university of groningen

University Medical Center Groningen

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.

Download date: 16-03-2025

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, Hoeve, & 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 trial), 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
Total	*60*	*54*	*90*

[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
Total		*54*

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 supervis`ion. 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. Is 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.