler, Pickrel & Brondino	1999	Experimental	6 (6)	Substance abuse & delinquency	11	62	Low
Huey, Henggeler, Brondino & Pickrel	2000	Non-experimental	No follow up	Delinquency	2	61	High
Huey, Henggeler, Brondino & Pickrel	2000	Non-experimental	No follow up	Delinquency	2	57	Low
Liddle, Dakof, Henderson, & Rowe	2011	Non-experimental	6 (4) & 12 (4)	Substance abuse	12	79	Low

Appendix C (Continued)

Authors	Publication year	Research design ^a	Follow-up (months) ^b	Outcome measure	<i>k</i> (# effect sizes)	Total score ITIPS-A ^c	Total Integrity level
Liddle, Dakof, Henderson, & Rowe	2011	Non-experimental	6 (4) & 12 (4)	Substance abuse	12	79	Low
Little, Hudson & Wilks	2002	Non-experimental	No follow up	Externalizing behavior & other behavior	4	51	Low
Schoenwald, Sheidow, Letourneau & Liao	2003	Non-experimental	No follow up	Externalizing behavior	2	56	High
Smith, Ureche, Davis, & Walters	2015	Non-experimental	No follow up	Substance abuse	2	61	High
Stambaugh, Mustillo, Burns, Stephens, Baxter, Edwards & Dekraai	2007	Experimental	No follow up	Other behavior	2	55	Low
Stewart, Felleman, & Arger	2015	Non-experimental	3-7 (1)	Substance abuse	2	63	High
Sundell, Hansson, Löfholm, Olsson, Gustle & Kadesjö	2008	Experimental	No follow up	Substance abuse, delinquency, externalizing behavior & other behavior	8	59	Low

Note.

^a Design refers to effect size calculation (pre-posttest or control-group) for the purpose of this meta-analysis.

^b Numbers between () indicate *n* = effect sizes applicable.

^c Total score ITIPS-A > 66 is indicated as adequate (Goense et al., 2014).

CHAPTER 4

Support Systems for Treatment Integrity

Published as:

Goense, P.B., Boendermaker, L., & van Yperen, T. (2015a). Support Systems for Treatment Integrity. *Research on Social Work Practice*, 1-5.

Abstract

Objective: This systematic review evaluates the content of effective support provided to practitioners of evidence-based interventions in order to establish and maintain treatment integrity. **Method:** Four articles covering six outcome studies are included in this review, these studies 1) adequately operationalized treatment integrity procedures and 2) showed sufficient effects of the support on levels of treatment integrity. **Results:** The results show that an effective support system consist of a minimum combination of training and ongoing support (supervision, consultation, coaching), preferably extended with booster sessions. **Conclusions:** The review provides specific information on effective content of this support which practitioners and intervention developers can use in order to establish and maintain treatment integrity of their planned interventions.

Keywords

Support systems, evidence-based interventions, treatment integrity

4.1 Introduction

Over the last decade much attention has been paid to the development and implementation of evidence-based intervention programs for troubled youths. One of the main difficulties with these evidence-based interventions are the disappointing treatment outcomes outside the research setting. Researchers frequently conclude that low treatment fidelity may be the cause of disappointing results of these intervention in the practice field (Schoenwald, Chapman, Sheidow, & Carter, 2009a; Sexton, & Turner, 2010; Tennyson, 2009). Treatment fidelity or integrity refers to the degree to which the practitioner utilizes an interventions prescribed procedures, avoids proscribed procedures (adherence) and proves the right level of skills and clinical acumen (competence) in delivering the components of the treatment (Barber, Triffeleman, & Marmar, 2007b; Perepletchikova, Treat, & Kazdin, 2007).

Research suggests that frequent and targeted support of practitioners is an effective way to establish and maintain treatment integrity (Kerby, 2006; Miko-lajczak, Stals, Fleuren, Wilde & Paulussen, 2009; Schoenwald et al, 2009b). Most evidence-based interventions therefore have incorporated specific demands on the support for practitioners that carry out the intervention. The support systems of these evidence-based interventions, however, differ from each other. Specific knowledge as to what should be the content of a support system or on the standard minimum rules for effective support, is lacking. As Beidas and Kendall (2010, p.26) conclude in their review on the training of practitioners in evidence-based practice: *“Despite the importance of EBP, we know less than preferred regarding how to best train therapists in EBP”*.

4.1.1 Support systems

What do we know about (effective elements in) support systems that contribute to establishing and maintaining treatment integrity? Literature suggests a short training of practitioners is not sufficient and it takes ongoing supervision, clinical consultation, and support over an extended period of time to allow for adherent and competent delivery of an intervention (Addis, 2002; Connor-Smith & Weisz, 2003; Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). Research on the optimal focus of the training suggests that it is preferable to focus on intervention principles, rather than teaching the details of manuals (Abramowitz, 2006; Hayes, 2002; Miller, Yahne, Moyers, Martinez, & Pirritano, 2004). When it comes to ongoing support as in supervision, consultation or coaching, literature suggest that a focus on a practitioners own functioning is the most effective way to enable them to deliver the intervention (Berger & Kleine, 2013; Voogden & Kuyenhoven, 2010).

In order to gain skills, the training and support of practitioners should consist of direct, active learning methods (Cross, Matthieu, Cerel, & Knox, 2007) as research on effective learning methods suggest that the sole use of indirect, passive, learning methods has limited effects on adapting behaviour (El-Tannir, 2002). Distinctive for direct learning methods are the opportunities for practice and feedback, mostly acquired through modelling and role-play (Beidas & Kendall, 2010).

The above described provides a general indication of effective (elements of) support systems for practitioners. The main goal of this review is to provide more specific information on the content of support and provide practical recommendations for intervention developers in order to establish and maintain treatment integrity of their planned interventions. This will be done by selecting and analysing studies that properly measure treatment integrity and show a positive and sufficient effect of the support on levels of treatment integrity. We have formulated three questions:

1. *What type of support is provided in studies with high levels of treatment integrity?*
2. *What is the focus of the support and which learning methods are applied?*
3. *What are the characteristics of effective supporters?*

4.2 Method

4.2.1 Literature search procedures

To identify studies for this review we searched the following databases:

- Academic Search Premier
- ERIC
- PsychINFO
- Web of Science (social sciences citation index)

The search terms used are: (*adherence OR competence OR integrity OR fidelity OR dissemination OR implementation OR efficacy*) AND (*evidence-based practice OR EBP OR research-based OR empirically supported*) AND (*training OR workshop OR supervision OR education OR feedback*) All databases are searched on abstract level.

4.2.2 Inclusion criteria

Included studies;

- Evaluate the effects of a support system for practitioners in an evidence-based intervention on dependent measure(s).
 - An intervention is considered evidence-based when it uses replicable procedures (codified in manuals) and has clear empirical support demonstrated in a randomized controlled trial (Kazdin & Weisz, 2003).
 - Dependent measures are self-reported and/or independently rated variables measures of adherence and competence⁷
- Are primary studies
- Are published in English, Dutch or German language.
- Are published between 2009 - February 2013. Herewith following up on the research of Beidas & Kendall (2010) that included articles published between 1990-2008.

4.2.3 Exclusion criteria

Studies are excluded if they;

- Have a purpose other than the evaluation of the effects of a support system of practitioners in an evidence-based intervention on dependent measure(s).
 - Including examination of mediators or moderators of therapeutic process, risk factors, cost effectiveness of the intervention, barriers to treatment implementation, characteristics of sample and treatment setting.
- Evaluated interventions that are not delivered by treatment agents (self-help therapies)
- Evaluated pharmacological interventions only

4.2.4 Study selection

Study selection has been done in a three step decision-making process; first step was the selection of studies that showed a positive relationship between the support system provided and the measurement of treatment integrity. This selection resulted in 27 articles covering 44 outcome studies on support conditions.

Second step was the rating of the 27 articles on the extent to which treatment integrity was operationalized in a proper way. To judge this, the ITIPS-A (Goense,

⁷The definitions of adherence and competence in the studies have to be equivalent to the definitions used in this review.

Boendermaker, van Yperen, Stams, & van Laar, 2014) was used. The ITIPS-A is an adapted version of the ITIPS developed by Perepletchikova (2006), for details on the adaptations made, see Goense et al. (2014). The ITIPS-A consists of 22 items, covering the domains of establishment, assessment, evaluation, and reporting of treatment integrity in outcome studies. Each item is rated on a 4-point scale. The ratings on the scale are based on multiple recommendations in the implementation literature (Goense et al., 2014) for a discussion of these in more detail, see Perepletchikova et al. (2007). The ITIPS-A has shown sufficient internal consistency on the domains of establishing (.66), assessing (.65) and evaluating (.64) and marginal consistency for the domain of reporting (.55) of treatment integrity and a sufficient interrater reliability (.633) in the study of Goense et al. (2014). Using the ITIPS-A the implementation of integrity procedures in outcome studies can be classified as inadequate, approaching adequacy, and adequate. Only studies that could be classified as adequate on the total score were included in this study. This selection resulted in 7 articles covering 13 outcome studies on support conditions.

Last step in the study selection procedure was a selection of the studies in which the treatment integrity scores after implementation of the support systems were considered sufficient by the authors of the articles. This resulted in 4 articles covering 6 outcomes studies.

4.2.5 Measure

Three trained coders used a manual to evaluate the elements of the support systems in the outcome studies on; the evidence-based intervention (key elements, intervention type, proof of effect), population in the outcome study, the practitioners, supervisors and consultants in the study (background, education etc), type of outcome study (RCT, N=1 etc), the training and/or supervision that was provided (duration, type of training elements, focus), and the results of the training on levels of treatment integrity.

4.3 Results

4.3.1 What type of support is provided in studies with a high level of treatment integrity?

All of the studies provide at least 2 types of support to the practitioners (table 1). The extend in which the support systems are provided in the studies differ. The most extensive training was provided in the three studies of Forgatch & DeGarmo (2011). Their training comprehended 21 workshop days. The other trainings in the studies are between 2 and 3,5 days. Two studies provided a booster session of one day. Two studies provided telephone coaching/ consultation. Supervision in all cases was provided in a group format.

Table 1
Type and extent of support systems in studies (N=6)

	Forgatch & DeGarmo (2011) Study 1	Forgatch & DeGarmo (2011) Study 2	Forgatch & DeGarmo (2011) Study 3	Hepner et al. (2011)	Godley et al. (2011)	Lu et al. (2012)
Intervention	10	PMTO	PMTO	CBT	ACRA	CBT
Training	21 days	21 days	21 days	2 days	3.5 days	2 days
Booster session				1 day		1 day
Supervision/ coaching/ consultation	Live group coaching, individual telephone consultation, monthly group coaching	See study 1	See study 1	Weekly group supervision 90-120 minutes	Biweekly telephone coaching calls	Weekly group supervision

4.3.2 What is the focus of the support and which learning methods are applied?

In table 2 the learning methods and content that are applied in the training are described. It shows that the training consist of a combination of indirect and direct learning (see section *support systems* for descriptions) methods with a little more indirect methods that are used. The focus of the trainings has been on; the theoretical foundation of the interventions, the techniques/elements to be applied in specific cases and key components and procedures.

Table 2

Learning methods and content applied in training

Indirect	Direct
Didactic presentation of information (workbooks, journal articles, chapters, materials)	Modelling of key components and procedures
Lectures on the theoretical foundation of the intervention	Role-play with feedback (on treatment integrity)
Directed reading of intervention manual	Practice in situations with increasing difficulty
Video demonstrations of techniques/elements/ treatment cases	Taking knowledge tests
Presentations	
Written feedback (on practice cases)	

Note. N=6.

The focus and the learning methods of the boostersessions are not clearly described in the studies. The study of Hepner, Hunter, Paddock, Zhou and Watkins (2011), only described that the one-day booster session addressed challenging issues that came up during the initial implementation of the intervention by the practitioners.

Information on the focus of the ongoing support (supervision/consultation/coaching) was provided in all of the studies. The focus was a combination of client progress, previous sessions and upcoming sessions. Table 3 shows the applied learning methods and content in the ongoing support.

Table 3*Learning methods and content applied in ongoing support*

Indirect	Direct
Written feedback on adherence and/or competence	Feedback client progress using monitoring data
	Reviewing audio-tapes
	Feedback/discussion on adherence and/or competence
	Discussion of challenges and potential solutions
	Role-plays to practice or demonstrate skills for implementing the program and addressing challenges

Note. N=6.

In the ongoing support many direct learning methods are applied. There is only one indirect learning method applied and that is the use of written feedback on levels of adherence and/or competence.

4.3.4 What are the characteristics of effective supporters?

Only four studies provide information on the supporters' characteristics. In the study of Godley, Garner, Smith, Meyers and Godley (2011) the trainers were the treatment developers. The characteristics in Forgatch and DeGarmo (2011) differ. In the first study the practitioners (G1) were trained by the intervention developer. In the second study the practitioners (G2) were trained by selected trainers that were themselves trained by the developer (G1). In the last study the practitioners were trained by G1 and G2.

All six studies provide some information on the characteristics of the supervisors. In two studies the supervisors were licensed clinical psychologists who all had experience in the intervention. In one study (Lu, Yanos, Gottlieb, Marcello Duva, Xie, Rosenberg, & Mueser, 2012) the supervisor in the time of the study was being trained in the intervention and treated one case herself. This supervisor was supported by an expert consultant. The coaches in the study of Godley et al. (2011) were model experts.

4.4 Discussion

The objective of this review was to provide specific information how to give content to support and thereby establish a framework for an effective support system directed at treatment integrity. This was done by analysing the support systems of studies that have properly measured treatment integrity and show a positive and sufficient effect of the support on levels of treatment integrity.

The results show that an effective support system consist of a minimum combination of training and ongoing support (supervision, consultation, coaching), preferably extended with booster sessions. These findings are congruent with the suggestions in the literature that a short training of practitioners (alone) is not sufficient to allow for adherent delivery of the intervention (Addis, 2002; Connor-Smith & Weisz, 2003; Fixsen et al., 2005; in Steinfeld, Coffman & Keyes, 2009).

The focus of the training was mostly on the theoretical foundation of the interventions and key component and procedures. Thereby the results support other research findings that suggest that an effective training has a focus on the (general) principles of the intervention, rather than teaching the details of manuals (Abramowitz, 2006; Hayes, 2002; Miller, Yahne, Moyers, Martinez, & Pirritano, 2004). When it comes to supervision, the results show it is recommended to focus on progress of clients, previous sessions and upcoming sessions.

The results show and effective training consist of a combination of indirect and direct learning methods. This is consistent with the literature that suggest that to gain knowledge can be achieved by indirect methods but to gain skills a combination with direct learning methods is necessary (Cross, Matthieu, Cerel, & Knox, 2007; El-Tannir, 2002; in Beidas & Kendall). The merely sole focus on direct learning methods in the ongoing support might be explained by the fact that when practitioners receive these forms of support, they have already sustained knowledge of the intervention in their prior training in the intervention and the focus shifts to the delivery of the intervention. The idea of shift of focus is supported by the results showing the distinction between the content of the direct learning methods in training and the content in the ongoing support. In the training the focus is on role-play, modelling and practice the skills and techniques of the intervention. In ongoing support feedback on client progress and levels of adherence and/or competence in the delivery of the intervention get more attention.

An interesting finding in the studies was on the characteristics of the persons delivering the support. The study of Forgatch and DeGarmo (2011) showed that all generations of trainers provided a training that showed a positive and sufficient effect on levels of treatment integrity. Looking at the characteristics of the supervisors one can conclude it is important a supervisor has experience in the

intervention but providing the intervention itself at the time of supervision does not seem necessary.

4.4.1 Limitations and future suggestions

This is the first review that has looked at the effects of support systems on treatment integrity while taking the operationalization of treatment integrity into account. Only studies that adequately operationalized treatment integrity procedures, and showed sufficient effects of the support on adherence levels were included. From the 7 articles and 13 outcome studies on support conditions that adequately implemented treatment integrity procedures, in only 4 articles covering 6 studies the support systems resulted in sufficient treatment adherence in the opinion of the authors of the article. This could be due to the fact that a lot of studies compared effects of different forms of support and therefore proved a lot of support ineffective.

However, 4 studies to draw an effective framework for a support system is very marginal. This review does however provide specific information on effective content of support (see table 2 and 3) which practitioners and intervention developers can use in order to establish and maintain treatment integrity of their planned interventions. It is important to research if the positive results of the studies can be replicated by providing similar support to practitioners of evidence-based interventions. This will not only extend the knowledge on how to best support practitioners in establishing and maintaining treatment integrity. More importantly, it has the potential to put a stop on the disappointing effects of evidence-based interventions in practice settings and result in the promising outcomes for clients they have proven in research settings. Given the influence of support systems on treatment integrity, as a general rule efficacy studies in the practice field should only be performed if support systems are working properly.

Acknowledgments

We thank Inge Busschers and Sanne Pronk for their contribution to the data collection and analyses.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research was made possible by Stichting Innovatie Alliantie (SIA), a Dutch public funding body for research done by Applied Universities in the Netherlands, in the form of a grant (2012–14–18P).

CHAPTER 5

Measuring Treatment Integrity: Use of and experience with measurements in child and youth care organizations

Under review (2nd round) at *Journal of Behavioral Health Services & Research* as:

Goense, P.B., Boendermaker, L., & van Yperen, T. (2016). Measuring Treatment Integrity: Use of and experience with measurements in child and youth care organizations.

Abstract

Performance feedback and supervision are essential to the adoption of evidence-based interventions with high treatment integrity in child and youth care organizations. Little is known about the use of treatment integrity measurements in these organizations. For this study, 12 interventions for children and young people in the Netherlands with externalizing behavioral problems were selected. For each intervention, an expert, two supervisors, and two therapists were approached for an interview. In total, 54 semi-structured interviews were conducted. The results show that almost all interventions used treatment integrity instruments (N=11, 91.7%). Only two used measurements for both QA procedures (certification and recertification) and supervision purposes. Therapists regard treatment integrity measurements as valuable when they are used for multiple purposes and feedback is provided. The results of this study suggest the feasibility of the use of measurements for multiple purposes. Collaborative action is required to develop instruments that effectively contribute to continuous improvement.

Keywords

Treatment integrity, instruments, quality assurance (QA), interventions, supervision

5.1 Introduction

Measuring treatment integrity is becoming an integral part of outcomes studies of evidence-based youth interventions (Fixsen & Ogden, 2014). This is not surprising, as high levels of treatment integrity are found to be related to positive outcomes for clients (Goense, Assink, Stams, Hovee, & Boendermaker, 2016a; Schoenwald, Chapman, Sheidow, & Carter, 2009a; Sexton, & Turner, 2010; Tennyson, 2009). Various instruments are used to measure levels of treatment integrity in these studies (Schoenwald & Garland, 2013). But as Schoenwald and Garland (2013) conclude in their review of treatment adherence measurement methods, “there is a gap that warrants bridging between adherence measurements methods devised for use primarily as independent variable checks in efficacy studies and those that can be used in diverse practice contexts.” (p. 154).

Little is known about the feasibility of the use of treatment integrity measurements in child and youth care organizations as part of quality assurance (QA) procedures, or as a tool to provide performance feedback to therapists. Details about the resources required for the implementation of integrity measurement methods are also rarely reported (Schoenwald & Garland, 2013).

This lack of knowledge is striking, since performance feedback is essential to the adoption of evidence-based interventions with high treatment integrity in child and youth care organizations. Ongoing supervision in which therapists’ interactions with clients are reviewed and discussed, is associated with treatment integrity (Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005; Garland & Schoenwald, 2013; Goense, Boendermaker, & van Yperen, 2015a; Hogue, Ozechowski, Robbins, & Waldron, 2013; Kerby, 2006; Miller, Sorensen, Selzer, & Brigham, 2006; Schoenwald, Sheidow, & Chapman, 2009b). The central question of this study was “*Are treatment integrity measurements used within child and youth care organizations, and if so, how and what are the experiences with the use of these types of measurements?*” This study provides information on the conditions that seem necessary to successfully implement this type of measurements in child and youth care organizations in order to improve the quality of service delivery to those who are in need.

