

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*. 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).

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.

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.