The role of therapist adherence within multisystemic therapy

FROM IMPLEMENTATION TO ALLIANCE

Aurelie Lange

From implementation to alliance:

The role of therapist adherence within Multisystemic Therapy

Aurelie Lange

This thesis has been facilitated by an unrestricted grant of ZonMW and with support of de Viersprong.

ISBN 978-94-6299-798-1
Design of the cover: Yannick Lange
Printing: Ridderprint, the Netherlands
Layout Design: Ridderprint, the Netherlands

From Implementation to Alliance
The role of therapist adherence within Multisystemic Therapy

Van implementatie tot alliantie: De rol van modeltrouw binnen Multi Systeem Therapie

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.A.P Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
Woensdag 31 januari 2018 om 13.30 uur

door

Aurelie Marc Caroline Marie Josephe Ghislaine Lange
geboren te Berlijn, Duitsland

Erasmus University Rotterdam

Promotiecommissie

Promotoren:
Prof.dr. J.J. van Busschbach
Prof.dr. R.H.J. Scholte

Overige leden:
Prof.dr. N.C.L. Mulder
Prof.dr. J.J. Asscher
Prof.dr. G.J. Overbeek

Copromotor:
Dr. R.E.A. van der Rijken

Content

Chapter 1

General introduction

"Without data on implementation, research cannot document precisely what program was conducted, or how outcome should be interpreted" (Durlak & Dupre, 2008, p. 340).

Monitoring to what extent an intervention or treatment is actually being delivered to clients is crucial when studying the effectiveness of the intervention. For this very reason, researchers studying the effectiveness of specific treatments started to assess whether therapists provided the treatment according to the model, thus with high therapist adherence (Schoenwald, 2011). The researchers wanted to make sure they could interpret the results of their evaluation studies as evidence for the effectiveness of the intervention. In subsequent efforts to disseminate and implement evidence-based interventions in new settings, therapist adherence evolved into a salient indicator of successful implementation (Durlak & DuPre, 2008; Fixsen, Naoom, Blase, Friedman, & Wallace, 2005; McLeod et al., 2013; Schoenwald & Garland, 2013).

Therapist adherence: the degree to which the therapist delivers the specified components of the intervention as intended (McLeod, Southam-Gerow, Tully, Rodríquez, & Smith, 2013; Perepletchikova & Kazdin, 2005).

Yet, implementing evidence-based interventions into clinical practice has proven challenging. Studies of evidence-based interventions conducted within everyday practice tend to achieve smaller effects than highly controlled evaluation studies (Henggeler 2011; Weisz, Ugueto, Cheron, & Herren, 2013). The assumption is that these smaller effects can be attributed to a lack of adherence. In other words, it is assumed these interventions consist of specific elements or techniques (i.e., specific factors) that are responsible for the treatment effects. Failure to implement or deliver these specific factors will lead to poorer treatment outcomes.

In contrast to specific factors, there are researchers who promote the primacy of common factors as the key to achieve positive clinical outcomes (e.g., Messer & Wampold, 2002; Sparks & Duncan, 2010). These scholars stress that it does not matter which intervention is employed, but instead, that therapists achieve positive behavioral outcomes through the general processes and factors underlying all psychotherapy, such as the working relationship (alliance), therapist allegiance (therapist's believe in the efficacy of the treatment), client motivation and hope for change, or therapist skills (Messer & Wampold, 2002; Sprenkle & Blow, 2004).

Alliance: the working relationship between the client and the therapist, which consists of the emotional bond and the agreement on the goals and tasks of treatment (Bordin, 1979; Hougaard, 1994).

Currently, many would agree that common therapeutic factors are needed in combination with specific techniques (e.g., Sexton & Kelley, 2010; Sprenkle & Blow, 2004).

However, few have attempted to study how these two components interrelate during treatment and jointly lead to outcomes. This dissertation addresses this gap. The aims are twofold. Firstly, we evaluate what factors affect therapist adherence scores when disseminating evidence-based interventions. Secondly, we assess the unique and joint role of adherence and alliance within treatment. These two aims will be discussed in more detail below.