5.1.1 Definition of treatment integrity

Treatment integrity encompasses two aspects: therapist adherence and therapist competence (Perepletchikova, Treat, & Kazdin, 2007). Therapist adherence can be described as the degree to which the therapist delivers the prescribed components of a specific intervention (i.e., the degree to which the delivery of an intervention is consistent with the intervention manual). Therapist competence is commonly described as the level of the therapist’s technical skills and judgment (timing and

appropriateness) in delivering the components of the intervention (Barber et al., 2006; Barber, Sharpless, Klostermann, & McCarthy, 2007a; Barber, Triffleman, & Marmar, 2007b; Goense, Boendermaker, van Yperen, Stams, & van Laar, 2014; Perepletchikova et al., 2007). McLeod and colleagues (2013b) divide competence into technical competence – namely delivering specific components of the intervention, such as the delivery of behavioral cognitive elements in interventions for youth with aggression problems – and common competence, that is, delivering common, non-specific components of interventions (e.g., alliance and creating positive expectancies) (McLeod, Southam-Gerow, Tully, Rodriguez, & Smith, 2013b).

5.1.2 Use and functions of integrity measurements

The use of treatment integrity measurements varies across contexts and stakeholders (Schoenwald et al., 2011). Table 1 outlines three types of usage. Now that the measurement of treatment integrity is becoming an integral part of outcomes studies, one of their uses is in efficacy and effectiveness studies. The aim of these measurements is to determine whether an intervention failed due to the failure of the intervention or components thereof, or due to the insufficient or inadequate application of the intervention (Schoenwald et al., 2011). Measurements of integrity in these studies can also provide detailed information on the challenge of disseminating evidence-based interventions to diverse minority populations. Instead of further extending mental health disparities by focusing efficacy studies only on majority populations, measuring treatment integrity in efficacy as well as effectiveness studies opens up opportunities to research how these interventions can be implemented to effectively address issues of minority populations (Nápoles, Santoyo-Olsson, & Stewart, 2013). In other words, if researchers want to know what works for whom, when, and why, they need to accurately measure the “what,” so that they can establish what works in a diverse population.

A second application lies in quality assurance. Intervention developers, purveyors, and stakeholders can use the ratings for certification or recertification purposes, and the organization can use them to train and support teams and individual therapists.

The third use of measurements are measurements used at the therapist and the team level. These measurements provide information on skills in delivering the interventions, and can be used to provide performance feedback to therapists. Feedback systems that focus on treatment integrity have the potential to create “learning organizations” when the data is used to make quality improvement decisions (Knox & Aspy, 2011; McLeod et al., 2013b).

Little is known about the assessment of integrity for the last two types of use. The aim of this study was to gather information on the utility of and experiences with integrity measurements for these two types of use.

Table 1

Possible uses, contents, and aims of measurements of treatment integrity

Use	Content	Aim
Efficacy and effectiveness studies	Information on relation between intervention/intervention elements and client outcomes	Information on elements strongly associated with the outcomes of the intervention
Intervention developer, purveyor, stakeholders	Evaluation of therapist- and agency-level clinical performance	Therapist and team certification/ recertification purposes; to design and adjust training and support; QA procedures
Therapist/team	Information on performance in delivering the intervention	Provide ongoing support to therapists to learn and develop skills/ performance feedback

5.1.3 The assessment of treatment integrity

Treatment integrity measurements vary widely in their format and informant source. The most common distinction is between direct and indirect measurements. Direct measurements, such as the rating of a videotaped session, are used to directly observe intervention delivery. Indirect methods are assessments made by therapists who rate their own integrity, assessments by other persons who rate what was done by a therapist (by means of interviews or questionnaires), or assessments of an assortment of products such as written assignments made by the therapist (Goense et al., 2014; Perepletchikova et al., 2007).

Compared to direct methods, the indirect methods are easier to use and more cost-effective. They are, however, susceptible to biases and distortions, because they are subject to the tendency to provide socially desirable answers and subjective recollections (Goense et al., 2014; Perepletchikova et al., 2007). The validity of therapist-reported ratings of integrity is disappointing, especially when compared to nonparticipant observational ratings (Hogue et al, 2013). Observational ratings by nonparticipant raters are the most rigorous method for rating integrity. Although this type of assessment is seen as the gold standard because it provides objective and highly specific information regarding therapist performance, it is

costly in terms of time and resources (Hogue et al., 2013, Schoenwald et al., 2011; Sutherland, McLeod, Conroy, & Cox, 2013).

To ensure the accuracy of the representation of the data obtained, data should be collected across intervention phases, situations, sessions, and clients, because these factors can influence the level of integrity of the therapist (Perepletchikova et al., 2007). Raters should also be trained in all of the major and minor intervention components, including the subtle aspects of the intervention and the intervention manual, in order to ensure adequate assessment of the integrity (Perepletchikova et al., 2007).

Instruments should assess the multiple components (adherence, and technical and common competences) of treatment integrity (Goense et al., 2014). Because therapists in child and youth care organizations might be experienced in delivering particular intervention methods acquired in previous therapeutic work, it is also recommended that measurements include “treatment differentiation,” (McLeod et al., 2013b) which is the degree to which the therapist delivers intervention methods that are not part of the specific intervention, and thus deviates from the planned intervention (Kazdin, 1994). In efficacy studies, measuring treatment differentiation is essential to determine additional intervention methods that attenuate or amplify intervention effects (McLeod et al., 2013b).

Instruments that use a Likert-type scale to assess the thoroughness and frequency of the delivery of the intervention are recommended (Sutherland et al., 2013). The breadth and depth of the delivery of an intervention component cannot be assessed with a dichotomous checklist. As Sutherland and colleagues (2013) quote McLeod, Islam, and Wheat (2013), “simply counting the frequency of the delivery of a component can misrepresent the therapeutic process by giving a higher weight to components that are used more often but may fairly weigh those used in a more thorough manner.” (p. 6).

There is growing enthusiasm about using research-based integrity instruments as QA tools in child and youth care organizations, on the premise that these instruments have strong psychometric properties (Schoenwald, 2011: in Hogue et al., 2013). However, systematic information on the use of psychometrically sound integrity instruments in child and youth care organizations is lacking. The gold standard for assessing treatment integrity discussed above, is a standard for measurements in a research context. The criteria are based on the need for valid and reliable measurements to objectively measure treatment integrity in efficacy trials. It is unknown whether this gold standard is also applicable to measurements that are to be used for different purposes. This study provides information about the use of and experiences with integrity measurements in child and youth care organizations, and discusses the conditions that seem necessary to successfully implement such measurements in these organizations.

5.2 Methods

5.2.1 Intervention selection

This study was part of a 30-month research project in the Netherlands, in which the content of the support systems of best-practice interventions was examined, in order to formulate tools for quality assurance in care as usual (Goense et al., 2015b). Here, a support system is defined as “the initial training, certification/re-certification, booster sessions, and ongoing support such as supervision.” Twelve interventions provided in the Netherlands for children and young people with externalizing behavioral problems were selected for this study (see Table 2 for an overview of the selected interventions). These particular interventions were selected for three reasons. First, they cover the whole spectrum of interventions for behavioral problems in children and young people in the Netherlands, and many organizations provide similar practice-based versions or apply parts of these interventions. Second, almost all of the interventions are evidence-based. In this context, evidence-based interventions are interventions that are theoretically based, well documented, protocolled, and structured, and are manualized and have gained empirical support in experimental or quasi-experimental research (Weisz, Jensen-Doss, & Hawley, 2006). Only Tools4U, TACt, and TACt-i have not gained such support. These interventions are adaptations of the evidence-based Washington State Aggression Replacement Training intervention, (Goldstein, Glick, Gibbs, 1998) and share the same key elements as this intervention. Because all interventions have a manual, the key elements are known, which makes it possible to measure treatment integrity. Third, these interventions are all provided in the Netherlands, and each works with more or less extensive QA procedures.

Table 2*Overview of the selected interventions*

Focus	Intervention name	Intervention type
Juvenile	Aggression Regulation Training (ART)	Group training
	Tools4U	Individual training
	Training Aggression Control (TACT)	Group training
	Training Aggression Control – Individual (TACT-i)	Individual training
Parent(s)	Triple-P	Group training
	Parent Management Training – Oregon (PMTO)	Individual training
	Incredible Years (IY)	Group training
	Parent–Child Interaction Therapy (PCIT)	Individual training
System / multi-system	Functional Family Therapy (FFT)	System therapy
	Multidimensional Family Therapy (MDFT)	Multi-system therapy
	Multi-system Therapy (MST)	Multi-system therapy
	Multidimensional Treatment Foster Care (MTFC)	Multi-system therapy

5.2.2 Participant characteristics

Dutch intervention experts, supervisors, and therapists employed by the licensed providers of the 12 selected interventions were eligible to participate in the study. The experts typically work at the national organization that holds the license to provide the interventions, and are responsible for the implementation and dissemination of the intervention within the country. They usually have dual roles, namely as team trainers and coaches, and as the supervisors' supervisor. The supervisors and therapists of these 12 interventions work in organizations for child and youth care that provide ambulatory, intramural, and extramural care to children, young people, and their families. Therapists provide parental skills training to parents, aggression replacement training to children and young people with externalizing behavior problems, or provide a system or multi-system intervention. See Table 2 for an overview of intervention focus (parent, juvenile, system/multi-system), name, and type.

5.2.3 Participant recruitment

Participants were recruited by purposive chain sampling between January and August 2013. The licensed providers of the interventions were contacted by the second author and asked to indicate (if possible) one expert, two supervisors, and two therapists who 1) could give a clear picture of the support system of their specific intervention, 2) had sufficient experience to have formed an opinion about the support system, and 3) would be willing and able to participate in the study. In order to have a minimum of at least two views on the support system, two supervisors and two therapists of each intervention were sought. Because of the limited number of intervention experts, the licensed providers were asked to indicate only one expert.

The persons indicated by the licensed provider were invited by the research team (by phone and email) to attend an interview about the support system of their interventions and their experience with this support system. Financial compensation was available upon request. All participants provided informed consent to be interviewed and audio-recorded.

5.2.4 Data collection

Qualitative data were collected through 60- to 120-minute semi-structured interviews held between September 2013 and January 2014. The lead author, two master's-level researchers, and one senior lecturer conducted the interviews using an interview protocol. The protocol was designed by the lead author and the second author to gather information on: 1) the content of the support system, 2) the reasons for implementing that specific type of support system, 3) the experiences with the support system, and 4) the conditions needed to successfully implement the support system (recommendations). Examples of guiding questions from the interview protocol related to the use of and experiences with treatment integrity instruments are provided in Figure 1. The interview protocol is available from the first author upon request. No demographic data (age, education, and years of experience) on the participants was obtained. Since supervisors and therapists were all licensed, and no hypotheses about possible differences due to educational background or experience were subjects of this study, data on demographic characteristics were considered not relevant.

The first interview conducted by the lead author was used to train the data collectors (two master's-level researchers and one senior lecturer) to use the interview protocol and prepare them for the participants' possible answers. In addition, each interviewer also collected intervention manuals and other written material related to the interventions in order to obtain information on the support systems before conducting the interviews. The data collectors were supervised

by the lead and second authors. During the interview period, monthly meetings were held, during which transcribed interviews were used to discuss the progress and maintain uniformity in interviewing. Each interview was audio-recorded and transcribed verbatim.

Content of the support system: Does this intervention use instruments to measure treatment integrity? If so: What dimensions does the instrument measure? Does that relate to the skills that professionals need to deliver the intervention, and if so, how? For what purpose are the instruments used? How are the instruments used for each purpose? Is feedback provided on the ratings, and if so, how?

Reasons for implementing the specific type of support system: Why was this instrument/these instruments chosen? Why is it/are they being used for this/these purposes?

Experiences with the type of support system: What is your experience with the use of the instruments for this purpose/these purposes? What is your experience with/opinion of the type of instrument that is being used? What are your experiences with the use of the instrument(s)?

Conditions needed to successfully implement the support system (recommendations): Are the instruments being used as intended? If not, what makes it hard to use the instruments as prescribed?

Figure 1. Examples of guiding questions from the interview protocol

5.2.5 Data analysis

Qualitative analyses were conducted by the first author, the two researchers holding Master's degrees, and one junior researcher. MaxQDA 11 (2016) was used for coding and analyzing the data. To develop the codebook, directed content analysis was used. Content analysis is a qualitative research method that involves a systematic process of coding and identifying themes or patterns in text data (Hsieh, & Shannon, 2005). Directed content analysis is a specific content analysis approach that uses existing theory or prior research to develop the initial coding scheme prior to analyzing the data. Additional codes are developed as the analysis proceeds, and the initial coding scheme is revised and refined (Hsieh, & Shannon, 2005). The strength of a directed approach to content analysis is that existing theory or prior research findings can be supported or extended (Hsieh, & Shannon, 2005). In this study, prior research findings on the content of effective support systems to establish and maintain treatment integrity were used to develop the interview protocol (Goense et al., 2015a). For example, research findings

suggest that support systems should consist of at least a combination of training and ongoing support (supervision), preferably extended with booster sessions, and that the support should focus on, for example, providing feedback on levels of treatment integrity (Goense et al., 2015a). Basic codes (e.g., content of support system: training, certification, supervision, and recertification) were identified based on the interview protocol. The first two transcripts were coded jointly by the two researchers holding Master's degrees using these codes; sub-codes and new codes were generated when additional domains emerged. The remaining interviews were then coded independently by the first author, the two researchers holding Master's degrees, and the junior researcher.

The initial results were shared with the second author, dependability was accomplished through the use of a logbook (audit trail), and weekly meetings were held to increase the trustworthiness of the data. Trustworthiness in qualitative research refers to the methodological provisions that are implemented to ensure rigorous analytical procedures (Morrow, 2005). To achieve credibility, the participants received the transcribed interviews as well as the results, for a member check.

5.3 Results

5.3.1 Participants

In total, 54 face-to-face interviews were conducted between September 2013 and January 2014. Of the participants, 23 were therapists, 10 were experts (of whom three were also national project leaders), and 21 were supervisors. Seven participants who had the professional role of expert as well as supervisor within their intervention, were indicated in this study as supervisors. See Table 3 for an overview of the intended number of interviews and the actual number of interviews that were held per role description.

Table 3*Intended number and actual number of interviews*

Role	Intended number	Actual number	
		N	%
Intervention expert	12	10	83.3
Supervisor ^a	24	21	87.5
Therapist	24	23	95.8
<i>Total</i>	<i>60</i>	<i>54</i>	<i>90</i>

^a All of the persons who were both supervisors and experts (n=7) are indicated here as supervisors

All participants were recruited by purposive chain sampling through the licensed provider, and in the opinion of the licensed provider they had enough experience to have formed an opinion about the support system (see Table 4 for an overview of participants per intervention). Financial compensation was provided to participants representing one specific intervention, and three other participants added minor additional commentaries on their interview after the member check.

Table 4*Scale of interventions and overview of interviews*

Intervention	Scale of interventions in the Netherlands (in 2014)	No. of interviews
PMTO	1 national project leader/expert	1
	13 supervisors, of whom 8 also experts	2
	±100 therapists	2
Triple-P	2 national project leaders/experts	2
	22 regional/local experts	1
	Unknown number of people who provide supervision/coaching	1
	±1100 therapists	2
Incredible Years	2 experts/supervisors	2
	±40 therapists	3
PCIT	2 experts	1
	± 26 therapists	1

Table 4 (Continued)

Intervention	Scale of interventions in the Netherlands (in 2014)	No. of interviews
ART	1 expert	1
	3–5 supervisors	2
	±250 (estimated) therapists	2
Tools4U	1 expert	1
	5 supervisors	2
	±44 therapists	2
TACT	2 experts/supervisors	2
	± 30 therapists who deliver TACT within child and youth care organizations	1
	±5 therapists who deliver TACT nationally as an alternative to detention	1
TACT-i	2 experts/supervisors	1
	± 30 therapists	3
FFT	2 experts	1
	3 supervisors	2
	± 50–70 therapists	2
MST	5 experts	1
	28 supervisors	2
	±30 teams each with on average 4 therapists and 1 supervisor (total ± 120 therapists)	2
MDFT	9 experts	1
	40 supervisors (1 per team)	2
	40 teams each with 4–5 therapists (total ±160 therapists)	2
MTFC	1 expert/supervisor	1
	2 supervisors	2
	2 teams each with 4–5 therapists who (together with the educational parents) deliver MTFC	0
<i>Total</i>		<i>54</i>

5.3.2 Overall use of treatment integrity instruments

All interventions except TACT (N=11, 91.7%) use treatment integrity instruments (see Table 5). From here on, we refer only to the 11 interventions that use instruments. Approximately half (N=6, 54.6%) of these interventions use measurements of treatment integrity for certification purposes, a third (N=4, 36.4%) do so for recertification, and over half (N=8, 72.7%) do so for supervision purposes.

Over half of the interventions (n=7, 58.3%) apply treatment integrity instruments for one purpose only, that is, certification (27.3%, Triple p, IY, PCIT), recertification (9.1%, Tools4U), or supervision (27.3%, ART, FFT, MST). Four interventions (36.4%) use measurements for more than one purpose, and two (PMTO and MDFT) use measurements for all purposes (see Table 5). Most ratings (N=7, 63.6%) are based on videotapes of live sessions with clients.

Table 5

Use of treatment integrity measurements

Intervention	Certification	Recertification	Supervision	Ratings based on
ART			✓	Live observation
Tools4U ^a		✓		Videotapes
TACT				
TACT-i ^a		✓	✓	Videotapes
Triple-P	✓			Role-play
PMTO	✓	✓	✓	Videotapes
Incredible Years	✓			Videotapes
PCIT	✓			Videotapes
FFT			✓	Process notes
MDFT	✓	✓	✓	Videotapes or live observation
MST			✓	Parent ratings
MTFC	✓		✓	Videotapes

^a Recertification for these interventions refers to a quality control during the delivery of the intervention (since the therapists have not been certified, they cannot be recertified)

5.3.3 Treatment integrity instruments for certification purposes

Use of instruments

Six interventions use instruments for the certification of therapists; none of these interventions focuses solely on juveniles. All but one of these instruments are rated by means of videotapes of sessions with clients (see Table 5). The instruments are rated by an external rater. These are usually the leading experts of the intervention, most of whom have been trained in rating videotapes of sessions (see Table 6).

Requirements concerning videotapes

The various interventions require differing numbers of tapes to be submitted for certification. The number required ranges from one tape of one session with a client, through four tapes of two families (two tapes per family) and eight tapes of two families (four tapes per family), to tapes of two groups of parents from the first session through to the last session (24 tapes in total). Three interventions have specific requirements concerning the kind of tapes that have to be submitted. The requirements generally include that tapes should include sessions in which certain topics or elements are delivered, or in which specific techniques are practiced (such as role-play with parents), or that tapes of different clients or training groups should be submitted (see Table 6). The reason for specific requirements concerning the type of tapes was explained by a supervisor: *“Because those are the core elements: it is the core that ‘makes’ the intervention. In practice settings most attention is also paid to those elements. They are the main topics of the intervention and are the effective components.”*

Table 6

Overview of requirements concerning videotapes and the raters of tapes for certification

Intervention	Number of tapes	Content of tapes	Rater
PMTO	4	Four tapes of two families (two tapes per family). Of each family there has to be a tape in which the item “encouragement” is explained, and a tape in which the parents reflect with the therapist on their implementation of the techniques. Two sessions on the item “setting boundaries” also have to be submitted. A role-play with parents on this subject must be included in the tape.	The rating is done according to the Fidelity of Implementation Rating System (FIMP). To become an FIMP rater and assess therapists for certification, one needs to follow a training and sit an exam.
Incredible Years	24	Tapes of two groups of parents from the first session through to the last session have to be submitted. 50% of parents must attend the entire training. The therapists have to provide a translation of documents and a translated transcript of the tapes and send everything in one package to the raters in the country of origin.	The experts of Incredible Years in the Netherlands assess the tapes and provide a written assessment in English.
PCIT	8	Eight tapes of two families (four tapes per family) have to be submitted. The tapes must include instruction sessions and coaching sessions of both the child-oriented and the parent-oriented phase. Therapists use a fixed format to record their reflections on their tape. They submit that reflection report including questions about their recordings and their own assessment.	The experts rate the tapes and provide feedback according to a fixed format.
MDFT	1	Not specified	Licensed provider rates the tape and provides written feedback.

Table 6 (Continued)

Intervention	Number of tapes	Content of tapes	Rater
MTFC	1	Not specified	Experts (also referred to as consultants) rate the tapes. Experts at different locations rate the same tapes on a regular basis to compare their scores and maintain uniformity in ratings.

Experiences with videotapes

The requirements concerning the content of the tapes prevent practitioners from selecting only videotapes of sessions they feel comfortable with. It can, however, also be impractical, because the timing of the recording can be too early in the process/stage of the intervention that therapists are in with their clients. Requirements concerning tapes also carry the risk that therapists work in a less client-oriented manner because they have to focus on getting a good videotape. One of the therapists said in this context: *“You have to record a good session. Sometimes I think about just doing the session with the client again, because I need to hand in a tape of that session. That, of course, is not the intention at all.”* When certification depends on only one tape, therapists find this stressful and impractical, especially when the tapes are used for this purpose only. See Figure 2 for quotations regarding experiences of certification with one tape.

One supervisor and two therapists of a specific intervention were negative about the instrument that is used to rate the tapes. In their opinion, the instrument focusses only on treatment adherence (and not competence). Both therapists and the supervisor commented that sessions in which not all elements are delivered can be very important and helpful to a client, but will score low on adherence. The supervisor stated: *“It rates adherence, which means it rates whether the elements you deliver in a session are the elements that are the core components of the intervention. I don’t always agree with this way of rating a session, because delivering all of those components does not necessarily mean it was a good session. The components should be addressed across the whole intervention program/training and I think it is quite unclear how that is taken into account.”*

Videotaping is not standard in each session. Therefore, making a tape for certification creates a lot of stress. For example, about how to work with the camera and the session that you should be taping. Therapists who are used to videotaping their sessions, have much less of a problem with it. (Supervisor)

No, it is just a snapshot. If I have one session that is the best I have ever had in my life, but for the rest I am horrible, they cannot judge that. (Therapist)

To get a good rating, I should discuss drugs, even with a young person who clearly stays away from drugs, so to speak. They like it when you discuss school and their relationship with their parents. Of course, it is not the intention of the certification that you discuss issues for reasons of certification. You just have to tape a session as it is. (Therapist)

Figure 2. Quotations regarding the experience of certification with only one tape

Experiences with feedback

One of the supervisors of an intervention where written feedback is provided on the videotapes reported finding this very valuable feedback, whereas the two therapists were less enthusiastic. The feedback is aimed at the static delivery of the intervention components, whereas the therapists said that many emotions are at play in a session and no feedback is provided on handling them. According to the therapists, the learning effects of this type of feedback are also minimal, because the feedback is provided by someone who does not know the therapists (or their work). A therapist of another intervention reported being very satisfied with the feedback that is provided: *“They provide you with an extended report. They have really put effort into it. That’s very nice. It’s a lot of work, but you get a lot of attention to what you have done.”*

Recommendations

Two therapists made specific recommendations concerning the process of certification based on treatment integrity. One recommended the use of live sessions rather than videotapes for certification: *“Well, I have thought about that. If I were the licensed provider, I would say we are not using videotapes anymore because you can mess around with them and you can put on a show that doesn’t cover the load. I would say go visit a live session as an examiner and see how that works. Then you have it pure and are part of the moments. Maybe even with a briefing on where are you with this family, what you have planned for this session. It fits with the intervention principles, what you are planning to*

do.” Another therapist recommended paying specific attention to the timing of the feedback when videotapes are submitted: *“It’s important that if you submit video material and you get feedback, that this feedback is provided before you start a new case. Because when you have already started and you get the feedback 3, 4 months later, it is not of much use anymore.”*

5.3.4 Treatment integrity instruments for recertification purposes

Use of instruments

Four interventions use treatment integrity instruments to recertify therapists. These interventions cover all the areas of focus (juvenile, parent, system/multi-system) of the interventions within this study. All instruments are rated by means of videotapes of sessions with clients (see Table 5). In all cases, the ratings are made by an external rater. This is usually the leading expert of the intervention, and most of the time the expert has been trained in rating videotapes (see Table 7). One therapists said in that respect: *“I find it logical that the licensed provider is responsible for the recertification, since they are the training institute. They set the requirements so they also have to check if you still fulfill the requirements. I think that of course my supervisor could also do that, but I think it’s a good thing it is done by an external institute. Because yes, the supervisor looks at it with a different eye and probably also with different interests.”*

Requirements concerning videotapes

When tapes are used for the recertification of therapists, therapists have to submit between one and three tapes every one or two years. Almost all interventions require the taping of certain types of sessions or topics. The licensed provider of one intervention suggest a number of topics from which a therapist can choose. For another intervention the supervisor has to approve the tape that is submitted (see Table 7).

Table 7

Overview of requirements concerning videotapes and the raters of tapes for recertification

Intervention	Number of tapes	Content of tapes	Rater
Tools4U	2 a year	The topics that have to be taped are determined each year. The tapes that are submitted need the supervisor's approval.	Licensed provider
TACT-i	3 a year	A tape has to be submitted of each of the intervention's topics (3 in total). The sessions that have to be taped are indicated. The therapist and supervisor both rate the tapes and submit their ratings.	Licensed provider
PMTO	1 a year	The lead educators provide a few topics from which the therapist can choose. The therapists tapes an entire session in which the topic is present.	The rating is done by an FIMP rater.
MDFT	1 every 2 years	Not specified	Licensed provider rates the tape and provides written feedback

Almost all supervisors and therapists agreed that recertification is necessary. A therapist said: *"I find it unbelievable that a lot of interventions are 'evidence-based' and have no follow-up or control system. I wonder how you can sustain the quality of your intervention that way. It means that everyone who can learn a little can do whatever in the name of an official evidence-based intervention without them even delivering the intervention itself."* When there are no recertification procedures, supervisors can lose track of the delivery of the intervention by their therapists. A supervisor: *"They have their certification, but after that it actually really begins. So we do want to be able to check that. We are now working on that. It's a simple fact that therapists very quickly drift off the intervention a bit."*

Experiences with videotapes

Although therapists were positive about the fact that recertification is part of their intervention, making a videotape can be stressful, impractical, and less client-oriented, especially when recertification is based on a single measurement moment (see Figure 3). As one therapist commented: *“It’s very exciting and it feels like all or nothing when you hand in your tape. That’s also how it’s being stated. You have to submit your best material.”* Therapists do not particularly like it when videotapes are used for recertification purposes only. One therapist suggested that it would be better to work with a learning curve that can be measured over a longer period: *“We’re practitioners, not movie makers. It would make it easier if you can look at tapes together or get the opportunity to learn from your tapes individually. Then they would know what kind of people they have working for them, and the measurements would be used not only for assessment, but also for learning. You can’t assess people on just one tape. You have to follow their practice on multiple aspects.”*

The moment of recertification can also come too early for therapists, especially if they have not provided a sufficient number of training sessions since their certification. As one therapist explained: *“I also know that there are many trainers who have only provided one or two training sessions this year. It’s actually almost not feasible that way. Because that means that first you’re still getting used to it, especially if you’re just starting out, okay, how do I do it all? And then, oh no, you’re already required to film a specific topic.”*

Intervention where videotape is used only for recertification

But you see that people like to attend supervision and booster sessions, because that is receiving information and the threshold is not too high. But they think of those tapes as simply too exciting, they think they won’t make it and they stop. (Supervisor)

It is a judgment, an exam. I can fail it and lose my license and then I lose my job, so they were like “aarghh!” (Supervisor)

The only thing that bothers me is that you get an assignment you have to commit to. So, for example, I have to tape a certain training element. Say, for instance, that the tape has gone awry. That means I can’t use it. So the next training session I have to do that training element again, while I actually didn’t want to do that training element. I find that annoying. (Therapist)

Figure 3. Quotations on videotapes used solely for recertification

Experiences with feedback

One supervisor commented on the experience with the provision of feedback on the ratings. This supervisor stated that the main purpose of the feedback is for therapists to also use the rating as a feedback moment, and not just for recertification purposes: *“And what you also have is that you give very extensive feedback, so really three sheets of A4 on domains that they know. [...] it really is feedback that they can use in the daily practice, and what is really important is that you make it personal. It’s not a list that rolls out of a system, but rather like, ‘Very nice, Peter, what you do there; you really see the children start shining.’ So provide concrete examples that really acknowledge and enlarge the personality of the therapist.”* One therapist who received feedback after a few weeks said about the timing of the feedback: *“It’s most comfortable when you sit with the rater. Then you can see everything and you can discuss it straight away. Or, for example, use livestreaming. Direct feedback is better than after a few weeks. The feedback really should be sooner.”*

Recommendations

The supervisor of an intervention that uses videotapes only for certification procedures, recommended the structural use of videotapes: *“And I think it would be good to do this very structurally. We could put more effort into that; it would make the annual rating less exciting because you’re used to doing it. Then you’d see that people don’t find it problematic anymore; the barrier would simply be low.”* The therapist who recommended using live sessions for certification also said that live sessions are better than videotapes for recertification: *“No, use a live session, for example. Yes, that is always exciting. [...] But knowing us, we’ll be open to that and understand the higher purpose, which is that we want to do our job the best we can.”*

5.3.5 Treatment integrity instruments for supervision purposes

Use of instruments

Seven interventions use information derived from instruments that measure treatment integrity during supervision. All three types of interventions (juvenile oriented, parent oriented, system/multi-system oriented) use these instruments (see Table 5). The main reason for using this information in supervision, is to provide feedback to therapists and to strengthen their professional competences. A therapist said: *“The aim of supervision is to get feedback on the delivery of the intervention, to see if you’re still doing it [the intervention] right. When nobody*

watches you, you can lose track.” A supervisor said: “The aim is to strengthen the professional’s competences, as well to improve him or her. There must be an opportunity to practice or to do some finishing touches, so that eventually it is delivered as intended: to strengthen the professional.”

Indirect ratings

Two interventions use indirect instruments in supervision. For one of these interventions, information on the session is provided by the therapist after each session with the client, and the supervisor collects this material to assess the delivery of the intervention. In this case, some supervisors make therapists also rate themselves in order to compare scores: *“I also let therapists rate themselves. Then we compare the ratings and we discuss whether or not the therapist recognizes the ratings that are made.”*

The second of the two interventions asks the clients for their opinion on the treatment integrity of the therapist. Each month, the clients are called and asked to rate the delivery of the core components of the intervention by their therapist on a 5-point Likert scale. During supervision, the supervisor and the therapist reflect on the data that have been collected over a certain period of time. A supervisor explained: *“The aim is not to look at the client level, but to look at the profile of the therapist. What are your strengths and how can you develop?”*

Direct ratings

Five interventions use direct instruments. Two of the five use these instruments to rate live sessions; the other three use them to rate videotapes. In the supervision of one intervention, the rating of videotapes is standard. The ratings can be done by the supervisor, by the professional, or by colleagues. When ratings are done by colleagues, the setting is usually a team session during which different colleagues rate the different components (elements of the instrument) that can be seen on the tape. Colleagues with a lot of experience in the intervention can sometimes rate multiple components or dimensions at the same time. Sometimes another component is chosen every supervision session, in order to rate all the components of the intervention at least once.

In individual supervision, there is the opportunity for therapists to choose a fragment of tape they have a personal question about, or supervisors can select a component that is not being delivered quite well enough and ask to show that on tape. Not everyone thought that watching a fragment chosen by a therapist is the best way to assess that therapist’s quality of delivery. A therapist commented: *“Selecting ten minutes you’re satisfied with and bringing that in to supervision, I think for an evidence-based intervention is not good. You should look at the*

session from beginning to end, in order to be able to see good and less good practices.”

Experiences with the use of instruments in supervision

A supervisor said the following about working with videotape ratings in supervision: *“That you just look with one another at tapes to see what someone is doing. And using such an instrument brings a certain focus. And working this way, I find those instruments very important to let therapists develop themselves.”* And a therapist said: *“I find it very informative. If I think of how my job would be if all of that didn’t happen, if I couldn’t zoom in on the details. You can do that when you have those tapes, because with those tapes you can look at details. You can see what you’ve done. What you’ve implemented and what the parents’ reaction is.”*

By using an instrument there is tangible information to start a conversation on the delivery of the intervention during supervision. The supervisor and therapists of one of the interventions have adjusted the instrument that was used to rate videotapes. The main reason was that the original instrument did not offer enough space for the therapist to reflect and was not useful in discussing the videotapes. The supervisor: *“We developed a new instrument because the old instrument was inadequate to discuss videotapes. You didn’t have a time indication or room for the therapists themselves to reflect. Such an instrument is necessary to discuss a tape well. Sometimes you see things that should be done differently in the tape, while a trainer later recovers. [...] Or when a trainer makes a less convenient choice, I find it important that the trainer sees this and is able to say ‘Next time I’ll do it differently.’ The previous instruments didn’t have room for that. Time indication is also relevant, because you can then see how long something has lasted.”*

Experiences with feedback

Therapists find the use of integrity instruments as feedback instruments in supervision very important: *“I also find it very important that feedback is provided. After all, you’re alone with the juvenile. I found filming to be very valuable, because you get feedback on what you normally don’t see.”* Another therapist mentioned how important it is to distinguish between using an instrument for feedback and using it to assess functioning: *“I think if an instrument focused on how you’re functioning and you’d also be judged if something went wrong, it would create enormous pressure. [...] If you know: end of the month they check whether we do it quite right, then it would be especially exciting what comes out*

and I just hope I did things well enough. So I'm not sure that necessarily benefits the quality."

Preferred type of feedback

Supervisors and therapists mentioned that the content of feedback that therapists prefer differs between junior and senior therapists. Therapists who have just started to deliver an intervention prefer feedback on adherence, whereas experienced therapists prefer feedback on common competences. A supervisor said: *"Starting therapists are completely occupied with implementing the intervention and adhering to the manual as much as possible. Their supervision and their questions are different [from those of experienced therapists]. We work on implementing the intervention with as much integrity as possible."* A therapist commented: *"When you start, the challenge is to follow the program as much as possible."* More experienced therapists have fewer questions about the application of the intervention, but more questions about the common competences required to deliver it: *"What's essential are group dynamics and your own actions: reflection on what is happening. The content [of the intervention] remains the same. You might have a question about it once and then you know it. But how do you ensure that it gets picked up? What do you do with yourself and the juvenile? That kind of question always remains, even if you master the intervention."*

5.4 Discussion

Little is known about the use of, or the feasibility of using, treatment integrity measurements in child and youth care organizations. The aim of this study was to provide information on the use of and experiences with integrity measurements in child and youth care organizations, and to discuss the conditions that seem necessary to successfully implement such measurements in these organizations. The results show that treatment integrity instruments are used for multiple purposes within child and youth care organizations. They are used as part of QA procedures (for certification and/or recertification) and to provide performance feedback to therapists (for supervision purposes). However, only four (36.4%) of the 11 interventions studied use measurements for more than one purpose.

For both certification and recertification, the vast majority of instruments are rated by means of videotapes of sessions with clients, with requirements concerning the type of sessions or topics to be taped and submitted. These requirements are in accordance with the implementation literature, which states that to ensure the accuracy of the representation of the obtained data, data should be collected across treatment phases, situations, sessions, and clients (Perepletchikova et

al., 2007). However, according to some therapists in this study sample, these requirements, which are formulated for research standards, are impractical. The therapists said that the timing of the recording is not always in line with the phase of the intervention they are in with their client.

For both certification and recertification, the ratings are done by an external rater, often the licensed provider of the intervention, who is trained in rating videotapes. The assessment of treatment integrity for this part of the QA procedure measures up to the gold standard of measurements for research purposes, since observational ratings by nonparticipant raters are seen as the most rigorous method for rating integrity, and the raters are trained (Hogue et al., 2013; Perepletchikova et al., 2007). However, the participants in this study whose videotapes are used for certification and/or recertification purposes only, find submitting videotapes stressful, impractical, and less client-oriented. In this respect, the therapists' suggestion to use the tapes to show a learning curve throughout a longer period of time should be valued.

The use of the instruments for supervision purposes is different from their use for certification and recertification. That integrity instruments are used for supervision purposes (N=7, 36.6%) is promising, since studies show that ongoing supervision in which therapists practice with clients are reviewed and discussed, is associated with high treatment integrity (Fixsen et al., 2005; Garland & Schoenwald, 2013; Goense et al., 2015a; Hogue et al., 2013; Kerby, 2006; Miller et al., 2006; Schoenwald et al., 2009b).

Instruments used only for supervision purposes are not always rated by means of videotapes of sessions with clients; instead, some ratings are based on live observations, process notes, or parents' ratings. The use of a direct assessment instrument (videotape ratings) in every supervision session is standard in only one intervention. This intervention, in which reviewing tapes is "normalized" in supervision, also uses videotapes for certification and recertification.

Standardizing the use of videotapes would overcome the practical problems that therapists face in meeting the requirements concerning sessions and topics, since they would have more material to choose from. This study indicates that the standard use of videotape ratings in supervision is not only possible, but also regarded as positive by both therapists and supervisors. For the gold standard of assessment of integrity to be applicable to both research and practice, it is recommended that videotape ratings are standard in QA procedures, and that these procedures always include supervision.

Many therapists and supervisors in this study stated that it is important to get the opportunity to learn. Therefore, integrity measurements should be used not only for assessment, but also for ongoing learning. Feedback on the assessment of treatment integrity is considered essential by both therapists and supervisors.

The implementation literature recommends the use of a Likert scale over a dichotomous checklist in order to be able to consider the breadth and depth of the delivery (Sutherland et al., 2013). The results of this study show that room for written reflection is important for the adoption of instruments in practice contexts. The need for feedback is also shown by studies that found that performance feedback is essential to the adoption of evidence-based interventions with high treatment integrity in child and youth care organizations (Fixsen et al., 2005, Miller et al., 2006; in Hogue et al., 2013).

According to participants in this study, therapists who are new to an intervention prefer feedback on adherence, whereas their more experienced colleagues prefer feedback on common competences. Supervisors who are responsible for feedback should take this into account in supervision sessions. Whether a relation exists between these preferences and the ability to differentiate between treatment methods acquired in previous therapeutic work and adherence, was not explored in this study. Given the importance of treatment differentiation to determine whether additional intervention methods are related to intervention effects (McLeod et al., 2013b), this is well worth investigating.

In efficacy and effectiveness studies, it is essential to distinguish between adherence, competence, and differentiation in order to understand how these components interact and influence intervention outcomes in both majority and minority populations (Goense et al., 2016a). This study shows that including components of treatment integrity (adherence and competence) in the instrument also corresponds to the preferences of professionals.

The main conclusion of this study is that therapists regard the use of treatment integrity measurements in child and youth care organizations as valuable and worth the time investment, but only when certain conditions are met. One of the most important of these conditions is that the measurements are used for multiple purposes and that feedback is provided. The implication of this finding is that the skills needed to provide constructive feedback based on instruments, needs further examination.

5.4.1 Limitations and recommendations for future research

Although videotape ratings by non-participant raters are seen as the gold standard for rating integrity in efficacy studies (Hogue et al, 2013, Schoenwald et al., 2011; Sutherland et al., 2013), it is unknown whether feedback to therapists based on videotapes leads to higher treatment integrity, compared to feedback based on indirect ratings. More research is needed to support the claim that the use of videotapes is the gold standard for boosting treatment integrity. However, solely

using videotapes will not lead to a better understanding of which intervention methods are regarded as most relevant by the recipients of interventions. This can be especially relevant when the interventions are used in populations that are culturally different from those for which they were developed, which in certain areas may happen regularly in day-to-day practice. It is recommended that efficacy and effectiveness studies also address the perception of the recipients, in order to understand how interventions can be adapted to suit culturally different populations.

It should be noted that a limited number of people per intervention were interviewed for this study, and due to the chain sampling procedure, selection bias is possible. Also no demographic data (age, education, and years of experience) of the participants was obtained. The purpose of this study was to gain a first insight into the experiences with and use of instruments in child and youth care organizations, and for that purpose the sampling was found sufficient by the authors. It does imply, however, that the findings in this study cannot be generalized to other populations or settings. More, and more systematic, research is needed, especially on the use of psychometrically sound instruments. Finally, this study did not address the use of and experiences with treatment integrity instruments for diverse populations. It remains a key challenge for future research to investigate whether, and if so, how integrity measurements should be adapted to support effective dissemination among diverse populations.

As mentioned, no demographic data on the participants was obtained, because there were no hypotheses about possible differences resulting from educational background or experience. Participants, however, mentioned that the content of feedback that therapists prefer differs between junior and senior therapists. This is an important topic for further study. To get the best outcomes, junior therapists and senior therapists may benefit from different QA systems.