I. What's in a score? Evaluation of factors affecting reliable assessment of therapist adherence after cross-national dissemination of an evidence-based intervention

Implementation refers to all activities designed to put the intervention into practice and ensure the intervention sustains through time in an ever changing environment (e.g., changes in staff, funding streams, or leaders; Fixsen et al., 2005). Studies have shown that failure to adequately implement an intervention can have detrimental effects on treatment outcomes (Durlak & Dupre, 2008; Sexton & Turner, 2010). A synthesis of research studies suggests that at least 60% of an evidence-based intervention should be implemented as intended to achieve desirable outcomes (Durlak & Dupre, 2008).

Therapist adherence is often one of the core measures to evaluate the success of the implementation of an intervention. Yet, despite this central role of therapist adherence measures, there is no consensus on how to assess adherence. For example, adherence instruments for evidence-based youth interventions targeting disruptive behavioral problems showed so much diversity that it proved impossible to classify them among meaningful dimensions (Schoenwald, Garland, Southam-Gerow, Chorpita, & Chapman, 2011). The content of the items obviously varied between interventions. However, the authors also noted marked differences regarding the level of detail of the items, the focus of the items (e.g., context, material, therapist behavior, or client behavior), the manner in which the items were scored (e.g., amount/duration, absence/presence, or quality of behavior), the moment of assessment (e.g., each session, or at specific time intervals), and by whom therapist adherence was scored. Moreover, the available instruments often lacked information on their psychometric properties (i.e., reliability and validity) as well as on their relationship with treatment outcomes, the latter being the primary reason why adherence is believed to be so important (Schoenwald & Garland, 2013).

The Therapist Adherence Measure (TAM-R) of Multisystemic Therapy (MST) is unique in the sense that it is one of the few instruments that *has* proven to be reliable and valid and to relate to positive treatment outcomes up to four years post-treatment (Huey, Henggeler, Brondino, & Pickrel, 2000; Schoenwald, Chapman, Sheidow, & Carter, 2009). MST is an evidence-based, intensive home- and community-based intervention for 12 to 18

years old adolescents with antisocial and/or delinquent behavioral problems (Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 2009). To ensure successful dissemination and implementation of MST at other sites, an elaborate quality-assurance system was developed, consisting, among other things, of training, supervision and consultation, and continuous data-monitoring (Henggeler & Schoenwald, 1999; Schoenwald, 2008). The TAM-R is a central instrument within this quality-assurance system, as the therapist is the primary linkage with the family and therefore critical in achieving desired outcomes for youths and families (Henggeler & Schoenwald, 1999).

Notwithstanding the amount of research on the TAM-R, findings regarding the adherence-outcome association were less consistent in European countries than in the United States (US), where MST was developed in the mid-1970s. While some European countries reported a positive adherence-outcome association (Ogden & Halliday-Boykins, 2004), others were unable to find an association (Butler, Baruch, Hickey, & Fonagy, 2011) or reported mixed findings (Sundell et al., 2008). Moreover, adherence-scores were systematically lower in other countries compared to the US (MST Institute, 2010), and this also applied to the Netherlands (based on Dutch dashboards). This was echoed in concerns expressed by Dutch therapists regarding the validity of the TAM-R. Thus, both researchers and therapists started to question what the TAM-R was actually measuring in Europe: 'What's in a score?'. Do these European adherence-scores reflect more than just therapist adherence? This thesis aimed to address this question from two different perspectives.

Basically, differences between scores collected in different countries can have two different causes. On the one hand, the observed differences between the US and other European countries may result from issues related to the instrument itself. Scholars have pointed to the necessity to assess equivalence of an instrument after translation and dissemination to another cultural context (Byrne & Van de Vijver, 2010; Van de Vijver & Tanzer, 2004). In other words, adherence-scores from different countries should be interpretable in the same way. One cannot a priori assume the American TAM-R to be equally valid and reliable in another country.

Equivalence: the absence of bias. Differences on the indicator (in this instance: the TAM-R) correspond to score differences on the underlying trait across groups (in this study: therapist adherence; Van de Vijver & Tanzer, 2004).