5.5 Implications for behavioral health

The results of this study indicate that the gap between the measurement methods devised for research purposes and the preferred use of such measurements in child and youth care organizations as part of QA procedures and supervision, is bridgeable. The use of the same instruments would make it possible to collect more, and more comparable, integrity data that could function as a data loop on multiple levels. Information on the conditions that seem necessary to successfully implement treatment integrity measurements in youth care practice, may stimulate the use of these instruments and thereby improve the quality of service delivery to those who are in need. The aim of this article was to provide this information. Collaborative action is required to develop these instruments so that they effectively contribute to continuous improvement.

Acknowledgments

We thank Sanne Pronk, Rosan Bakker, Lianne Bertling, Inge Busschers, and Joyce Berentsen for their contribution to the data collection and analyses.

Funding

The research included in this article was financially supported by Stichting Innovatie Alliantie (a Dutch funding body for research conducted by applied universities in the Netherlands) in the form of a grant (2012-14-18P).

Compliance with Ethical Standards

Conflict of Interest

The authors declare no potential conflicts of interest with respect to the research, authorship, or publication of this article.

CHAPTER 6

Stimulating quality of service delivery in mental health care settings for youth. Suggestions for integrating support systems

Under review (2nd round) at *Administration and Policy in Mental Health and Mental Health Services Research* as:

Goense, P.B., Boendermaker, L., & van Yperen, T. (2016). Stimulating quality of service delivery in mental health care settings for youth. Suggestions for integrating support systems.

Abstract

Frequent and targeted support to professionals is essential to control for quality delivery of services. In youth care settings, there is limited time and capability to implement all of the support systems that are suggested by program developers. With the pressure and responsibility to provide services with high quality and low costs, organizations strive to effectively and efficiently integrate different support systems. In this point of view we discuss the potential of integrating support systems around overlapping common, contextual and structural factors of interventions that are delivered in youth care setting.

Keywords

Support Systems; Quality Delivery; Interventions; Factors

6.1 Introduction

Over the last few years it has become clear that the quality of the delivery of services to youth with externalizing behavioral problems and their families is associated with client outcomes (Goense, Assink, Stams, Boendermaker, & Hoeve, 2016a; Lipsey, 2009; Schoenwald, Chapman, Sheidow, & Carter, 2009a; Tennyson, 2009). Quality of delivery indicates that it is not only important that services are being delivered (adherence to the manuals) but also that services are skillfully, timely and appropriately delivered (competence in delivery) (Barber et al., 2006; Barber, Sharpless, Klostermann, & McCarthy, 2007a; Barber, Triffleman, & Marmar, 2007b; Mcleod, Southam-Gerow, Tully, Rodriguez, & Smith, 2013b; Perepletchikova, Treat, & Kazdin, 2007). Most mental health care settings for youth in the Netherlands, as probably elsewhere, typically provide a wide range of services for an equally wide range of target groups. They offer practice-based approaches in the guidance of juveniles and/or families, practice-based group training programs, as well as evidence-based interventions that mostly have been developed abroad. It is common for the professionals working in these settings to provide multiple interventions and/or training programs. Professionals in a semi-residential setting for example, offer a practice-based approach in the guidance of the juveniles on the group such as the solution-focused approach, and at the same time provide an evidence-based intervention, such as the Washington State Aggression Replacement Training (Goldstein, Glick, & Gibbs, 1998), to specific juveniles.

There is a growing number of studies that indicate that the quality of the delivery of services can be positively stimulated by providing frequent and targeted support to professionals (Hogue, Ozechowski, Robbins, & Waldron, 2013; Kerby, 2006; Mikolajczak, Stals, Fleuren, Wilde, & Paulussen, 2009; Miller, Sorensen, Selzer, & Brigham, 2006; Schoenwald, Sheidow, & Chapman, 2009b). For quality control, most evidence-based interventions have incorporated specific support systems for the professionals that deliver the intervention. The manuals of the practice-based approaches and training programs also increasingly describe which support is needed. However, in the youth care settings, there is limited time and capability to implement all of these support systems. With the pressure and responsibility to provide services with high quality and low costs, youth care settings want to effectively and efficiently integrate the support systems. The aim of this point of view is to discuss the potential of integrating support systems around overlapping factors of interventions⁸.

⁸ From this point onwards we will refer to interventions as a collective word for practice-based approaches, training programs and evidence-based interventions.

6.2 Factors of interventions

The most conventional distinction in intervention literature is between specific and common factors of interventions. Specific factors apply in the context of specific goals and target populations. Specific factors are based on existing theoretical knowledge on the risk and protective factors that play a role in the onset and persistence of the specific problem that the intervention intends to target. The specific factors can be differentiated in contextual factors (i.e. target population, theoretical framework), structural factors (i.e. length, duration, format) and therapeutic techniques, see Figure 1 (Blasé & Fixsen, 2013; van Yperen, Veerman, & Van den Berg, 2015). Chorpita and Daleiden, (2009) define a therapeutic technique⁹ as ‘a discrete clinical technique or strategy (e.g., time out, relaxation) used as part of a larger intervention plan’ (p. 569). Common factors contribute to client outcomes, regardless of the type of intervention and target population. Examples of common factors are the working alliance, creating positive expectancies and responsivity (see Figure 1) (Andrews & Bonta, 1994; McLeod et al., 2013b).

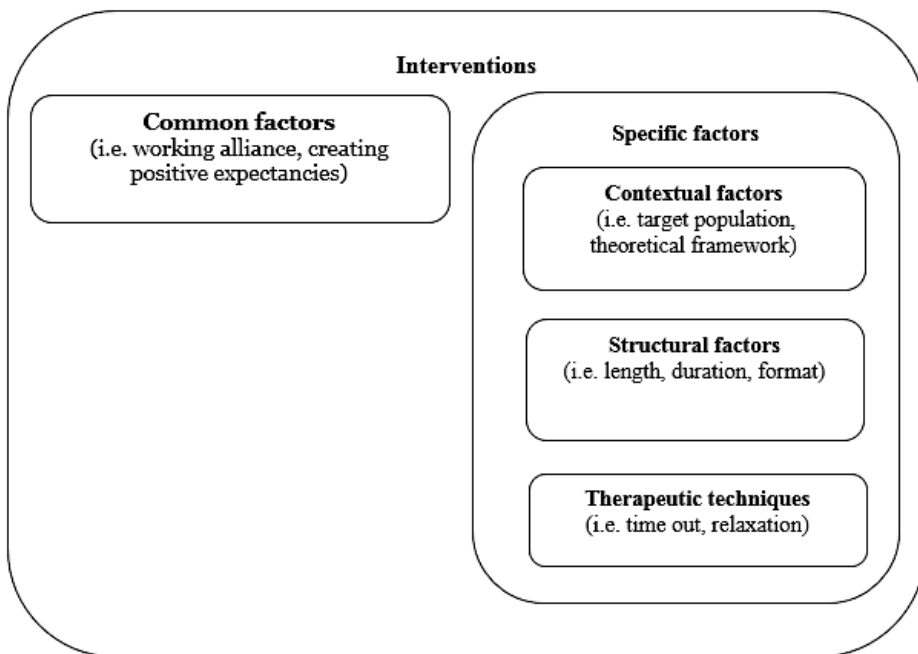


Figure 1. Factors of interventions

⁹ Chorpita and Daleiden (2009) refer to therapeutic techniques as practice elements.

All interventions have the common factors. Many interventions also tend to overlap on the contextual and structural factor level. For instance, many interventions target juveniles with externalizing behavior problems (contextual factor) and many interventions are provided in group format (structural factor). The common and overlapping factors of interventions could be a starting point for a solution to the accumulation problem of support systems.

6.3 Support systems targeting factors

There are multiple ways to define interventions and thus many potential criteria for identification of factors that overlap between interventions (Garland, Hawley, Brookman-Fraze, & Hurlburt, 2008). The identification of factors ranges between macro level analyses characterized by broad theoretical orientation of interventions, to micro level analyses of each verbal and nonverbal behavior that is part of the intervention (Garland et al., 2008). One of the difficulties the field is facing in its attempt to identify overlapping factors is the lack of knowledge about existing practices. As mentioned before, local providers do not solely provide evidence-based interventions but commonly provide a diverse set of care, programs and help to their clients. Although attempts are being made, there still is little research characterizing treatment process in usual care (Garland et al., 2008). This limits the ability of the field to identify overlapping factors that are applicable to local providers. At the same time, many interventions are proprietary of nature. In that light, it is important to discuss how realistic it would be for local providers (or anyone else in that respect) to develop support systems on the factors represented by these major interventions that are already providing such training and supervision.

First of all it is important to note that many local providers do not provide proprietary interventions (Barrett & Ollendick, 2004; Kazdin & Weisz, 2003; Weisz, Jensen, & McLeod, 2005). A number of reasons have been suggested for the nonuse of these interventions, including the expenses of training and supervision requirements (Weisz et al., 2005 In: Mazzucchelli & Sanders, 2010). For local providers, integrating support systems of proprietary interventions, is not possible yet. However, we think that the restrictions on training and supervision of proprietary interventions is untenable in the near future and it is important to anticipate on that. The emerging trend in the field towards the clustering of evidence-based interventions is an important indication to believe change is at hand. Many scholars are examining overlapping factors of interventions (i.e. Barth & Liggett-Creel, 2014; Chorpita, Daleiden, & Weisz, 2005; Garland et al., 2008). Some even have suggested to use the profiles of the clustered factors for training and supervision purposes of professionals (see Garland et al., 2008). The clustering of interventions is in line with the understanding that 'one size does

not fit all' (Stewart, Felleman, & Arger, 2015) and that flexible use of interventions is necessary to meet the needs of clients. If this line of thinking continues, we believe this will create opportunities to redesign the training and supervision of professionals in the field. Up to the point that local providers are involved in organizing the support around these interventions within their own settings.

An important guidepost for the integration of support systems, are the factors that most professionals have questions about or need support on, because they attain to the professionals' needs. Previous research suggests that some factors of interventions require competencies that are perceived by professionals as more difficult to learn and implement than others (Goense, Pronk, Boendermaker, Bakker, Ruitenber, & Bertling, 2015b). These competencies are related to a) common factors of interventions, that require general competencies of professionals and b) factors that require competencies (skills) of professionals in the contextual and structural level (see Figure 1). For example, the most difficult part of delivering group interventions (structural factor) is handling group interactions, keeping pace, motivating the participants to work together in role-play and respond to individual needs within the group context (Goense et al., 2015b).

We suggest youth care settings to focus the support of professionals around these factors. Start for instance on providing a training on handling group interactions and motivating participants to work together in role-play, for all professionals in your youth care setting that provide interventions in group format to juveniles with externalizing behavior problems. After provision of support to professionals on overlapping common, organizational and structural factors, integrating support around overlapping therapeutic techniques would be the next logical step.

6.4 Discussion

Youth care settings are confronted with a multitude of support systems and professionals are confronted with an accumulation of support sessions. Settings strive to effectively and efficiently integrate these different support systems. As a first step, we suggest to integrate support around overlapping common, contextual and structural factors of interventions. Integrating support systems based on competencies that professionals need for the delivery of multiple interventions serves two purposes. First, it enables youth care settings to organize support more efficiently. This meets the desire of youth care settings and professionals to prevent the stacking of support systems and sessions. Second, it lays a foundation for improved practice. More efficient organization of support to professionals around these factors increases the likelihood that the implementation of these factors are stimulated on a broad scale within a setting.

Integrating support systems on overlapping factors is a total new way of stimulating quality of services delivery which brings about new research tasks. It requires a joint effort of program developers, disseminators, youth care settings and researchers. The results of integrating support systems for the settings, professionals and client outcomes, needs to be examined. We think the integration of support systems on overlapping factors does more justice to the capabilities of youth care settings to provide frequent and targeted support to their professionals. With that, it is the first step in the right way of stimulating quality of delivery on a broad scale.

Compliance with Ethical Standards

Funding

There has been no funding for this study.

CHAPTER 7

General discussion

7.1 Introduction

Evidence concerning psychosocial interventions for children and young people with externalizing behavior problems has been amassed at an impressive pace in recent years (Southam-Gerow & Prinstein, 2014). Interventions that have been proven effective are now considered vehicles through which the knowledge of “what works” can be applied in practice. Outcomes for children, young people, and their families, however, have not improved in line with these advances in knowledge.¹⁰ This deficit has become known as the “implementation gap,” that is, the difference between the knowledge of “what works” and the application of this knowledge in real-life practice (Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005). The implementation gap raises the following questions about the application of interventions:

1. What does it mean to apply interventions – as vehicles of the knowledge of “what works” – and how is this operationalized in outcome studies?
2. Does the application of these interventions make a real difference to the end-users of the services?
3. What types of support for professionals can strengthen implementation processes?

This dissertation answered these three questions and presented knowledge of factors that contribute to bridging the implementation gap. In answering these questions, the focus was on whether professionals are delivering the interventions as intended. To realize this aim, five objectives were formulated, namely:

1. To examine the adequacy of the implementation of treatment integrity procedures in outcome studies of interventions targeting externalizing behavior problems of youth.
2. To examine the moderating effect of level of treatment integrity on the reduction of youth antisocial behavior after an intervention.
3. To examine the essential ingredients of support for youth care professionals to enable them to deliver the intended intervention with sufficient treatment integrity.
4. To examine the experiences and use of treatment integrity instruments within child and youth care organizations.
5. To devise a potential way to integrate support systems for professionals around overlapping factors of interventions.

¹⁰ A meta-analysis of RCTs that had tested youth evidence-based interventions in more clinically representative contexts, pitting them against usual care, showed a mean effect size of 0.29 for these interventions (Weisz et al., 2013a).

An overview of the research questions, objectives, corresponding chapters, and brief conclusions is presented in Table 7.1 (section 7.1.4).

7.2 Summary of the most important results

7.2.1 What does it mean to apply interventions – as vehicles of the knowledge of “what works” – and how is this operationalized in outcome studies?

In child and youth care, evidence-based interventions are considered vehicles through which the knowledge of “what works” can be applied in practice. These interventions target specific populations with specific problems. The interventions are theoretically based, well-documented,¹¹ protocolled, structured, and manualized, and have gained empirical support in experimental or quasi-experimental research (Weisz, Jensen-Doss, & Hawley, 2006). Applying the content of interventions as intended is referred to as treatment integrity (Perepletchikova, Treat, & Kazdin, 2007). The first objective of the research underlying this dissertation was to examine the operationalization of treatment integrity procedures in outcome studies of evidence-based interventions for children and young people with externalizing behavior problems. Treatment integrity procedures provide information about the establishment, assessment, evaluation, and reporting of treatment integrity.

The results of the systematic review of outcome studies that were performed prior to May 2012 (Chapter 2), show that treatment integrity was rarely assessed. The systematic search resulted in only 24 articles covering 29 studies that actually assessed treatment integrity. Almost all studies (80%) that were included in the review approached adequacy in implementing treatment integrity procedures. Therapist competence was reported far less often than therapist adherence. Many assessments of treatment integrity relied on self-reports by professionals, which precluded reports on the training of raters and indicators of interrater reliability. The findings were also limited in scope: Almost half (45%) of the studies were about the same intervention (multisystem therapy; MST), mostly examined by the same researchers. The present study shows that although measuring treatment integrity in outcome studies is found to be important, it is often missing or not examined under adequate circumstances. The lack of studies adequately assessing treatment integrity undermines the confidence we can have in statements made about the relationship between treatment integrity and intervention outcomes.

¹¹ “Well-documented” includes the documentation of clinical expertise and patient values with regard to the intervention, as evidence-based practice is the integration of best research evidence with clinical expertise and patient values (Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000, p 1).

Without adequate integrity measurements, the actual delivery of interventions remains unknown and no statements can be made about the relationship between treatment integrity and outcomes.

7.2.2 Does the application of these interventions make a real difference to the end-users of the services?

When the actual delivery of the intervention is measured adequately, statements can be made about the association between treatment integrity and client outcomes. Previous meta-analyses on the effects of interventions for juveniles exhibiting antisocial behavior, have suggested that delivering an intervention with a high level of integrity is associated with positive client outcomes (Lipsey, 2009; Tennyson, 2009). However, these meta-analyses did not take into account the quality of the treatment integrity procedures of the included studies. The second objective of this research was to meta-analytically examine, in a multilevel model, the moderating effect of level of treatment integrity on the reduction of youth antisocial behavior after an intervention. The operationalization of treatment integrity procedures of the primary studies was taken into account.

Treatment integrity was found to be a significant moderator of the reduction in client's antisocial behavior after an intervention (Chapter 3). A significant difference was found between the effect sizes of studies with a high level of treatment integrity and those with a low level of treatment integrity. Significant medium to large effects of evidence-based interventions were found when the level of treatment integrity was high ($d = 0.633$). Non-significant small/marginal effects were found when the integrity level was low ($d = 0.143$). In the study, a notably small number of studies ($k = 6, 35.3\%$) were found in which the intervention was implemented with high levels of treatment integrity.

Other significant moderators were study design, intervention type, and intervention duration. A multivariate analysis was performed to examine the effect of level of treatment integrity on client outcomes, adjusting for intervention characteristics (intervention type and intervention duration). The results showed that the association between high levels of treatment integrity and positive client outcomes stayed the same when controlled for the other significant moderators. This indicates the specific contribution of high levels of treatment integrity to client outcomes over and above the effect of intervention characteristics. The results of this study indicate that delivering evidence-based interventions with high treatment integrity is critical and should be stimulated.

7.2.3 What types of support for professionals can strengthen implementation processes?

Essential ingredients of support for child and youth care professionals

Delivering an intervention with a high level of integrity is a difficult task. One of the questions around the implementation gap is therefore: “What types of support for professionals can strengthen implementation processes?” Research suggests that providing professionals with frequent and targeted support is an effective way to establish and maintain treatment integrity (Kerby, 2006; Mikolajczak, Stals, Fleuren, Wilde & Paulussen, 2009; Schoenwald et al., 2009b). Specific knowledge of what the content of a support system should be, or of the standard minimum rules for effective support, was lacking. The third objective of the research underlying this dissertation was to extend the knowledge of how best to support professionals in establishing and maintaining treatment integrity in planned interventions.

The support systems of evidence-based interventions that showed a positive and sufficient effect of the support on levels of treatment integrity were analyzed in a systematic review (Chapter 4). The results indicated that an effective support system minimally consists of a combination of training and ongoing support (supervision, consultation, coaching), preferably extended with booster sessions. The findings are congruent with the suggestions in the literature that merely giving practitioners a short training course is not sufficient to enable adequate delivery of the intervention (Addis, 2002; Connor-Smith & Weisz, 2003; Fixsen et al., 2005; in Steinfeld, Coffman & Keyes, 2009). After initial training, professionals need continued support to ensure treatment integrity. The support should focus on, among other things, providing feedback on levels of treatment integrity, preferably through the use of direct instruments (such as video, audio, and/or live sessions).

Experiences and use of treatment integrity instruments within child and youth care organizations

Various instruments are used to measure levels of treatment integrity in outcome studies of evidence-based interventions (Schoenwald & Garland, 2013). However, little is known about the feasibility of the use of treatment integrity measurements in child and youth care organizations as part of quality assurance procedures, or as a tool to provide performance feedback to therapists. The fourth objective of this research was to gain knowledge of the use of treatment integrity

measurements within child and youth care organizations, and knowledge of the experiences with the use of these types of measurements. Given that we found that support should focus on, among other things, providing feedback on levels of treatment integrity (see Chapter 4), this objective became even more relevant.

The results of the qualitative study (Chapter 5) show that the 12 selected interventions use treatment integrity instruments for multiple purposes. Instruments are used as part of quality assurance procedures (for certification and/or recertification) and to provide performance feedback to therapists (for supervision purposes). The vast majority of instruments are rated by means of videotapes of sessions with clients, with requirements concerning the type of sessions or topics to be taped. The use of a direct assessment instrument (videotape ratings) in every supervision session was found to be standard in only one intervention (Parent Management Training – Oregon). The therapists of this intervention were positive about the use of the videotape ratings.

The present study showed that one of the most important conditions for professionals to perceive measurements as valuable, is that the instruments are used for multiple purposes. However, only 36.4% of the interventions included in this study used measurements for more than one purpose. The second condition to perceive measurements as valuable is that feedback to professionals is provided. Participants in the study mentioned that the content of feedback that professionals prefer differs between junior and senior professionals: Professionals who have just started to deliver the intervention prefer feedback on adherence, whereas their more experienced colleagues prefer feedback on competence.

The overall conclusion of the qualitative study (Chapter 5) is that instruments for the measurement of treatment integrity can be standardly used in supervision and for multiple purposes, and that both therapists and supervisors prefer using the instruments that way. At the same time, the results show that almost none of the interventions standardly use the instruments in supervision or for multiple purposes. Combined with the results presented in Chapter 4, it is clear that much action needs to be undertaken in the provision of support to these professionals. There is a need for structural direct learning (by using videotapes) in supervision, instruments need to be used for multiple purposes, and there needs to be differentiation in the type of feedback provided to professionals. The implication of this is that the skills needed to provide constructive feedback based on instruments, needs further examination.

A potential way to integrate support systems for professionals around overlapping factors of interventions

The most crucial aspect in the attempt to bridge the implementation gap is to provide knowledge on possible ways to deal with the limitations present in child and youth care organizations. It is insufficient to provide information on what support is needed for professionals without knowing what support is realistic for professionals. The question becomes: “What types of support for professionals can strengthen implementation processes *and* be adopted in real-life child and youth care organizations?” One of the major difficulties with the provision of support systems to professionals is that child and youth care organizations have limited time and capability to provide such systems. Child and youth care organizations are confronted with a multitude of mandatory support systems, and professionals are confronted with an accumulation of support sessions. These organizations are under pressure to provide, and are responsible for providing, high-quality, low-cost services, and thus strive to effectively and efficiently integrate support systems.

The last objective of this research was to devise a potential way to integrate support systems for professionals that take into account these organizations’ capacities and incapacities. As a first step in integrating support for professionals, it is suggested (Chapter 6) that support be organized around common and overlapping (structural and contextual) factors of interventions. Organizing support based on the competences that professionals need to deliver multiple interventions serves two purposes. First, it enables child and youth care organizations to organize support more efficiently, while also reducing the accumulation of support sessions for professionals. This fulfils the desire of both child and youth care organizations and professionals to prevent the stacking of support systems and sessions. Second, it lays a foundation for improved practice. More efficient organization of support for professionals around common and overlapping factors increases the likelihood that the implementation of these factors is stimulated on a broad scale. Integrating support around overlapping therapeutic techniques would be the next logical step. Integrating support systems around overlapping factors is a totally new way of stimulating the quality of services delivery, which leads to new research tasks. The effects of integrating support systems need to be examined for child and youth care organizations, professionals, and clients.

7.2.4 Overview of research questions, objectives, corresponding chapters, and brief conclusions

The research questions, objectives, and corresponding chapters, together with a brief conclusion per chapter, are presented in Table 7.1.

Table 7.1
Overview of research questions, objectives, corresponding chapters, and brief conclusions

Research question	Objective	Corresponding chapter	Brief conclusion
1. What does it mean to apply interventions – as vehicles of the knowledge of “what works” – and how is this operationalized in outcome studies?	To examine the adequacy of the implementation of treatment integrity procedures in outcome studies of interventions targeting externalizing behavior problems of youth.	Chapter 2	Applying the content of interventions is referred to as treatment integrity. Although measuring treatment integrity is important, it is often missing or not examined under adequate circumstances. This weakens the confidence in statements made in many studies about the association between treatment integrity and intervention outcomes.
2. Does the application of these interventions make a real difference to the end-users of the services?	To meta-analytically examine the moderating effect of level of treatment integrity on the reduction of youth antisocial behavior after an intervention.	Chapter 3	Treatment integrity was found to be a significant moderator of the reduction of client antisocial behavior after an intervention. Significant medium to large effects of evidence-based interventions were found when the level of treatment integrity was high ($d = 0.633$). Non-significant small/marginal effects were found when the level of integrity was low ($d = 0.143$).

Table 7.1 (Continued)

Research question	Objective	Corresponding chapter	Brief conclusion
3. What types of support for professionals can strengthen implementation processes?	To examine the essential ingredients of support for child and youth care professionals to enable them to deliver the intended intervention with sufficient treatment integrity.	Chapter 4	An effective support system minimally consists of a combination of training and ongoing support (supervision, consultation, coaching), preferably extended with booster sessions. The support should focus on, among other things, providing feedback on levels of treatment integrity, preferably through the use of direct instruments (such as video, audio, and/or live sessions).
	To examine the experiences and use of treatment integrity instruments within child and youth care organizations.	Chapter 5	Instruments are used as part of quality assurance procedures (certification and/or recertification) and to provide performance feedback to professionals (supervision purpose). One of the most important conditions for professionals to perceive measurements as valuable, is that feedback is provided and the instruments are used for multiple purposes. The content of feedback that professionals prefer depends on their experience: Professionals who have just started to deliver the intervention prefer feedback on adherence, whereas experienced therapists prefer feedback on competence.
	To discuss a potential way to integrate support systems for professionals around overlapping factors of interventions.	Chapter 6	As a first step in integrating support for professionals, it is suggested to organize support around common and overlapping contextual and structural factors of interventions. More efficient organization of support for professionals around common, organizational, and structural factors increases the likelihood that implementation of these factors is stimulated on a broad scale.

7.3 Limitations

This dissertation answered questions about the implementation gap concerning what we know and what we do for children and young people with externalizing behavior problems and their families. Some limitations of the underlying research need to be mentioned. The first is the generalization of findings in Chapters 2, 3, and 5 to a broader set of interventions (such as practice-based approaches) and different target populations within child and youth care organizations. These three chapters focus on interventions aimed at children and young people with externalizing behavior problems and their families. Child and youth care organizations also provide care, treatment, and interventions to populations confronted with other difficulties, for example, children and young people with internalizing problems. In addition to the evidence-based interventions discussed in Chapters 2 and 3, child and youth care organizations also provide practice-based approaches that are developed through practice and in light of the outcomes of that practice. Whether the findings presented in this dissertation are applicable to practice-based interventions and different target populations needs examination.

The second limitation is the lack of generalizability of the findings in Chapters 2 and 3 to the whole range of interventions that focus on children and young people with externalizing behavior problems. The main focus of the research was on the treatment integrity of interventions for children and young people with externalizing behavior problems and their families. The studies included in the analyses in Chapters 2 and 3 are limited to those that have taken treatment integrity into account. The studies in these chapters show that only a small number of studies focusing on children and young people with externalizing behavior problems actually assess treatment integrity. In addition, studies that do take treatment integrity into account commonly focus on the same intervention.

The third limitation is that this research did not establish under which conditions a high level of treatment integrity is associated with positive client outcomes. In the multilevel meta-analysis study (Chapter 3), no differentiation between efficacy and effectiveness studies was made. It is therefore unknown whether the association between level of treatment integrity and client outcomes is different for these types of studies.

The fourth limitation is that it was not possible to identify the contribution that each aspect of treatment integrity makes to client outcomes. The measurements of treatment integrity have to reflect the delivery as intended. It is generally recognized in the literature that treatment integrity encompasses therapist adherence and therapist competence (Perepletchikova, Treat, & Kazdin, 2007). However, in the first study in this dissertation (Chapter 2), only three of the included outcome studies addressed both aspects. Of the 29 studies that assessed treatment integrity, 24 assessed treatment integrity solely as therapist adherence. Competence,

at least up until the moment of that study, was an absolute outlier in treatment integrity measurements. This is surprising given the importance of competence in implementing an intervention in accordance with a client's needs (see section 7.3 for a further discussion). The lack of measurements of competence in outcome studies also had an effect on the meta-analytic study that is included in this dissertation (Chapter 3). In that study, we could not differentiate between levels of adherence and competence, because most studies included in this meta-analysis did not make this differentiation. Therefore, it remains unknown which indicators of competence compromised treatment progress and had an impact on intervention outcomes. The outcome studies also had not incorporated treatment differentiation in their measurements. Thus, treatment differentiation could not be taken into account either. It remains unknown whether additional treatment methods may have decreased or increased the treatment effects of the interventions included in Chapter 3 (McLeod et al., 2013b). In short, previous studies did not take into account certain aspects of treatment integrity, possibly because the awareness of the distinction of aspects of treatment integrity and the tools to measure these were not (yet) available.

The last limitation is the reliability of the data of primary studies that were used to indicate levels of treatment integrity in this dissertation. An adequate operationalization of treatment integrity procedures was an inclusion criterion for outcome studies in two of the studies presented in this dissertation (Chapters 3 and 4). This criterion, however, does not guarantee that valid and reliable measurement instruments were used, or that the conditions under which measurements were made reflect the actual delivery as intended. The adequacy levels of implementation of treatment integrity procedures were obtained with the Implementation of Treatment Integrity Procedures Scale, adapted version (ITIPS-A) (Goense, 2013). This instrument can only be used for descriptive purposes. It gives a mere overall evaluation of the implementation of procedures on treatment integrity in a study. This limitation implies that the results of Chapters 3 and 4 have to be interpreted with caution.

7.4 Implications for clinical practice

The results of, and the knowledge gained during, this research have implications for the daily practices of child and youth care organizations. First of all, child and youth care organizations in the Netherlands and elsewhere typically provide a wide range of advice, guidance, help, and care, much of which is practice-based. The factors within the practice-based approaches that are to target the problems in many cases have not yet been made explicit. Over the last decade, most Dutch child and youth care organizations have made a great effort to describe their interventions, develop a theoretical underpinning, and collect information on goal

attainment and outcomes for clients (for an overview, see Databank Effectieve Jeugdinterventies, 2016).¹² Still, little is known about the content, delivery, or client outcomes of the vast majority of practice-based approaches and care as usual either in the Netherlands or abroad (Garland, Hawley, Brookman-Frazer, & Hurlburt, 2008). The results of the present research imply that child and youth care organizations have to be explicit, or more explicit, about what they are doing in care as usual, which practice-based approaches they are implementing, and the outcomes they achieve with clients. This demands a proper definition of what is being done, as practice cannot improve what it does not measure accurately, and it cannot measure what it does not define (inspired by Durlak & Dupre, 2008). It is only by being explicit about what is done, that reflection on the outcomes is possible and efforts can be made to sustain this possibly effective care. A proper description of the content of care also creates the possibility to disseminate it on a broader scale.

To be able to be explicit about the actual delivery of services, it is recommended that child and youth care organizations use treatment integrity instruments. These instruments should be developed for care as usual, which in most cases is a practice-based approach. In the development of these instruments, the focus should be on the core factors of each of the services that child and youth care organizations provide. The overlapping factors between these services have to be examined. As discussed in Chapter 6, many interventions tend to overlap on common, contextual, and structural factors, and possibly also on therapeutic techniques. Focusing on overlapping factors will mean that child and youth care organizations will not have to develop treatment integrity instruments for every single type of care they provide, which will decrease the administrative burden on both professionals and supervisors.

This dissertation shows that when treatment integrity data are used as part of quality assurance procedures and for supervision purposes, the data can function in a data loop. To contribute effectively to the continuous improvement of the services that are delivered, treatment integrity data should always be used together with client outcomes. Outcome data enable reflection on what is done and what the results of these actions are for clients. The combination of these data also provides essential information about the effectiveness of interventions that are being delivered under clinical conditions.

To be able to use treatment integrity information within a data feedback loop in clinical practice, an effort has to be made to incorporate support systems in which this is possible. This requires first and foremost that there is support for professionals. Child and youth care organizations should integrate the support sessions for professionals around overlapping common, contextual, and struc-

¹² www.nji.nl/interventies, see also www.youthpolicy.nl

tural factors of interventions. As discussed in Chapter 6, integrating support systems based on the competences that professionals need to deliver multiple interventions serves two purposes. First, it enables child and youth care organizations to organize support more efficiently. This meets the desire of both child and youth care organizations and professionals to prevent the stacking of support systems and sessions. Second, it lays a foundation for improved practice. More efficient organization of support for professionals around these factors increases the likelihood that the delivery of these factors are stimulated on a broad scale within a setting. Integrating support around overlapping therapeutic techniques would be the next logical step.

The support for professionals should focus not only on the actual delivery of the core factors of the care (adherence), but also, and maybe even more so, on the skillful, responsive, and flexible delivery of these factors (competence in delivery). As stated by Weisz et al. (2013b), in the actual youth ecosystem, in which so many real-world factors are at play, the client's needs can change mid-treatment, requiring the professional to shift his or her focus. This requires the professional to attend to the client's needs, without drifting from the intended intervention. The knowledge and skills necessary for professionals to flexibly deliver the core factors of interventions in ways that best meet the needs of clients, has to be the content of the support for these professionals. Therefore, it is essential to have supervisors and professionals who are capable of reflecting on treatment integrity and outcomes of clients.

7.5 Recommendations for future research

For over a long period of time primary outcome studies of evidence-based interventions for children and young people with externalizing behavior problems have not been taking treatment integrity measurements into account. With the growing attention to treatment integrity, the demands on the operationalization of treatment integrity procedures will increase and they should therefore receive more attention in outcome studies. Treatment integrity procedures reflect the establishment, assessment, evaluation, and reporting of treatment integrity. With respect to the establishment of treatment integrity, this dissertation has indicated the importance of providing professionals with frequent and targeted support. Thus, outcome studies should always report on treatment integrity in light of the support systems that are in place.

Second, scholars suggest that treatment integrity encompasses two aspects, namely therapist adherence and therapist competence (Perepletchikova, Treat, & Kazdin, 2007). It is recommended that in the assessment of treatment integrity, future research should distinguish between the different aspects of treatment integrity. Only by distinguishing treatment integrity in adherence and competence

can one fully understand how these different aspects interact and have an impact on client outcomes. It is crucial that valid and reliable instruments be developed and used to measure levels of adherence and competence in order to make any statements of value concerning this construct. Proper measurement of treatment integrity will provide crucial information about the content of training and support for professionals who are responsible for the delivery of these interventions.

Third, it is recommended that future studies of evidence-based interventions also measure treatment differentiation. As discussed in this dissertation, professionals may also be experienced in delivering particular treatment methods acquired in previous therapeutic work, that are not part of the specific intervention under study (McLeod et al., 2013b). The degree to which the professionals use these other treatment methods, and consequently deviate from the planned intervention, is referred to as treatment differentiation (Kazdin, 1994). Without measuring treatment differentiation, it is not possible to examine additional treatment methods that may decrease or increase treatment effects (McLeod et al., 2013b).

Finally, and as also stated by Weisz (2015), there should be more emphasis on developing and testing interventions in representative clinical conditions for which the interventions are ultimately intended. The clients in representative conditions often have a comorbidity of problems and are difficult to motivate for treatment. At the same time, resources for training and supporting professionals are minimal in these settings (Broeders, van der Helm, & Stams, 2015). As Weisz (2015, p. 131) stated: “It would be an interesting paradox if research designed to improve clinical care has produced treatments that do not fit – and may not cope so well with – the very clinical care it was intended to improve.” Researching interventions in conditions that are representative of the daily practices of child and youth care organizations, will generate more externally valid evidence on intervention effects, mechanisms of change, and factors that moderate effects. This type of research will generate knowledge that is applicable to the daily practices of child and youth care organizations (Weisz, 2015).

7.6 Concluding remarks

The primary studies included in this dissertation illustrate the complexities and challenges faced when conducting implementation research. On the positive side, this dissertation also illustrates the increasing knowledge in the field and the potential to sophisticate implementation research methodology and procedures.

Among the main challenges of implementation science is the disparity between optimal implementation circumstances and the limitations present in the practice field of child and youth care organizations. As also put forward at the 2015 Global Implementation Conference, going to scale and achieving collective and social impact requires 1) shared objectives for creating contexts that are hospitable to evidence-based programs, and 2) capacity building of key stakeholder groups to develop and institutionalize the infrastructure needed for the full and effective use of evidence-based interventions. Key stakeholder groups, funders, policy-makers, program developers, researchers, and service providers need to make both transformative and incremental changes in order to co-produce the visible infrastructure needed to support the effective implementation of evidence-based programs (Global Implementation Conference, 2015).

In this dissertation, I have addressed a few of the important questions around the implementation gap in child and youth care. As social circumstances change, implementation processes will change. This means there will be no end to the effort that needs to be made to effectively implement effective care. Collaboration and sharing of knowledge within and beyond country, scientific domain, and service setting borders is essential to gain a greater understanding of this ever-changing subject. The Global Implementation Initiative (GII) and European Implementation Collaborative (EIC), as well as the national networks around implementation, are great examples of collective efforts to increase the social impact of effective care. To achieve a greater social impact, it is essential to bridge the implementation gap. With this dissertation, I hope to have contributed to the knowledge necessary to bridge the gap, and thus to making “what works” work for children and young people with externalizing behavior problems and their families, both in the Netherlands and abroad.

References

- Abramowitz, J. (2006). Toward a functional analytic approach to psychologically complex patients: A comment on Ruscio and Holohan. *Clinical Psychology: Science and Practice*, *13*, 163–166.
- Addis, M.E. (2002). Methods for disseminating research products and increasing evidence-based practice: Priorities, obstacles, and future directions. *Clinical Psychology: Science and Practice*, *9*, 367–368.
- Albrecht, G., & Spanjaard, H. (2007). *Tools4U: Training cognitieve en sociale vaardigheden als taakstraf voor minderjarigen: Handleiding voor trainers*. Duivendrecht: PI Research.
- Algemene Rekenkamer (2012). *Werkwijze in jeugdgevangenis verbeterd; Justitie weet (nog) niet wat het effect op jeugdigen is. Behandeling veel duurder, maar budget lager door minder jeugdcriminaliteit [Work process improved in juvenile detention; Ministry of Law does not (yet) know the effects for juveniles. Treatment is much more expensive, but the budget is lower because of less juvenile delinquency]* (Persbericht 22-03-2012). Retrieved from http://www.rekenkamer.nl/Nieuws_overzicht/Persberichten/2012/03/Werkwijze_in_jeugdgevangenis_verbeterd_Justitie_weet_nog_niet_wat_het_effect_op_jeugdigen_is
- Andrews, D. A., & Bonta, J. (1994). *The psychology of criminal conduct*. Cincinnati: Anderson.
- Aos, S., M. Miller, & E. Drake. (2006). *Evidence-Based Public Policy Options to Reduce Future Prison Construction, Criminal Justice Costs, and Crime Rates*. Olympia: Washington State Institute for Public Policy, Document No. 06-10-1201
- Asscher, J. J., Deković, M., Manders, W. A., van der Laan, P. H., Prins, P. J., & van Arum, S., & Dutch MST Cost-Effectiveness Study Group (2014). Sustainability of the effects of Multisystemic Therapy for juvenile delinquents in the Netherlands: Effects on delinquency and recidivism. *Journal of Experimental Criminology*, *10*(2), 227–243. 10.1007/s11292-013-9198-8 (2014).
- Assink, M., van der Put, C.E., Hoeve, M., de Vries, S.L.A., Stams, G.J.J.M., & Oort, F.J. (2015). Risk factors for persistent delinquent behavior among juveniles: A meta-analytic review. *Clinical Psychology Review*, *42*, 47–61. doi:10.1016/j.cpr.2015.08.002
- Barber, J. P., Gallop, R., Crits-Christoph, P., Frank, A., Thase, M. E., Weiss, R. D., & Gibbons, B. C. (2006). The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: Results from the national institute drug abuse collaborative cocaine treatment study. *Psychotherapy Research*, *16*(2), 229–240. doi:10.1080/10503300500288951

- Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007a). Assessing intervention competence and its relation to therapy outcome. *Professional Psychology: Research and Practice, 38*(5), 493-500. doi:10.1037/0735-7028.38.5.493
- Barber, J., Triffleman, E., & Marmar, C. (2007b). Considerations in treatment integrity: Implications and recommendations for PTSD research. *Journal of Traumatic Stress, 20*(5), 793-805. doi:10.1002/jts.20295
- Barnoski, R. (2004). *Outcome evaluation of Washington State's research-based programs for juvenile offenders*. Olympia, WA: Washington State Institute for Public Policy.
- Barrett, P. M., & Ollendick, T. H. (Eds.). (2004). *Handbook of interventions that work with children and adolescents: Prevention and treatment*. New York: John Wiley.
- Barth, R.P., & Liggett-Creel, K. (2014). Common components of parenting programs for children birth to eight years of age involved with child welfare services. *Children and Youth Services Review, 40*, 6-12. doi: 10.1016/j.childyouth.2014.02.004
- Beidas, R. S., & Kendall, P. C. (2010). Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice, 17*, 1-30.
- Berger, M., & Kleine, K. (2013). *Handreiking reflecteren is leren*. Utrecht: Nederlands Jeugdinstituut (NJI).
- Blasé, K., & Fixsen, D. (2013). *Core intervention components. Identifying and operationalizing what makes programs work*. Washington, DC: Office of Human Services Policy, US Department of Health and Human Services.
- Boendermaker, L. (2011). *Implementeren is reflecteren. Evidence-based werken en de implementatie van interventies in de jeugdzorg*. Amsterdam: Lectoraat implementatie in de jeugdzorg/Hogeschool van Amsterdam.
- Brestan, E.V., & Eyberg, S.M. (1998). Effective psychosocial treatments of conduct disordered children and adolescents: 29 years, 82 studies, and 5,272 kids. *Journal of Clinical Child Psychology, 27*, 180-189.
- Broeders, R., van der Helm, G.H.P., & Stams, G.J.J.M. (2015). Preventie van gesloten plaatsing van jongeren met ernstige gedragsproblemen in Friesland: een pleidooi voor een gerichte aanpak. *Orthopedagogiek: Onderzoek en Praktijk, 54*, 318-330.