On the other hand, these cross-national differences may pertain to the delivery of the treatment. Thus, the differences may also reflect true differences in adherence. A recent study in MST provided evidence that could account for lower adherence in countries outside the US (Löfholm, Eichas, & Sundell, 2014). They showed that therapist adherence was influenced by the amount of experience with delivering the intervention. As MST has been developed in the US, there was inevitably more experience with providing MST in the US than in other countries providing this intervention. These two perspectives, lack of

equivalence or true differences, led to the following **research questions**:

1. Is the Dutch TAM-R equivalent to the original American version?
2. How does experience affect therapist adherence?

The answers to these research questions will provide some insight into factors affecting therapist adherence scores after cross-national dissemination. By focusing on the instrument itself and on the delivery of the intervention, the topic will be addressed from different perspectives, thus providing a more comprehensive understanding.

II. People or protocol: The unique and joint contribution of adherence and alliance to optimal treatment outcomes

Theoretically, adherence to an evidence-based intervention would be expected to relate to treatment outcomes, as the key ingredients of an evidence-based intervention should be delivered as intended to achieve the desired treatment outcome (Mihalic, 2004). There are good examples of this positive association (e.g., Robbins et al., 2011; Schoenwald et al., 2009; see Goense, Assink, Stams, Boendermaker, & Hoeve, 2016 for a meta-analysis). Yet, given the lack of reliable and valid instruments, the lack of congruence on how adherence should be measured, and the fact that the intervention should be effective in the first place, null-findings have also been reported (e.g., Weck et al., 2013; see Webb, DeRubeis, & Barber, 2010 for a meta-analysis).

Some scholars, therefore, emphasize common factors rather than specific factors to explain the effectiveness of treatments. The most well-known common factor within psychotherapy probably is alliance. Both practitioners and scholars agree that alliance is relevant for achieving positive outcomes (Friedlander, Escudero, Heatherington, & Diamond, 2011; Karver, Handelsman, Fields, & Bickman, 2006; Stamoulos et al., 2016). Although some have posited that alliance itself is curative, most believe alliance to be one of the factors needed in a successful treatment (McLeod et al., 2013; Shirk, Karver, & Brown, 2011).

Several models have described how alliance and the specific ingredients may collaborate to lead to optimal outcomes. A strong alliance may motivate and encourage the client, whereas adherent implementation of the specific ingredients may provide the actual tools and techniques to foster therapeutic change. On top of their individual contribution, alliance and adherence may enhance one another: Whereas a strong alliance may be a precondition for the adherent implementation of the intervention, adherence may foster confidence in the therapist's skills, and thereby deepen the client–therapist alliance (Goldfried & Davila, 2005; Hill, 2005).

Nonetheless, concerns have been raised regarding adherence to evidence-based interventions (Gyani, Shafran, Rose, & Lee, 2015; Nelson, Steele, & Mize, 2006). Some practitioners have pointed to the inflexibility of treatment protocols and have stressed the

need to adapt to the unique situation of each client. Others have expressed their fear that evidence-based protocols may hamper the therapeutic relationship, instead of enhancing it, as proposed by the above models (Gyani, et al., 2015; Nelson et al., 2006). Thus, practitioner clinical expertise and theoretical models seem to contradict one another.

Unfortunately, few studies so far have actually studied how alliance and adherence interrelate, and the available evidence provides contradictory findings. Some studies did not find any association between alliance and adherence (Hukkelberg & Ogden, 2013; The Multisite Violence Prevention Project, 2014), whereas others showed alliance to predict adherence (Tschuschke et al., 2015; Weck, Grikscheit, Jakob, Höfling, & Stangier, 2015). In the absence of a strong alliance, a rigid focus on adherence may either lead to further deterioration of the alliance and interfere with therapeutic change (Barber et al., 2006; Castonguay, Goldfried, Wiser, Raue, & Hayes, 1996), or may 'save' a treatment with low alliance, leading to positive treatment outcomes (Webb et al., 2012).