- Bruns, E. J., Suter, J. C., Force, M. M., & Burchard, J. D. (2005). Adherence to wraparound principles and association with outcomes. *Journal of Child & Family Studies, 14*(4), 521-534. doi:10.1007/s10826-005-7186-y
- Burns, B.J., Hoagwood, K., & Mrazek, P.J. (1995). Effective treatment for mental disorders in children and adolescents. *Clinical Child and Family Psychology Review, 2*, 199-254.
- Carr, A. (Ed.). (2000). *What Works With Children and Adolescents: A Critical Review of Psychological Interventions With Children, Adolescents and Their Families*. Florence, KY: Taylor & Francis/Routledge.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*(40), 1-9. doi:10.1186/1748-5908-2-40
- Chapman, J. E., & Schoenwald, S. K. (2011). Ethnic similarity, therapist adherence, and long-term multisystemic therapy outcomes. *Journal of Emotional & Behavioural Disorders, 19*(1), 3-16. doi:10.1177/1063426610376773
- Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta analyses: A structural equation modeling approach. *Psychological Methods, 19*(2), 211-229. <http://dx.doi.org/10.1037/a0032968>.
- Chorpita, B.F., & Daleiden, E.L. (2009). Mapping evidence-based treatments for children and adolescents: application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology, 77*, 566 - 579.
- Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2005). Identifying and selecting the common elements of evidence based interventions: A distillation and matching model. *Mental Health Services Research, 7*(1), 5-20.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, M. A., Piquero, A. R., & Jennings, W. G. (2010). Crime costs across offender trajectories: Studying the costs of crime across offender trajectories. *Criminology & Public Policy, 9*(2), 279-305. <http://dx.doi.org/10.1111/j.1745-9133.2010.00627.x>.
- Cross, W., Matthieu, M., Cerel, J., & Knox, K. (2007). Proximate outcomes of gatekeeper training for suicide prevention in the workplace. *Suicide and Life-Threatening Behavior, 37*, 659-670.
- Connor-Smith, J. K., & Weisz, J. R. (2003). Applying treatment outcomes research in clinical practice: Techniques for adapting interventions to the real world. *Child and Adolescent Mental Health, 8*, 3-10.

- 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*(1), 23-45.
- Davis, D.A., Thomson, M.A., Oxman, A.D., & Haynes, R.B. (1992). Evidence for the effectiveness of CME A review of 50 randomized controlled trials. *JAMA. 1992;268:1111Y1117*
- De Swart, J. J. W. , van den Broek, H., Stams, G. J. J. M., Asscher, J. J., van der Laan, P.H., Holsbrink-Engels, G. A., & van der Helm, G. H. P. (2012). The effectiveness of institutional youth care over the past three decades: A meta-analysis. *Children and Youth Services Review, 34*, 1818-1824.
- De Valk, S., Van der Helm, G.H.P., Beld, M., Schaftenaar, P., Kuiper, C., & Stams, G.J.J.M. (2015). Does punishment in secure residential youth care work? An overview of the evidence. *Journal of Children's Services, 10*, 3-16.
- Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist, 54*, 755-764.
- Dobson, D., & Cook, T. J. (1980). Avoiding type III errors in program evaluation: results from a field experiment. *Evaluation and Program Planning, 3*, 269-376.
- Durlak, J., & DuPre, E. (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*(3), 327-350.
- Eames, C., Daley, D., Hutchings, J., Whitaker, C. J., Bywater, T., Jones, K., & Hughes, J. C. (2010). The impact of group leaders' behaviour on parents acquisition of key parenting skills during parent training. *Behaviour Research and Therapy, 48*(12), 1221-1226. doi:10.1016/j.brat.2010.07.011
- Eames, C., Daley, D., Hutchings, J., Whitaker, C. J., Jones, K., Hughes, J. C., & Bywater, T. (2009). Treatment fidelity as a predictor of behaviour change in parents attending group-based parent training. *Child: Care, Health & Development, 35*(5), 603-612. doi:10.1111/j.1365-2214.2009.00975.x
- Effland, V., Walton, B., & McIntyre, J. (2011). Connecting the dots: Stages of implementation, wraparound fidelity and youth outcomes. *Journal of Child & Family Studies, 20*(6), 736-746. doi:10.1007/s10826-011-9541-5
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple graphical test. *British Medical Journal, 315*, 629-634.
- El-Tannir, A. (2002). The corporate university model for continuous learning, training, and development. *Education & Training, 44*, 76-81.

- Farrington, D. P., & Welsh, B. C. (2006). A half-century of randomized experiments on crime and justice. In M. Tonry (Ed.), *Crime and Justice* (pp. 55-132). Chicago: University of Chicago Press.
- Fixsen, D.L., & Ogden, T. (2014). Editorial. Facing the challenges of implementation. *Zeitschrift für Psychologie*, *222*(1), 1-3.
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.
- Forgatch, M.S., & DeGarmo, D.S. (2011). Sustaining Fidelity Following the Nationwide PMTO™ Implementation in Norway. *Prevention Science*, *12*, 235–246. doi: 10.1007/s11121-011-0225-6
- Fulda, K. G., Lykens, K. K., Bae, S., & Singh, K. P. (2009). Unmet mental health care needs for children with special health care needs stratified by socioeconomic status. *Child and Adolescent Mental Health*, *14*(4), 190–199. doi:10.1111/j.1475-3588.2008.00521.x
- Garland, A.F., Hawley, K.M., Brookman-Frazee, L., & Hurlburt, M.S. (2008). Identifying Common Elements of Evidence-Based Psychosocial Treatments for Children’s Disruptive Behavior Problems. *Journal of American Academy of Child and Adolescent Psychiatry*, *47*(5), 505-514.
- Garland, A.F., & Schoenwald, S.K. (2013). Use of Effective and Efficient Quality Control Methods to Implement Psychosocial Interventions. *Clinical Psychology: Science and Practice*, *20*, 33–43.
- Gendreau, P., Goggin, C., & Smith, P. (1999). The forgotten issue in effective correctional treatment: Program implementation. *International Journal of Offender Therapy and Comparative Criminology*, *43*(2), 180-187.
- Glisson, C., Schoenwald, S. K., Hemmelgarn, A., Green, P., Dukes, D., Armstrong, K. S., & Chapman, J. E. (2010). Randomized trial of MST and ARC in a two-level evidence-based treatment implementation strategy. *Journal of Consulting and Clinical Psychology*, *78*(4), 537-550. doi:10.1037/a0019160
- Global Implementation Conference (2015). CALL FOR ABSTRACTS: 3rd Biennial Global Implementation Conference Dublin Convention Centre – May 26 (Pre-Conference Day) – May 28, 2015.
- Godley, S.H., Garner, B.R., Smith, J.E., Meyers, R.J., & Godley, M.D. (2011). A Large-Scale Dissemination and Implementation Model for Evidence-Based Treatment and Continuing Care. *Clinical Psychology: Science and Practice*, *18*, 67-83.

- Goense, P.B. (2013). *ITIPS-A, manual for coding studies on implementation of procedures for treatment integrity, adapted version*. Amsterdam University of Applied Sciences (Internal publication).
- Goense, P.B., Assink, M., Stams, G.J.J.M., Boendermaker, L., & Hoeve, M. (2016a). *Making 'What Works' Work: A meta-analytic study of the effect of treatment integrity on outcomes of evidence-based interventions for juveniles with antisocial behavior*. Manuscript provisionally accepted for publication.
- Goense, P.B., Boendermaker, L., & van Yperen, T. (2016b). *Measuring Treatment Integrity: Use of and experiences with measurements in routine youth care practice settings*. Manuscript submitted for publication.
- Goense, P.B., Boendermaker, L., & van Yperen, T. (2015a). Support Systems for Treatment Integrity. *Research on Social Work Practice*, 1-5.
- Goense, P., Boendermaker, L., van Yperen, T., Stams, G.J., & van Laar, J. (2014). Implementation of Treatment Integrity Procedures: An Analysis of Outcome Studies of Youth Interventions Targeting Externalizing Behavioral Problems. *Zeitschrift für Psychologie*, 222(1), 12–21.
- Goense, P. B., Pronk, S., Boendermaker, L., Bakker, R., Ruitenberg, I., & Bertling, L. (2015b). *Leren op de werkvloer. Organisatie en inhoud van leren op de werkvloer*. Delft: Eburon.
- Goldstein A.P., Glick, B. & Gibbs, J.C. (1998). *Aggression Replacement Training; A comprehensive intervention for aggressive youth* (rev. ed.). Champaign, IL.: Research Press.
- Graham, C., Carr, A., Rooney, B., Sexton, T., & Satterfield, R. L. W. (2014). Evaluation of functional family therapy in an Irish context. *Journal of Family Therapy*, 36, 20–38. doi: 10.1111/1467-6427.12028
- Hayes, S. (2002). Getting to dissemination. *Clinical Psychology: Science and Practice*, 9, 410–415.
- Helmond, P., Overbeek, G., & Brugman, D. (2015). An examination of program integrity and recidivism of a cognitive-behavioral program for incarcerated youth in The Netherlands. *Psychology, Crime & Law*. doi: 10.1080/1068316X.2014.989164
- Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology*, 65(5), 821-833.
- Henggeler, S. W., Pickrel, S. G., & Brondino, M. J. (1999). Multisystemic treatment of substance-abusing and dependent delinquents: Outcomes, treat-

- ment fidelity, and transportability. *Mental Health Services Research*, 1(3), 171-184.
- Hepner, K.A., Hunter, S.B., Paddock, S.M., Zhou, A., & Watkins, K.E. (2011). Training addiction counselors to implement CBT for depression. *Administration and Policy in Mental Health*, 38, 313–323. doi:10.1007/s10488-011-0359-7
- Hogue, A., Dauber, S., Samuolis, J., & Liddle, H. A. (2006). Treatment techniques and outcomes in multidimensional family therapy for adolescent behavior problems. *Journal of Family Psychology*, 20(4), 535-543. doi:10.1037/0893-3200.20.4.535
- Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting & Clinical Psychology*, 76(4), 544-555.
- Hogue, A., Ozechowski, T.J., Robbins, M.S., Waldron, H.B. (2013). Making Fidelity an Intramural Game: Localizing Quality Assurance Procedures to Promote Sustainability of Evidence-Based Practices in Usual Care. *Clinical Psychology: Science and Practice*, 20, 60-77.
- Holth, P., Torsheim, T., Sheidow, A. J., Ogden, T., & Henggeler, S. W. (2011). Intensive quality assurance of therapist adherence to behavioral interventions for adolescent substance use problems. *Journal of Child & Adolescent Substance Abuse*, 20(4), 289-313. doi:10.1080/1067828X.2011.581974
- Houben, M., Van den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*. <http://dx.doi.org/10.1037/a0038822> (Advance online publication).
- Hox, J. (2002). *Multilevel analysis: Techniques and applications* (2nd ed.). Mahwah: Lawrence Erlbaum Associates Publishers.
- Hsieh, H., & Shannon, S.E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15, 1277–1288.
- Huey, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology*, 68(3), 451-467.
- Kalidien, S.N., de Heer- de Lange, N.E., & van Rosmalen, M.M. (2010). *Criminaliteit en Rechtshandhaving 2010. Ontwikkelingen en samenhangen*. Ministerie van Justitie.

- Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among U.S. children: Variation by ethnicity and insurance status. *American Journal of Psychiatry*, 159, 1548–1555. doi:10.1176/appi.ajp.159.9.1548
- Kazdin, A. (1994). Methodology, design, and evaluation in psychotherapy research. In A. E. Bergin, & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 19–71). Oxford, UK: John Wiley & Sons.
- Kazdin, A.E., & Weisz, J.R. (1998). Identifying and developing empirically supported child and adolescent treatments. *Journal of Consulting and Clinical Psychology*, 66, 19-36.
- Kazdin, A.E., & Weisz, J.R. (2003). *Evidence-based psychotherapies for children and adolescents*. New York: The Guilford Press.
- Kerby, N., T. (Ed.). (2006). *Helping others help children; clinical supervision of child psychotherapy*. Washington DC.: American Psychological Association (APA).
- Knotter, M.H., Wissink, I.B., Moonen, X.M.H., Stams, G.J.J.M., & Jansen, G.J. (2013). Staff's attitudes and reactions towards aggressive behaviour of clients with intellectual disabilities: A multi-level study. *Research in Developmental Disabilities*, 34, 1397-1407.
- Knox, L. M., & Aspy, C. B. (2011). Quality improvement as a tool for translating evidence-based interventions into practice: What the youth violence prevention community can learn from healthcare. *American Journal of Community Psychology*, 48(1–2), 56–64. doi:10.1007/s10464-010-9406-x.
- Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Liddle, H. A., Dakof, G. A., Henderson, C., & Rowe, C. (2011). Implementation outcomes of multidimensional family therapy-detention to community: A reintegration program for drug-using juvenile detainees. *International Journal of Offender Therapy and Comparative Criminology*, 55(4), 587-604.
- Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: a meta analytic overview. *Victims and offenders*, 4, 124-147.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

- Little, E., Hudson, A., & Wilks, R. (2002). The efficacy of written teacher advice (tip sheets) for managing classroom behaviour problems. *Educational Psychology, 22*(3), 251-266.
- Lu, W., Yanos, P.T., Gottlieb, J.D., Marcello Duva, S., Xie, H., Rosenberg, S.D., & Mueser, K.T. (2012). Use of Fidelity Assessments to Train Clinicians in the CBT for PTSD Program for Clients With Serious Mental Illness. *Psychiatric Services, 63*, 785-792.
- MaxQDA 11 (2016). *Software for qualitative data analysis, 1989-2016, VERBI Software – Consult – Sozialforschung GmbH*. Berlin: Germany.
- Mazzucchelli, T.G., & Sanders, M.R. (2010). Facilitating Practitioner Flexibility Within an Empirically Supported Intervention: Lessons From a System of Parenting Support. *Clinical Psychology: Science and Practice, 17*, 238–252.
- McGlinchey, J. B., & Dobson, K. S. (2003). Treatment integrity concerns in cognitive therapy for depression. *Journal of Cognitive Psychotherapy: An International Quarterly, 17*, 299-318.
- McLeod, B. D., Islam, N. Y., & Wheat, E. (2013a). Designing, conducting, and evaluating therapy process research. In J. Comer & P. Kendall (Eds.), *The Oxford handbook of research strategies for clinical psychology* (pp. 142-164). New York, NY: Oxford University Press.
- McLeod, B.D., Southam-Gerow, M.A., Tully, C.B., Rodriguez, A., & Smith, M.M. (2013b). Making a Case for Treatment Integrity as a Psychosocial Treatment Quality Indicator for Youth Mental Health Care. *Clinical Psychology, 20*(1), 1-23.
- Mikolajczak, J., Stals, K., Fleuren, M. A. H., Wilde, E. J. d., & Paulussen, T. G. W. M. (2009). *Kennisynthese van condities voor effectieve invoering van jeugdinterventies*. Leiden / Utrecht: TNO Kwaliteit van Leven / Nederlands Jeugdinstituut (NJI).
- Miller, W. R., Sorensen, J. L., Selzer, J. A., & Brigham, G.S. (2006). Disseminating evidence based practices in substance abuse treatment: A review with suggestions. *Journal of Substance Abuse Treatment, 31*, 25–39. doi:10.1016/j.jsat.2006.03.005
- Miller, W., Yahne, C., Moyers, T., Martinez, J., & Pirritano, M. (2004). A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting and Clinical Psychology, 72*, 1050–1062.
- Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247-266.

- Morrow, S.L. (2005). Quality and trustworthiness in qualitative research in counseling psychology. *Journal of Counseling Psychology*, 52, 250-260.
- Nápoles, A.M., Santoyo-Olsson, J., & Stewart, A.L. (2013). Methods for Translating Evidence-Based Behavioral Interventions for Health-Disparity Communities. *Preventing Chronic Disease*, 10. doi: <http://dx.doi.org/10.5888/pcd10.130133>
- Ollendick, T.H., & King, N.J. (2000). *Empirically Supported Treatments for Children and Adolescents, 2nd ed.* New York: Guilford; 2000:386Y425.
- Parhar, K. K., Wormith, J. S., Derkzen, D. M., & Beauregard, A. M. (2008). Offender coercion in treatment: A meta-analysis of effectiveness. *Criminal Justice and Behavior*, 35, 1109-1135.
- Perepletchikova, F. (2006). Retrieved April, 2013, from www.treatmentintegrity.com
- Perepletchikova, F., & Kazdin, A. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology-Science and Practice*, 12(4), 365-383. doi:10.1093/clipsy/bpi045
- Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, 75(6), 829-841.
- Peters, G.-J. Y (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist*, 16 (2).
- R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (Retrieved from <http://www.Rproject.org/>).
- Robbins, M. S., Feaster, D. J., Horigian, V. E., Puccinelli, M. J., Henderson, C., & Szapocznik, J. (2011). Therapist adherence in brief strategic family therapy for adolescent drug abusers. *Journal of Consulting and Clinical Psychology*, 79(1), 43-53.
- Rosenthal, R. (1994). Statistically describing and combining studies. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). New York: Russell Sage Foundation.
- Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, 118(2), 183.

- Sackett, D.L., Straus, S.E., Richardson, W.S., Rosenberg, W., & Haynes, R.B. (2000). *Evidence-Based Medicine: How to Practice and Teach EBM, 2nd edition*. Edinburgh: Churchill Livingstone.
- Sanders, M. R. (2008). Triple P-Positive Parenting Program as a public health approach to strengthening parenting. *Journal of Family Psychology, 22*(3), 506–517. doi:10.1037/0893-3200.22.3.506.
- Sanetti, L. M. H., Gritter, K. L., & Dobey, L. (2011). Treatment integrity of interventions with children in the school psychology literature from 1995 to 2008. *School Psychology Review, 40*, 72-84.
- Schoenwald, S. K. (2011). It's a bird, It's a plane, it's ...fidelity measurement in the real world. *Clinical Psychology: Science and Practice, 18*(2), 142–147. doi:10.1111/j.14682850.2011.01245.x
- Schoenwald, S. K., Carter, R. E., Chapman, J. E., & Sheidow, A. J. (2008). Therapist adherence and organizational effects on change in youth behavior problems one year after multisystemic therapy. *Administration and Policy in Mental Health and Mental Health Services Research, 35*(5), 379-394. doi:10.1007/s10488-008-0181-z
- Schoenwald, S. K., Chapman, J. E., Sheidow, A. J., & Carter, R. E. (2009a). Long-term youth criminal outcomes in MST transport: The impact of therapist adherence and organizational climate and structure. *Journal of Clinical Child and Adolescent Psychology, 38*(1), 91-105. doi:10.1080/15374410802575388
- Schoenwald, S. K., & Garland, A. F. (2013). A review of treatment adherence measurement methods. *Psychological Assessment, 25*(1), 146–156. doi:10.1037/a0029715
- Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S.L., Sheidow, A.J., & Southam-Gerow, M. A. (2011). Toward the Effective and Efficient Measurement of Implementation Fidelity. *Administration and Policy in Mental Health, 38*, 32-43.
- Schoenwald, S. K., Sheidow, A. J., & Chapman, J. E. (2009b). Clinical supervision in treatment transport: Effects on adherence and outcomes. *Journal of Consulting & Clinical Psychology, 77*(3), 410-421. doi:10.1037/a0013788
- Schoenwald, S. K., Sheidow, A. J., Letourneau, E. J., & Liao, J. (2003). Transportability of multisystemic therapy: Evidence for multilevel influences. *Mental Health Services Research, 5*(4), 223-239. doi:10.1023/A:1026229102151
- Schoenwald, S. K., Sheidow, A. J., & Letourneau, E. J. (2004). Toward effective quality assurance in evidence-based practice: Links between expert consultation, therapist fidelity, and child outcomes. *Journal of Clinical Child & Adolescent Psychology, 33*(1), 94-104.

- Sexton, T., & Turner, C. W. (2010). The effectiveness of functional family therapy for youth with behavioral problems in a community practice setting. *Journal of Family Psychology : JFP : Journal of the Division of Family Psychology of the American Psychological Association (Division 43)*, 24(3), 339-348.
- Smith, D.C., Ureche, D.J., Davis, J.P., & Walters, S.T. (2015). Motivational Interviewing With and Without Normative Feedback for Adolescents With Substance Use Problems: A Preliminary Study. *Substance Abuse*, 36(3), 350-358. doi:10.1080/08897077.2014.988838
- Southam-Gerow, M.A., & Prinstein, M.J. (2014). Evidence Base Updates: The Evolution of the Evaluation of Psychological Treatments for Children and Adolescents. *Journal of Clinical Child & Adolescent Psychology*, 43(1), 1-6. doi:10.1080/15374416.2013.855128
- Stambaugh, L. F., Mustillo, S. A., Burns, B. J., Stephens, R. L., Baxter, B., Edwards, D., & DeKraai, M. (2007). Outcomes from wraparound and multisystemic therapy in a center for mental health services system-of-care demonstration site. *Journal of Emotional and Behavioral Disorders*, 15(3), 143-155.
- Stanger, C., Ryan, S.R., Scherer, E.A., Norton, G.E., & Budney, A.J. (2015). Clinic- and Home-based Contingency Management Plus Parent Training for Adolescent Cannabis Use Disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 54(6), 445-453.
- Steinfeld, B. I., Coffman, S. J., & Keyes, J. A. (2009). Implementation of an evidence-based practice in clinical setting: What happens when you get there? *Professional Psychology: Research and Practice*, 40, 410-416. doi:10.1037/a0015677
- Stewart, D.G., Felleman, B.I., & Arger, C.A. (2015). Effectiveness of Motivational Incentives for Adolescent Marijuana Users in a School-Based Intervention. *Journal of Substance Abuse Treatment*, 58, 43-50.
- Sundell, K., Hansson, K., Löfholm, C. A., Olsson, T., Gustle, L. H., & Kadesjö, C. (2008). The transportability of multisystemic therapy to Sweden: Short-term results from a randomized trial of conduct-disordered youths. *Journal of Family Psychology: JFP: Journal of the Division of Family Psychology of the American Psychological Association (Division 43)*, 22(4), 550-560.
- Sutherland, K.S., McLeod, B.D., Conroy, M.A., & Cox, J.R. (2013). Measuring Implementation of Evidence-Based Programs Targeting Young Children at Risk for Emotional/Behavioral Disorders: Conceptual Issues and Recommendations. *Journal of Early Intervention*, 35(2): 129-149. doi: 10.1177/1053815113515025

- Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R., & Jones, D. R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal*, 320, 1574–1577.
- Tennyson, H.R. (2009). Reducing Juvenile Recidivism: A meta-analysis of Treatment Outcomes. *School of professional Psychology. Paper 109*.
- Toffalo, D. A. D. (2000). An investigation of treatment integrity and outcomes in wraparound services. *Journal of Child and Family Studies*, 9(3), 351–361.
- Turnell, A., & Edwards, S. (1999). *Signs of safety: A solution and safety oriented approach to child protection casework*. New York/London: Norton.
- Van den Noortgate, W., López-López, J. A., Marin-Martinez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, 45(2), 576–594. <http://dx.doi.org/10.3758/s13428-012-0261-6>.
- Van den Noortgate, W., López-López, J. A., Marin-Martinez, F., & Sánchez-Meca, J. (2014). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods*. <http://dx.doi.org/10.3758/s13428-014-0527-2> (Advance online publication).
- Van der Helm, G.H.P., Boekee, I., Stams, G.J.J.M., & Van der Laan, P.H. (2011) Fear is the key: Keeping the balance between flexibility and control in a Dutch Youth Prison. *Journal of Children's Services*, 6, 248–263.
- Van Yperen, T.A. (Ed.). (2010). *55 vragen over effectiviteit. Antwoorden voor de jeugdzorg*. Utrecht: Nederlands Jeugdinstituut.
- Van Yperen, T.A., Veerman, J.W., & van den Berg, G. (juni 2015). *Elementen die er toe doen. Overzicht van begrippen over werkzame elementen en een voorstel voor een indeling*. Den Haag: ZonMw.
- Veerman, J.W., & van Yperen, T.A. (2007). Degrees of freedom and degrees of certainty: A developmental model for the establishment of evidence-based youth care. *Evaluation and Program Planning*, 30, 212–221.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48.
- Voogden, M., & Kuyvenhoven, M. (2010). Onderzoek naar de effecten van supervisie. *Tijdschrift Supervisie en Coaching*, 27, 3–14.
- Walker, H. M., Golly, A., McLane, J. Z., & Kimmich, M. (2005). The oregon first step to success replication initiative: Statewide results of an evaluation of the program's impact. *Journal of Emotional & Behavioral Disorders*, 13(3), 163–172.

- Waltz, J., Addis, M., Koerner, K., & Jacobson, N. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*(4), 620-630.
- Weisburd, D. (2003). Ethical practice and evaluation of interventions in crime and justice: The moral imperative for randomized trials. *Evaluation Review, 27*, 336-354.
- Weissberg, R. P., Kumpfer, K. L., & Seligman, M. (2003). Prevention that works for children and youth - an introduction. *American Psychologist, 58*(6-7), 425-432.
- Weisz, J.R. (2015). Bridging the Research-Practice Divide in Youth Psychotherapy: The Deployment-Focused Model and Transdiagnostic Treatment. *Verhaltenstherapie, 25*, 129–132. doi: 10.1159/000430432
- Weisz, J. R., Jensen, A. L., & McLeod, B. D. (2005). Development and dissemination of child and adolescent psychotherapies: Milestones, methods, and a new deployment-focused model. In E. D. Hibbs & P. S. Jensen (Eds.). *Psychosocial treatments for child and adolescent disorders: Empirically based strategies for clinical practice* (pp. 9–39). Washington, DC: American Psychological Association.
- Weisz, J. R., Jensen-Doss, A., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist, 61*, 671–689.
- Weisz, J. R., Kuppens, S., Eckshtain, D., Ugueto, A. M., Hawley, K. M., & Jensen-Doss, A. (2013a). Performance of evidence-based youth psychotherapies compared with usual clinical care: A multilevel meta-analysis. *JAMA Psychiatry, 70*(7), 750–761. <http://dx.doi.org/10.1001/jamapsychiatry.2013.1176>.
- Weisz, J.R., Ugueto, A.M., Cheron, D.M., & Herren, J. (2013b). Evidence-Based Youth Psychotherapy in the Mental Health Ecosystem. *Journal of Clinical Child & Adolescent Psychology, 42*(2), 274-286. doi:10.1080/15374416.2013.764824
- Wibbelink, C. J. M., & Assink, M. (2015). *Handleiding voor het uitvoeren van een drie-level meta-analyse in R*. [Manual for conducting a three-level meta-analysis in R]. Amsterdam, the Netherlands: Universiteit van Amsterdam.

Nederlandse samenvatting

Het overbruggen van de implementatiekloof

Een studie over duurzame implementatie van interventies in de zorg voor jeugd

Inleiding

De kennis over wat er werkt voor kinderen en jongeren die problemen hebben die zich uiten in gedrag als veel ruzie maken, tegendraads zijn, vechten (zogenoemde externaliserende gedragsproblemen), heeft de laatste jaren een ware vlucht genomen (Southam-Gerow & Prinstein, 2014). Dit heeft er echter niet toe geleid dat de aanpak van deze gedragsproblemen in de praktijk in dezelfde mate een vooruitgang heeft laten zien (Weisz et al., 2013a). Dit onderscheid tussen de kennis over wat er werkt en de toepassing en het gebruik van deze kennis in de praktijk, heet ook wel de implementatiekloof (Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005). In *evidence-based* interventies zit veel kennis over ‘wat werkt’ verpakt. Deze interventies zijn theoretisch onderbouwd, goed gedocumenteerd¹³, geprotocolleerd en gestructureerd, hebben een handleiding en er is op basis van (quasi-) experimenteel onderzoek, empirisch bewijs voor de werkzaamheid van deze interventies (Weisz, Jensen-Doss, & Hawley, 2006). Deze interventies geven daarmee de professional houvast in zijn of haar werk. Het toepassen van de inhoud van een interventie zoals beschreven, wordt in de literatuur over implementatie aangeduid als behandelintegriteit (Perepletchikova, Treat, & Kazdin, 2007).

Het doel van het onderzoek beschreven in dit proefschrift is om bij te dragen aan de kennis die noodzakelijk is voor het overbruggen van de implementatiekloof. De focus is daarbij gericht op het uitvoeren van een interventie zoals bedoeld (behandelintegriteit). Om het doel te bereiken is gebruik gemaakt van verschillende onderzoeksmethoden.

Dit proefschrift bestaat uit twee systematische literatuurstudies, een multilevel meta-analyse, een kwalitatieve studie en een opiniërend artikel. In de eerste twee studies (hoofdstuk 2 en 3) ligt de nadruk op de kennis over de toepassing van *evidence-based* interventies. In een systematisch literatuuronderzoek is onderzocht hoe behandelintegriteit is geoperationaliseerd in uitkomststudies en door middel van een multilevel meta-analyse is onderzocht of een hoge mate van behandelintegriteit daadwerkelijk positieve uitkomsten voor jongeren tot gevolg heeft. De derde studie gaat in op de essentiële elementen in de ondersteuning aan

¹³ Goed gedocumenteerd betekent dat klinische expertise en cliëntvoorkeuren met betrekking tot de interventie zijn gedocumenteerd. *Evidence-based practice* is immers de integratie van het best beschikbare wetenschappelijke bewijs met klinische expertise en cliëntvoorkeuren (Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000, p 1).

professionals om hen in staat te stellen interventies daadwerkelijk goed toe te passen. Dit is uitgezocht aan de hand van een systematisch literatuuronderzoek (hoofdstuk 4). In uitkomststudies wordt veelal gebruik gemaakt van instrumenten om de mate van behandelintegriteit inzichtelijk te maken (Schoenwald, & Garland, 2013). Informatie over de toepassing van deze instrumenten in de praktijk ontbreekt echter. Daarom zijn de ervaringen met en het gebruik van dergelijke instrumenten in jeugdzorgorganisaties aan de hand van interviews met professionals en supervisors onderzocht (hoofdstuk 5). Ten slotte is een opiniërend artikel (hoofdstuk 6) geschreven om de, tijdens het schrijven van dit proefschrift, opgedane kennis en ideeën ten aanzien van de inrichting van ondersteuning aan professionals met een breder publiek te delen.

De onderzoeksvragen in de studies zijn beantwoord met data verzameld binnen het programma van de Hogeschool van Amsterdam: Lectoraat Kwaliteit en Effectiviteit in de Zorg voor Jeugd, Faculteit Maatschappij en Recht. De beschreven studies in de hoofdstukken 4 en 5 vallen daarbij binnen het project 'Wie helpt de hulpverlener?' dat financieel is ondersteund door Stichting Innovatie Alliantie (subsidienummer 2012-14-18P).

Belangrijkste bevindingen

Hoofdstuk 2 beschrijft het systematische literatuuronderzoek naar de operationalisering van behandelintegriteit in uitkomststudies van interventies voor jongeren met externaliserende gedragsproblemen. Onderzoekers concluderen regelmatig, dat een lage mate van behandelintegriteit de oorzaak kan zijn van teleurstellende resultaten van interventies in uitkomststudies (Schoenwald, Chapman, Sheidow, & Carter, 2009a; Sexton, & Turner, 2010; Tennyson, 2009). Het doel van de studie was om te achterhalen of uitkomststudies behandelintegriteit goed operationaliseren. Een goede operationalisering van behandelintegriteit is immers noodzakelijk om daadwerkelijk een inzicht in de toepassing van de interventie te verkrijgen.

De resultaten laten zien dat het meten van behandelintegriteit in uitkomststudies van interventies voor jongeren met externaliserende gedragsproblemen tot aan april 2012 uitzonderlijk was. In totaal zijn slechts 24 artikelen met daarin 29 primaire studies gevonden die daadwerkelijk behandelintegriteit hebben onderzocht. Met behulp van de Implementation of Treatment Integrity Procedures Scale – Adapted (ITIPS-A) (Goense, 2013) is de operationalisering van behandelintegriteit in de geïncludeerde studies beoordeeld. Een groot gedeelte van de studies die zijn opgenomen (80%), hebben de operationalisering op een niveau uitgewerkt die grenst aan voldoende. De bevindingen waren beperkt in generaliseerbaarheid aangezien bijna de helft (45%) van de studies zich richt op

één en dezelfde interventie – Multisysteem therapie (MST) - in de meeste gevallen onderzocht door dezelfde onderzoekers.

Het meest opvallende was dat behandelcompetentie (*competence*, het responsief kunnen toepassen van de interventie zonder verlies van de essentiële elementen) beduidend minder is onderzocht dan behandeltrouw (*adherence*, het toepassen van alle onderdelen van de interventie) terwijl behandelintegriteit bestaat uit beide onderdelen (Barber et al., 2006; Barber, Sharpless, Klostermann, & McCarthy, 2007a; Barber, Triffleman, & Marmar, 2007b; Mcleod, Southam-Gerow, Tully, Rodriguez, & Smith, 2013b; Perepletchikova, Treat, & Kazdin, 2007).

De studie laat zien dat ondanks dat veel onderzoekers het belang van het meten van behandelintegriteit in studies onderschrijven, behandelintegriteit in veel studies niet is gemeten of niet goed is gemeten. Met name mist de aandacht voor de *competence*. Het gebrek aan studies die behandelintegriteit op de juiste manier hebben geoperationaliseerd, ondermijnt het vertrouwen dat kan worden gelegd in uitspraken over de relatie tussen behandelintegriteit en de uitkomsten van interventies. Zonder goed onderzoek naar behandelintegriteit blijft het onbekend hoe interventies daadwerkelijk worden uitgevoerd en kunnen er geen uitspraken worden gedaan over de relatie tussen behandelintegriteit en de uitkomsten van interventies.

In **hoofdstuk 3** wordt de multilevel meta-analyse beschreven die is uitgevoerd met enkel studies waarin de operationalisering van behandelintegriteit volgens de Implementation of Treatment Integrity Procedures Scale – Adapted (ITIPS-A) (Goense, 2013) voldoende adequaat was. De search voor deze studies is na het onderzoek zoals beschreven in hoofdstuk 2, geüpdatet tot en met november 2015. Het doel van deze meta-analyse was om te onderzoeken wat het modererende effect¹⁴ is van de mate van behandelintegriteit op de afname van antisociaal gedrag van jongeren na een interventie.

Behandelintegriteit blijkt een significante moderator van de afname van het antisociale gedrag van jongeren na een interventie. Ook is een significant verschil gevonden tussen de effectgroottes van studies met een hoge mate van behandelintegriteit en studies met een lage mate van behandelintegriteit. Een significant medium-tot-groot effect van *evidence-based* interventies is gevonden voor studies waarin de mate van behandelintegriteit hoog was ($d = 0.633$). Een niet-significant klein/marginaal effect is gevonden voor *evidence-based* interventies waar de behandelintegriteit laag was ($d = 0.143$). De resultaten laten ook zien dat de relatie tussen een hoge mate van behandelintegriteit en uitkomsten van de

¹⁴ Een modererend effect wil zeggen dat wordt nagegaan in hoeverre de variabele behandelintegriteit een significante bijdrage levert aan de relatie tussen de interventie en de uitkomsten voor jongeren.

interventie hetzelfde blijft wanneer gecontroleerd wordt voor andere significante moderatoren zoals het type studie, type interventie en de duur van de interventie.

In weinig van de studies die zijn opgenomen in de meta-analyse ($k = 6$, 35.3%), was de mate van behandelintegriteit zoals gemeten in die studies, daadwerkelijk hoog. De resultaten van deze studie laten zien dat het toepassen van een *evidence-based* interventie met een hoge mate van behandelintegriteit cruciaal is, maar ook dat die hoge integriteit niet vaak voorkomt. Dat maakt de noodzaak om behandelintegriteit hoog op de agenda te zetten hoog. Een *evidence-based* interventie gebruiken is geen garantie voor effect, de interventie goed gebruiken geeft veel meer kans op dat effect.

Hoofdstuk 4 beschrijft het systematische literatuuronderzoek over hoe professionals het beste ondersteund kunnen worden in het behandelinteger toepassen van een interventie. In deze studie zijn ondersteuningssystemen geanalyseerd die een positief effect hadden op de mate van behandelintegriteit aan de interventie. De resultaten laten zien dat een effectief ondersteuningssysteem minimaal bestaat uit een combinatie van training en doorlopende ondersteuning (supervisie, consultatie, coaching), bij voorkeur gecombineerd met zogenaamde boostersessies (opfrustrainingen). De resultaten laten zien dat in de structurele ondersteuning gebruik moet worden gemaakt van actieve leermethodes. De leermethodes moeten zich daarbij onder andere richten op het geven van feedback op de mate van behandelintegriteit van de professional, idealiter door gebruik te maken van video, audio en/of live sessies.

De studie beschreven in hoofdstuk vier laat zien dat het structureel ondersteunen van professionals, door bijvoorbeeld een supervisor, bijdraagt aan een behandelintegere uitvoering van een interventie. Het is dus cruciaal dat jeugdzorgorganisaties na het trainen van hun professionals in de interventies, zorg dragen voor een voortdurende ondersteuning van professionals gericht op de uitvoering van die interventies.

Hoofdstuk 5 beschrijft of en hoe instrumenten voor het meten van behandelintegriteit worden gebruikt in jeugdzorgorganisaties en wat de ervaringen van professionals en supervisoren hiermee zijn. De resultaten van de kwalitatieve studie laten zien dat bij vrijwel alle interventies waar professionals van zijn geïnterviewd, instrumenten worden gebruikt om de mate van behandelintegriteit te bepalen. De meeste instrumenten worden door professionals zelf, hun supervisor of een externe beoordelaar ingevuld aan de hand van video-opnamen van professionals die aan het werk zijn met hun cliënt(en). Het gebruik van video-opnamen in elke supervisie sessie was slechts het geval bij één interventie (PMTO). De professionals van die interventie, zijn hier positief over. De instrumenten worden over het algemeen voor verschillende doeleinden gebruikt. Vaak worden ze gebruikt als onderdeel van de kwaliteitsbewakingsprocedures (certificering 54.6% of

hertificering 36.4%) of in de supervisie om feedback te geven aan professionals over hun handelen (72.7%). Slechts 36.4% van de interventies die in de studie zijn geïnccludeerd, gebruiken de instrumenten voor meer dan één doel.

De geïnterviewde professionals en supervisors stellen dat de belangrijkste voorwaarde voor professionals om de afname van de instrumenten als waardevol te beschouwen, is dat deze worden gebruikt voor meerdere doeleinden en dat er feedback op de scores op de instrumenten wordt gegeven. De inhoud van de feedback die professionals prefereren verschilt volgens een aantal geïnterviewde professionals tussen startende en meer ervaren professionals. Startende professionals zouden meer behoefte hebben aan feedback op behandeltrouw (*adherence*) terwijl ervaren professionals meer behoefte hebben aan feedback op (algemene) competenties (*common competence*) zoals het aangaan van een alliantie met de cliënt.