One of the weaknesses of most previous studies is that they assessed alliance and adherence on only one occasion, thereby ignoring the changing and dynamic nature of both aspects. Both alliance and adherence can change over time and evidence suggests that these changes are related to treatment outcomes (Chiapa et al., 2015; Owen, Miller, Seidel, & Chow, 2016; Robbins et al., 2011; Stiles et al., 2004). Hence, the way in which they interrelate and uniquely or jointly contribute to outcome may also change over time.
Thus, the second aim of this dissertation is to assess the unique and joint role of adherence and alliance within treatment using a longitudinal approach to the assessment of alliance and adherence. To address this aim the following **research questions** will be answered:

3. How do adherence and alliance interrelate over time during treatment?
4. How do adherence and alliance develop over time during treatment?
5. How do adherence and alliance uniquely and jointly contribute to short- and long-term treatment outcomes?

The answers to these questions will help to develop a better understanding of the role of alliance and adherence within treatment, which may inform clinical practice and help practitioners effectively combine the two to reach optimal treatment outcomes.

Study Design and Samples

For this dissertation, routinely collected data from Dutch organizations providing MST was used. As described above, MST has integrated an elaborate quality-assurance system, including monitoring of adherence on a monthly basis and assessing the MST ultimate outcomes (i.e., no out-of-home placement, no police contact, and adolescent attending school or work) at the end of treatment. As such, all organizations providing MST in the Netherlands are required to report on these aspects for all cases on the MST Institute website (MSTi; www.MSTInstitute.org). In addition to this compulsory data-collection, the

Dutch Network Partner, MST-the Netherlands, has advised Dutch organizations providing MST to take part in a more elaborate routine outcome monitoring procedure, collecting not only the information required for the MST Institute website, but also a broader range of measures including information on client characteristics, behavioral problems, and parental stress at the start and end of the treatment, and treatment outcomes at six, twelve, and eighteen months post-treatment. This information is stored in an online database (BergOp; www.bergop.net).

Client characteristics and the compulsory MST treatment outcomes are reported by the therapist, after consultation with the primary caregiver and other sources of information if needed. Primary caregivers are requested to complete a questionnaire on child behavioral problems and parental stress at the start and end of treatment. Therapist adherence and post-treatment outcomes are collected through telephone interviews with the primary caregiver participating in MST. Although MST-the Netherlands advises organizations to make use of the independent call center Kwestion, specialized in collecting these data for MST, organizations are free to choose how they collect these data, as long as they are not collected by the therapists themselves.

Data collection for the MST Institute website started at the implementation of MST in the Netherlands in 2005. Initially, 4 Dutch teams, distributed over 2 organizations, started providing MST. The additional data collection started after foundation of MST-the Netherlands in 2008. The number of Dutch teams and organizations changed over time, including, for example, 24 teams distributed over 6 organizations in 2014. The participating organizations, the period of data collection, and the instruments used in this dissertation varied depending on the differing purposes of the studies described. For all studies, the organizations were asked for consent to share their data.

Outline of this Dissertation

In the following section the outline of this dissertation will be described. In **Chapter 2** the first research question is answered (i.e. the equivalence of the Dutch TAM-R to the original US TAM-R). Equivalence can be described as the opposite of bias and is a prerequisite for score comparison across cultural or national groups (Van de Vijver & Tanzer, 2004). **Chapter 3** looks into the second research question by exploring how experience affects therapist adherence scores. For this purpose, a Swedish study was replicated, which tested the associations between the experience with the MST treatment model at different levels (therapist, team, and country-wide), therapist adherence, and treatment outcomes.

To study the unique and joint role of adherence and alliance and answer research question 3, 4 and 5 **Chapter 4** first describes a study to assess the underlying factor structure of the Dutch TAM-R. Previous research suggested that the TAM-R may consist of multiple factors. The chapter shows that the Dutch TAM-R consists of two aspects: Therapist

adherence and the client-therapist working alliance. Chapter 5 and 6 subsequently use these factors of the TAM-R to evaluate the unique and joint contribution of adherence and alliance to outcome. **Chapter 5** addresses research question 4 by detailing how adherence and alliance interrelate during treatment. In **Chapter 6** research question 4 and 5 are answered as we investigate how therapist adherence and alliance develop over time and how this development is uniquely and jointly related to short- and long-term treatment outcomes. We close, in **Chapter 7**, with a summary of and reflection on the main findings, practical implications, limitations, and recommendations for future research.

References

Barber, J. P., Gallop, R., Crits-Christoph, P., Frank, A., Thase, M. E., Weiss, R. D., & Gibbons, M. B. C. (2006). The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: Results from the National Institute Drug Abuse Collaborative Cocaine Treatment Study. *Psychotherapy Research, 16*, 229-240.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research, and Practice, 16*, 252-260.

Butler, S., Baruch, G., Hickey, N., & Fonagy, P. (2011). A randomized controlled trial of Multisystemic Therapy and a statutatory therapeutic intervention for young offenders. *Journal of the American Academy of Child & Adolescent Psychiatry, 50*, 1220-1235.e2.

Byrne, B. M., & van de Vijver, F. J. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing, 10,* 107-132.

Castonguay, L. G., Goldfried, M. R., Wiser, S., Raue, P. J., & Hayes, A. M. (1996). Predicting the effect of cognitive therapy for depression: A study of unique and common factors. *Journal of Consulting and Clinical Psychology, 64*, 497-504.

Chiapa, A., Smith, J. D., Kim, H., Dishion, T. J., Shaw, D. S., & Wilson, M. N. (2015). The trajectory of fidelity in a multiyear trial of the family check-up predicts change in child problem behavior. *Journal of Consulting and Clinical Psychology, 83*, 1006-1011.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327-350.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature.* Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Friedlander, M. L., Escudero, V., Heatherington, L., & Diamond, G. M. (2011). Alliance in couple and family therapy. *Psychotherapy, 48*, 25-33.

Goense, P. B., Assink, M., Stams, G-J., Boendermaker, L., & Hoeve, M. (2016). Making 'what works' work: A meta-analytic study of the effect of treatment integrity on outcomes of evidence-based interventions for juveniles with antisocial behavior. *Aggression and Violent Behavior, 31*, 106-115.

Goldfried, M. R., & Davila, J. (2005). The role of relationship and technique in therapeutic change. *Psychotherapy: Theory, Research, Practice, Training, 42*, 421-430.

Gyani, A., Shafran, R., Rose, S., & Lee, M. J. (2015). A qualitative investigation of therapists' attitudes towards research: Horses for courses? *Behavioural and Cognitive Psychotherapy, 43*, 436-448.

Henggeler, S. (2011). Efficacy studies to large-scale transport: The development and validation of Multisystemic Therapy programs. *Annual Review of Clinical Psychology, 7*, 351-381.

Henggeler, S. W., & Schoenwald, S. K. (1999). The role of quality assurance in achieving outcomes in MST programs. *Journal of Juvenile Justice and Detention Services, 14*, 1-17.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (2009). *Multisystemic therapy for antisocial behavior in children and adolescents.* New York: Guilford.

Hill, C. E. (2005). Therapist techniques, client involvement, and the therapeutic relationship: Inextricably intertwined in the therapy process. *Psychotherapy: Theory, Research, Practice, Training, 42*, 431-442.

Hougaard, E. (1994). The therapeutic alliance - A conceptual analysis. *Scandinavian Journal of Psychology, 35*, 67-85.

Huey, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in Multisystemic Therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology, 68*, 451-467.

Hukkelberg, S. S., & Ogden, T. (2013). Working alliance and treatment fidelity as predictors of externalizing problem behaviors in parent management training. *Journal of Consulting and Clinical Psychology, 81*, 1010-1020.

Karver, M. S., Handelsman, J. B., Fields, S., & Bickman, L. (2006). Meta-analysis of therapeutic relationship variables in youth and family therapy: The evidence for different relationship variables in the child and adolescent treatment outcome literature. *Clinical Psychology Review, 26*, 50-65.