Op basis van de studie beschreven in hoofdstuk 5 kan geconcludeerd worden dat instrumenten waarmee behandelintegriteit wordt gemeten standaard gebruikt kunnen worden in de supervisie, voor meerdere doeleinden gebruikt kunnen worden en dat dit ook de voorkeur heeft van professionals. In de praktijk gebeurt dit allen echter nog maar zelden, zo blijkt uit deze studie. In samenhang bezien met de resultaten van de studie uit hoofdstuk 4 is duidelijk dat er nog veel valt te verbeteren als het gaat om de ondersteuning van deze professionals. Er moet meer gebruik worden gemaakt van structureel direct leren - aan de hand van video-opnamen - tijdens de supervisie, instrumenten moeten voor meerdere doelen worden ingezet en er moet gedifferentieerd worden naar behoefte van professionals met betrekking tot de inhoud van de feedback. Deze bevindingen impliceren dat het nuttig is om meer onderzoek te doen naar de vaardigheden die nodig zijn om op een effectieve en constructieve manier, met gebruikmaking van instrumenten, feedback te geven op de toepassing van een interventie.

Ten slotte wordt in **hoofdstuk 6** opiniërend ingegaan op het vraagstuk van hoe om te gaan met de integratie van verschillende ondersteuningssystemen in één organisatie. Het is immers onvoldoende om informatie te verschaffen over welke ondersteuning voor professionals er nodig is, zonder daarbij te kijken naar welke ondersteuning realistisch is. Jeugdzorgorganisaties worden namelijk geconfronteerd met een grote hoeveelheid aan (verplichte) ondersteuningssystemen behorende bij interventies. Doordat professionals vaak meerdere interventies uitvoeren, hebben ze te maken met meerdere ondersteuningssystemen, en worden daardoor geconfronteerd met een opeenstapeling van ondersteuningssessies. Het doel van het opiniërende artikel was daarom de discussie te beginnen over hoe ondersteuningssystemen gericht op de implementatie van interventies voor professionals kunnen worden geïntegreerd binnen jeugdzorgorganisaties.

Om het integreren van ondersteuning van professionals te bevorderen wordt voorgesteld de ondersteuning in eerste instantie te organiseren rond algemene, en overlappende factoren van interventies. De integratie houdt kort gezegd in dat de ondersteuning wordt georganiseerd rond competenties die alle professionals nodig hebben voor het goed uitvoeren van hun werk (de algemene factoren) en competenties die professionals nodig hebben voor het uitvoeren van meerdere interventies (de overlappende factoren die te maken hebben met de structuur en context van de interventie). Deze wijze van integratie van ondersteuning biedt de mogelijkheid om de ondersteuning efficiënter te organiseren, waardoor een opeenstapeling van ondersteuningssessies voor professionals wordt voorkomen. Daarnaast legt het een fundament voor een algeheel betere uitvoeringspraktijk. Efficiëntere organisatie van de ondersteuning aan professionals rond algemene, contextuele en structurele factoren van interventies vergroot de kans dat de uitvoering van deze factoren onder meer professionals wordt gestimuleerd. Het integreren van ondersteuning rond overlappende factoren is een geheel nieuwe manier om de kwaliteit van de zorg voor jeugd te stimuleren. Het komt tegemoet aan de wens van jeugdzorgorganisaties en professionals om het stapelen van ondersteuning en ondersteuningssessies te voorkomen en is daarmee een potentiële manier om de implementatiekloof op dit gebied te overbruggen. In tweede instantie is van belang te onderzoeken hoe deze ondersteuning kan worden gecombineerd met ondersteuning rond de specifieke competenties. Meer in het algemeen is het nodig dat het effect van het integreren van ondersteuning op de jeugdzorgorganisaties, professionals en cliënten verder wordt onderzocht.

Bijdragen van dit onderzoek

De vier studies en het opiniërende artikel in dit proefschrift dragen bij aan de kennis over succesvolle implementatie van interventies voor jeugdigen met externaliserende gedragsproblemen. Deze kennis is noodzakelijk voor het realiseren van meer effect van interventies voor kinderen en jongeren met gedragsproblemen, in Nederland en daarbuiten.

De voornaamste bevinding uit dit proefschrift is dat het goed toepassen van kennis over ‘wat werkt’ zoals uitgewerkt in *evidence-based* interventies, samenhangt met positieve uitkomsten voor jongeren. Deze bevinding is in lijn met eerdere onderzoeken (Lipsey, 2009; Tennyson, 2009) naar de relatie tussen behandelintegriteit en uitkomsten en versterkt deze door de gehanteerde striktheid omtrent de inclusie van studies. Een tweede belangrijke bevinding is dat een goede ondersteuning van professionals hen in staat stelt om interventies met een voldoende mate van behandelintegriteit uit te voeren. Deze bevinding onderstreept het belang van een goed functionerend ondersteuningssysteem waarin aandacht is voor behandelintegriteit. De ondersteuning van professionals moet

zich daarbij niet alleen richten op de daadwerkelijke behandeltrouw (*adherence*) aan de kernfactoren van de geboden zorg, maar ook en misschien zelfs nog meer op het vaardig, tijdig en goed inzetten van deze factoren (*competence*). Weisz et al. (2013b) stellen dat in de daadwerkelijke leefwereld van jeugdigen waarin zoveel factoren een rol spelen, de cliënt zijn of haar behoeften halverwege een behandeling kunnen veranderen, wat van een professional vereist dat hij of zij de focus verlegt. Dit vereist een professional die kan inspelen op de behoefte van de cliënt zonder daarbij af te wijken van de bedoelde interventie. De ondersteuning aan de professionals moet daarom gericht zijn op de toepassing van kennis en vaardigheden die noodzakelijk zijn om flexibel de werkzame factoren van een interventie in te zetten. Supervisoren en professionals bekwaam in het reflecteren op behandelintegriteit zijn hierbij essentieel.

De bevindingen impliceren dat expliciet wordt gemaakt wat de toepassing van een interventie behelst. Het is cruciaal dat er hierbij valide en betrouwbare instrumenten worden gebruikt en ontwikkeld om de mate van behandelintegriteit inzichtelijk te maken, zowel ten aanzien van *evidence-based* interventies als voor de gebruikelijke zorg. Alleen met valide en betrouwbare instrumenten kunnen er waardevolle uitspraken worden gedaan over de behandelintegriteit. Goede metingen van behandelintegriteit verschaffen belangrijke informatie voor de inhoud van training en ondersteuning aan professionals die verantwoordelijk zijn voor de uitvoering van de interventies.

Om effectief bij te dragen aan een continue verbetering van de hulpverlening dienen data over behandelintegriteit altijd in samenhang met uitkomsten (zoals afname van gedragsproblematiek) van cliënten te worden gebruikt. Dit maakt het mogelijk om te reflecteren op wat er daadwerkelijk aan hulp is ingezet en wat de resultaten daarvan zijn voor cliënten. Het onderzoeken van interventies in representatieve condities is daarin essentieel. Zoals gesteld door Weisz (2015) “Het zou een interessante paradox zijn wanneer onderzoek dat bedoeld is om de klinische praktijk te verbeteren, behandelingen produceert die niet passen bij – en misschien ook niet goed omgaan met – dezelfde klinische praktijk die ze voorgenomen hadden te verbeteren.” (letterlijke vertaling, p. 131). Het onderzoeken van interventies in situaties die representatief zijn voor de alledaagse praktijk zal meer valide bewijs opleveren over de effecten, werkzame mechanismen en factoren die het effect modereren. Het resultaat zal zijn dat de kennis die deze onderzoeken opleveren beter toepasbaar zal zijn in die alledaagse praktijk (Weisz, 2015).

Slotwoord

De studies in dit proefschrift illustreren de complexiteit van en uitdagingen rondom het toepassen van kennis van ‘wat werkt’. Tegelijkertijd laten de studies ook de toenemende kennis in het veld zien en de potentie om de implementatiekloof tussen de kennis over wat er werkt en de toepassing en het gebruik van deze kennis in de praktijk, op bepaalde delen te overbruggen. Omdat maatschappelijke omstandigheden zullen veranderen, zullen implementatieprocessen veranderen. Dit betekent dat er bij voortdurende inspanningen nodig zijn om op een effectieve manier effectieve zorg te implementeren. Het samenwerken en delen van kennis in binnen en buitenland, tussen wetenschappelijke domeinen en dienstverleningssectoren is essentieel om een beter begrip van dit onderwerp te krijgen. De Global Implementation Initiative (GII) en European Implementation Collaborative (EIC) alsmede de nationale netwerken die zich bezig houden met implementatie, zijn goede voorbeelden van gezamenlijke inspanningen voor een grotere sociale impact van effectieve zorg. Om meer effect met effectieve interventies te bereiken, is het overbruggen van de implementatiekloof essentieel. Ik hoop met dit proefschrift een bijdrage te hebben geleverd aan de kennis die noodzakelijk is om de implementatiekloof te overbruggen en daarmee bij te dragen aan het laten werken van ‘wat werkt’ voor kinderen en jongeren met gedragsproblemen en hun families in Nederland, en daarbuiten.

Dankwoord

Promoveren is doorzetten, het is vallen, opstaan en met goede moed weer doorgaan. Dat kan je niet alleen, en dat heb ik gelukkig ook niet alleen hoeven doen. Daarom wil ik op deze plek iedereen bedanken die me de afgelopen jaren heeft gesteund tijdens het werken aan dit proefschrift.

Leonieke Boendermaker, mijn leidinggevende bij de Hogeschool van Amsterdam en co-promotor, ik ben je waanzinnig dankbaar voor alles wat je in gang hebt gezet om mij deze kans te geven. Ondanks alles ben je altijd achter me blijven staan en dat vertrouwen waardeert ik enorm. Jij hebt me laten ontwikkelen van een jonge onderzoeker tot waar ik nu ben. Dit had ik niet zonder jou kunnen doen, dank je wel.

Tom van Yperen, mijn promotor. Ik weet niet of je je realiseert hoe erg ik me gezegend voel dat jij mijn promotor bent geweest. Je bent eerlijk, scherp, slim en je immer bewust van de praktijk. Ik heb genoten van de samenwerking en hoop dat we die in de toekomst blijven voortzetten.

Daarnaast dank aan de leden van de leescommissie, Prof. dr. Erik Knorth, Prof. dr. Menno Reijneveld en Prof. dr. Geert-Jan Overbeek, voor jullie bereidheid en tijd om dit proefschrift te lezen en te beoordelen.

In het bijzonder wil ik ook Germie van den Berg bedanken. Als projectmedewerker bij het NEJA heb ik je leren kennen en sindsdien ben je mijn inspirator. Jij wist al dat ik ging promoveren voordat ik dat zelf wist. Je hebt altijd in me geloofd en ik hoop dat je trots bent op dit eindresultaat.

Altra Onderwijs & Jeugdhulp, Jeugdbescherming Regio Amsterdam en Spirit Jeugd & Opvoedhulp, zonder onze samenwerkingsprojecten was dit proefschrift niet tot stand gekomen. Ik ben geweldig trots op wat we allemaal al hebben gedaan en nog steeds aan het doen zijn omtrent leren op de werkvloer. Ik wil ook alle professionals die mee hebben gewerkt bedanken, jullie input is onmisbaar geweest. Sanne Pronk, Inge Busschers, Carolien Konijn en alle versterktrouwen, nergens leer ik meer dan met jullie te brainstormen. Lets keep up the good work!

Medeauteurs van de meta-analyse, Geert-Jan Stams, Machteld Hoeve en Mark Assink, dank voor jullie inzet bij deze studie. Mijn doorzettingsvermogen is nog nooit zo op de proef gesteld. Door jullie overtuiging, vastberadenheid en kennis is het echt een studie geworden om trots op te zijn.

Collega's van de Hogeschool van Amsterdam en in het bijzonder van het lectoraat 'Kwaliteit en Effectiviteit in de Zorg voor Jeugd', wat een gedeelde vreugde dit proefschrift. Altijd kon ik terecht voor een praatje of een brainstorm, soms vanaf de andere kant van de wereld maar altijd dichtbij. Dank voor de fijne samenwerking!

Initiatiefgroep en leden van het Nederlands Implementatie Collectief, de uitwisselingen over implementatie hebben me scherp gehouden op deze materie en maken me immer enthousiast. Ik geloof er heilig in dat je met gebundelde krachten het verst komt. Dank jullie wel dat we dit samen doen.

My dear fellow board members of the European Implementation Collaborative, Bianca Albers, Stella Owens, Knut Taraldsen and Tom Jefford, you are a true inspiration to me. I hope we can continue to contribute in sharing implementation knowledge for a substantive amount of time.

Lieve vrienden uit Amsterdam, Den Haag en al die andere plekken op deze wereld, zonder jullie was ik gek geworden en dankzij jullie ben ik gestoord. Jullie hebben allemaal op jullie eigen manier een bijdrage geleverd aan dit proefschrift. De wintersporten, bootcampsessies, etentjes, boottrippjes, immer nice days festivals en vrijdagmiddag-, zondagmiddag- en wat dan ook voor middagborrels, zijn een goede uitlaatklep geweest. Ik heb jullie waanzinnig lief, dank jullie wel dat jullie er altijd voor me zijn.

Lieve mam, pap en Lau. It ain't over till the fat lady sings, en mijn hemel wat hebben jullie het af en toe met mij te verduren gehad. Altijd zijn jullie er voor me geweest, hebben jullie geluisterd en wijze raad gegeven. Dit is niet alleen mijn mijlpaal, dit is onze mijlpaal. Schwesternin, of ik nou achter je aan sjeesde op jouw/mijn snowboard, happy hour met jou en de familie aan het vieren was op Curaçao of gewoon lekker met Marti en jou op de bank hing, die momenten samen relativeerden de hele boel weer bij elkaar. Pa dank voor je altijd wijze woorden, scherpe argumentatie en goede raad. Ik ben niet alleen een wereldreiziger maar vooral een doorzetter omdat jij dat ook bent. Lieve Mam, jij bent het lichtpunt van mijn leven, onbevooroordeeld, lief en met een tomeloos vertrouwen. Door jou ken ik mezelf.

Nee ik ben je niet vergeten, ik zal je nooit vergeten. Want al ben je soms uit mijn oog, je bent nooit uit mijn hart. Lieve Patrick dank je wel dat je in mij en in ons gelooft, mi ta stima bo hopi hopi veel.

About the author



Pauline Goense was born on April the 2nd, 1986 in 's-Gravenhage, The Netherlands. After completing her secondary school at Maerlant Lyceum The Hague in 2004 and a year of travelling through Australia, she studied Behavior and Society at the University of Amsterdam. In 2008 she obtained her bachelor's degree and enrolled in the Master Program Studies of Forensic Child and Youth Care Sciences at the same university. During this Master Program she spent half a year in the Dutch Caribbean for her research internship at a juvenile justice care facility on the island of Curacao. In 2009 Pauline obtained her Master Degree

and enrolled in the transition program Public Law which enabled her to enroll in the Master Program Public Law, curriculum Criminal Law the year after. In 2011 she obtained the Master Degree in Criminal Law at the University of Amsterdam. Between 2010 and 2012 Pauline was also the project coordinator of the Netwerk Effectieve Jeugzorg Amsterdam (NEJA), worked as a (substitute) youth care worker in a closed youth care facility for children with severe behavioral problems (de Koppeling) and worked as a junior policy advisor at the department 'Research and Development' of Spirit, a youth care organization in Amsterdam.

Since 2011 Pauline holds a position as a researcher and lecturer at the Amsterdam University of Applied Sciences. She works in the research group 'Implementation and Effectiveness of Youth Care Services'. A research proposal that was granted by Stichting Innovatie Alliantie (SIA) in 2012 kick started the PhD project of Pauline which was later (2013) made official in collaboration with the University of Groningen. Pauline is co-initiator of the Netherlands Implementation Collaborative (NIC) and board member of the European Implementation Collaborative (EIC).

List of publications

International journals

- Goense, P.B., Assink, M., Stams, G.J.J.M., Boendermaker, L., & Hoeve, M. (2016). *Making 'What Works' Work: A meta-analytic study of the effect of treatment integrity on outcomes of evidence-based interventions for juveniles with antisocial behavior*. Provisionally accepted for publication.
- Goense, P.B., Boendermaker, L., & van Yperen, T. (2016). *Measuring Treatment Integrity: Use of and experience with measurements in routine youth care practice settings*. Manuscript submitted for publication.
- Goense, P.B., Boendermaker, L., & van Yperen, T. (2016). *Stimulating quality of service delivery in mental health care settings for youth. Suggestions for integrating support systems*. Manuscript submitted for publication.
- Goense, P.B., Boendermaker, L., & van Yperen, T. (2015b). Support Systems for Treatment Integrity. *Research on Social Work Practice*, 1-5.
- Goense, P., Boendermaker, L., van Yperen, T., Stams, G.J.J.M., Laar, J. (2014). Implementation of Treatment Integrity Procedures: An Analysis of Outcome Studies of Youth Interventions Targeting Externalizing Behavioral Problems. *Zeitschrift für Psychologie*, 222, 12–21.

Books

- Goense, P.B., Pronk, S., Boendermaker, L., Bakker, R., Ruitenbergh, I., Bertling, L. (2015b). *Leren op de werkvloer. Organisatie en inhoud van leren op de werkvloer*. Delft: Eburon.

Book chapters

- Boendermaker, L., & Goense, P.B. (*accepted*). Meten van behandelintegriteit als basis voor effectiviteit. In J.W. Veerman & T. van Yperen (Eds.), *Zicht op Effectiviteit* (2th ed.). Manuscript accepted for publication.
- Boendermaker, L., & Goense, P.B. (*accepted*). Uitvoeren van interventies in de jeugdzorg. In M. Berk, A. Hoogenboom, M. de Kleermaeker & K. Verhaar (Eds.), *De jeugdprofessional in ontwikkeling. Handboek voor professionals in het jeugd domein* (2th ed.). Manuscript accepted for publication.

Rapports

- Boendermaker, L., Boomkens, C., Goense, P.B. & Steffens, H. (2012). *De implementatie van MDFT bij Spirit in Amsterdam. Procevaluatie over de periode 2010-2011*. Amsterdam: HvA / Lectoraat Implementatie in de jeugdzorg.
- Goense, P.B., Pronk, S., Boendermaker, L., & Berentsen, J. (2014). *Wie helpt de hulpverlener? Ondersteuning bij het uitvoeren van interventies in de jeugdzorg. Rapportage deelonderzoek 1*. Amsterdam: Hogeschool van Amsterdam.

Other

- Goense, P.B. (2013). *ITIPS-A, manual for coding studies on implementation of procedures for treatment integrity, adapted version*. Amsterdam University of Applied Sciences (Internal publication).
- Goense, P.B. (2013). *Tevredenheidsvragenlijst ondersteuning jeugdzorgprofessionals*. Amsterdam: Hogeschool van Amsterdam: Lectoraat Kwaliteit en Effectiviteit in de Zorg voor Jeugd.
- Goense, P.B., & Ruitenbergh, I. (2015). *Toolkit Videofeedback in de Zorg voor Jeugd*. Delft: Eburon.
- Konijn, C., Brussen, A., van Tol, T., Zuurveld, P., & Goense, P. (2013). *Move2Learn/Groep. Sociale en cognitieve vaardigheidstraining voor een groep jongeren van 12-18 jaar. Trainershandleiding*. Amsterdam: Spirit.
- Meershoek, K., Tarée, R., Ruitenbergh, I, van Andel, H., Goense, P.B. (2015, juni 3). *Wie helpt de hulpverlener? Instructiefilm opnemen video-materiaal* [Video file]. Available from <https://www.youtube.com/watch?list=PLqQz5Oy8HYpOWf9YE4sa8u7zp1YZHB1GJ&v=C8rPIfm-VFqo>

International conference presentations¹⁵

September 5th 2012

Treatment Integrity in Juvenile Intervention Research: Analysis of the construct.

European Scientific Association on Residential & Family Care for Children and Adolescents (EUSARF), Glasgow, Scotland.

February 6th 2013

Treatment Integrity in Youth Intervention Outcome Studies: Analysis of the implementation of treatment integrity procedures.

Nordic Conference on Implementation of Evidence-based Practice, Linköping, Sweden.

September 18th 2014

Delivery of an intervention: A framework for the measurement of treatment integrity.

Biennial Australian Implementation Conference, Sydney, Australia.

February 4th 2015

Effective support systems for treatment integrity.

Nordic Conference on Implementation of Evidence-based Practice, Bergen, Norway.

27th May 2015

Measuring treatment integrity: use of and experiences with instruments in everyday settings.

Global Implementation Conference, Dublin, Ireland.

¹⁵ Only presentations of which the content was related to this dissertation are included