Löfholm, C. A., Eichas, K., & Sundell, K. (2014). The Swedish implementation of Multisystemic Therapy for adolescents: Does treatment experience predict treatment adherence? *Journal of Clinical Child & Adolescent Psychology, 43*, 643-655.

McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodriguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. *Clinical Psychology: Science & Practice, 20*, 14-32.

Messer, S. B., & Wampold, B. E. (2002). Let's face facts: Common factors are more potent than specific therapy ingredients. *Clinical Psychology: Science & Practice, 9*, 21-25.

Mihalic, S. (2004). The importance of implementation fidelity. *Report on Emotional & Behavioral Disorders in Youth, 4*, 83-105.

MST Institute (2010). *2010 MST Data Report*. Retrieved July 2015 from http://www.mstinstitute.org.

Nelson, T. D., Steele, R. G., & Mize, J. A. (2006). Practitioner attitudes toward evidence-based practice: Themes and challenges. *Administration and Policy in Mental Health and Mental Health Services Research, 33*, 398-409.

Ogden, T. & Halliday-Boykins, C. A. (2004). Multisystemic treatment of antisocial adolescents in Norway: Replication of clinical outcomes outside of the US. *Child and Adolescent Mental Health, 9*, 77-83.

Owen, J., Miller, S. D., Seidel, J., & Chow, D. (2016). The working alliance in treatment of military adolescents. *Journal of Consulting and Clinical Psychology, 84*, 200-210.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change. Issues and research recommendations. *Clinical Psychology: Science and Practice, 12*, 365-383.

Robbins, M. S., Feaster, D. J., Horigian, V. E., Puccinelli, M. J., Henderson, C., & Szapocznik, J. (2011). Therapist adherence in Brief Strategic Family Therapy for adolescent drug abusers. *Journal of Consulting and Clinical Psychology*, *79*, 43-53.

Schoenwald, S. K. (2008). Toward evidence-based transport of evidence-based treatments: MST as an example. *Journal of Child and Adolescent Substance Abuse, 17*, 69-91.

Schoenwald, S. K. (2011). It's a bird, it's a plane, it's fidelity measurement in the real world. *Clinical Psychology, Science and Practice, 18,* 142-147.

Schoenwald, S. K., Chapman, J. E., Sheidow, A. J., & Carter, R. E. (2009). Long-term youth criminal outcomes in MST transport: The impact of therapist adherence and organizational climate and structure. *Journal of Clinical Child & Adolescent Psychology, 38*, 91-105.

Schoenwald, S. K. & Garland, A. F. (2013). A review of treatment adherence measurement methods. *Psychological Assessment, 25*, 146-156.

Schoenwald, S. K., Garland, A. F., Southam-Gerow, M. A., Chorpita, B. F., & Chapman, J. E. (2011). Adherence measurement in treatments for disruptive behavior disorders: Pursuing clear vision through varied lenses. *Clinical Psychology: Science & Practice, 18*, 331-341.

Sexton, T. L. & Kelley, S. D. (2010). Finding the common core: Evidence-based practices, clinically relevant evidence, and core mechanisms of change. *Administration and Policy in Mental Health and Mental Health Services Research, 37*, 81-88.

Sexton, T. & Turner, C. W. (2010). The effectiveness of functional family therapy for youth with behavioral problems in a community practice setting. *Journal of Family Psychology, 24*, 339-348.

Shirk, S. R., Karver, M. S., & Brown, R. (2011). The alliance in child and adolescent psychotherapy. *Psychotherapy, 48*, 17-24.

Sparks, J. A. & Duncan, B. L. (2010). Common factors in couple and family therapy: Must all have prizes? In Duncan, B. L., Miller, S. D., Wampold, B. E., & Hubble, M. A. (Eds.), *The heart and soul of change: Delivering what works in therapy, 2nd ed.* (pp. 357-391). Washington, DC, US: American Psychological Association.

Sprenkle, D. H. & Blow, A. J. (2004). Common factors and our sacred models. *Journal of Marital and Family Therapy, 30*, 113-129.

Stamoulos, C., Trepanier, L., Bourkas, S., Bradley, S., Stelmaszczyk, K., Schwartzman, D., & Drapeau, M. (2016). Psychologists' perceptions of the importance of common factors in psychotherapy for successful treatment outcomes. *Journal of Psychotherapy Integration, 26*, 300-317.

Stiles, W. B., Glick, M. J., Osatuke, K., Hardy, G. E., Shapiro, D. A., Agnew-Davies, R., Rees, A., & Barkham, M. (2004). Patterns of alliance development and the rupture-repair hypothesis: Are productive relationships U-shaped or V-shaped? *Journal of Counseling Psychology, 51*, 81-92.

Sundell, K., Hansson, K., Löfholm, C. A., Olsson, T., Gustle, L-H, & Kadesjö, C. (2008). The transportability of Multisystemic Therapy to Sweden: Short-term results from a randomized trial of conduct-disordered youths. *Journal of Family Psychology, 22*, 550-560.

The Multisite Violence Prevention Project (2014). Implementation and process effects on prevention outcomes for middle school students. *Journal of Clinical Child & Adolescent Psychology, 43*, 473-485.

Tschuschke, V., Crameri, A., Koehler, M., Berglar, J., Muth, K., Staczan, P., et al. (2015). The role of therapists' treatment adherence, professional experience, therapeutic alliance, and clients' severity of psychological problems: Prediction of treatment outcome in eight different psychotherapy approaches. Preliminary results of a naturalistic study. *Psychotherapy Research, 25*, 420-434.

Van de Vijver, F. & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliqué, 54,* 119-135.

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*, 200-211.

Webb, C. A., DeRubeis, R. J., Dimidjian, S., Hollon, S. D., Amsterdam, J. D., & Shelton, R. C. (2012). Predictors of patient cognitive therapy skills and symptom change in two randomized clinical trials: The role of therapist adherence and the therapeutic alliance. *Journal of Consulting and Clinical Psychology, 80*, 373-381.

Weck, F., Grikscheit, F., Jakob, M., Höfling, V., & Stangier, U. (2015). Treatment failure in cognitive-behavioural therapy: Therapeutic alliance as a precondition for an adherent and competent implementation of techniques. *British Journal of Clinical Psychology, 54*, 91-108.

Weck, F., Rudari, V., Hilling, C., Hautzinger, M., Heidenreich, T., Schermelleh-Engel, K., & Stangier, U. (2013). Relapses in recurrent depression 1 year after maintenance cognitive-behavioral therapy: The role of therapist adherence, competence, and the therapeutic alliance. *Psychiatry Research, 210*, 140-145.

Weisz, J. R., Ugueto, A. M., Cheron, D. M., & Herren, J. (2013). Evidence-based youth psychotherapy in the mental health ecosystem. *Journal of Clinical Child & Adolescent Psychology, 42*, 274-286.

I

What's in a score?

Evaluation of factors affecting reliable assessment of therapist adherence after cross-national dissemination of an evidence-based intervention

Chapter 2

The lack of cross-national equivalence of a therapist adherence measure (TAM-R) in Multisystemic Therapy (MST)

Published as **Lange, A.M.C**., Busschbach, J.J.V., van Geffen, W., Timman, R., Scholte, R.H.J., & van der Rijken, R.E.A. (2015). The lack of cross-national equivalence of a therapist adherence measure (TAM-R) in Multisystemic Therapy (MST). *European Journal of Psychological Assessment*, 32, 312-325.

Abstract

This two-study report investigates the equivalence of the Dutch Therapist Adherence Measure Revised (TAM-R) to the US original. The TAM-R is a questionnaire measuring therapist adherence to the treatment model of Multisystemic Therapy (MST). Monitoring of therapist adherence is essential for evidence-based interventions to ensure the quality of the delivered treatment. International implementation of the TAM-R assumes cross-national equivalence, even though this assumption has never been investigated. In study 1 Rasch analysis was applied to 1875 Dutch TAM-R reports and the response category frequency distributions of the items of 1875 US TAM-R reports. Response frequencies were more heavily skewed in the US compared to the Netherlands and several items showed Differential Item Functioning (DIF). Study 2 investigated whether adaptations to the translation of the items and response categories could improve equivalence. For this purpose, 237 families were randomly allocated to 1 of 3 versions (original TAM-R, adapted items only, adapted items and response categories) and the analyses from study 1 were replicated. Results indicated that equivalence was not improved by the adapted translations. The article concludes with a discussion of several potential other sources of bias, such as differences in sample characteristics, implementation of MST, and response styles.

Therapist adherence to the treatment model is increasingly being recognised as a crucial aspect of effective interventions, as adherence to the evidence-based protocol is necessary to guarantee successful dissemination across multiple settings. To closely monitor therapist adherence, reliable instruments should be incorporated in the treatment (Perepletchikova, 2011; Schoenwald, 2011). Moreover, with cross-national dissemination of treatments, these instruments should be equivalent across cultural or linguistic groups and scores should be interpretable in the same way (Van de Vijver & Tanzer, 2004; Byrne & Van de Vijver, 2010). The current study aims to investigate the equivalence of the Dutch Therapist Adherence Measure-Revised (TAM-R; Henggeler, Borduin, Schoenwald, Huey, & Chapman, 2006) to the United States (US) original.

The TAM-R was developed in the US to assess therapist adherence to the treatment model of Multisystemic Therapy (MST). MST is an evidence-based intervention for youth with serious antisocial behaviour, which incorporates an elaborate quality assurance system to ensure the same quality of treatment (and thereby achievement of desired outcomes) across sites. This system contains multiple layers of continuous data-driven and qualitative feedback loops to monitor and support treatment and implementation fidelity at the level of the therapist, supervisor, consultant, and organisation (Henggeler & Schoenwald, 1999; Schoenwald, 2008). Being the primary linkage with the family, the therapist is critical in achieving desired outcomes for youths and families and the quality assurance system therefore centrally evolves around supporting and sustaining therapist adherence (Henggeler & Schoenwald, 1999). The TAM-R is used to monitor therapist adherence. Scores on the TAM-R are used to inform qualitative support from supervisors and consultants (Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 2010). MST Services (MSTS; www.mstservices.com) was founded to support communities in the development and implementation of MST using this quality assurance system (Henggeler, 2011).

In the large-scale dissemination of MST both within and outside of the US, MST Services also paid attention to linguistically and conceptually equivalent translation of all instruments and materials (Schoenwald, Heiblum, Saldana, & Henggeler, 2008). However, as far as we know, it has never been studied whether equivalence was actually achieved. Nonetheless, therapist adherence monitoring is conducted worldwide using an adherence threshold of .61, meaning that 61% of the items of the TAM-R should be scored as indicating adherence. Moreover, therapists should pass this threshold in 80% of their cases (MST Institute, 2012). By setting these targets internationally, international equivalence of the TAM-R is assumed, yet without the necessary evidence base. Lack of evidence of the equivalence of the TAM-R hampers international comparison, as it cannot be assumed that these scores can validly be compared across countries, nor that they can be measured against the same targets. The current study aims to address this gap by investigating the equivalence of the US and the Dutch TAM-R.

Equivalence will be evaluated using the theoretical framework by Van de Vijver and Tanzer (2004). Equivalence can be described as the opposite of bias and is a prerequisite for score comparison across cultural or national groups. Bias occurs if score differences on the indicators (in this instance the TAM-R) do not correspond to differences on the underlying trait (in this study: therapist adherence). Van de Vijver and Tanzer (2004) list several typical sources of bias, such as poor item translation, sample differences, differences in the administration and instructions of the instrument, differential familiarity with the material and response categories, differential response styles (e.g. social desirability or extremity scoring), differences in the underlying construct, or differential appropriateness of the behaviours tapped by the items.

In study 1 the equivalence of the TAM-R is investigated using Rasch analysis. This analytical approach is particularly suitable for studying cross-national equivalence as it allows identifying items functioning differentially across groups (Bond & Fox, 2007). Study 2 examines the influence of one typical source of bias, namely poor item and response category translation. Aim is to test an adapted translation of the Dutch TAM-R, which attempted to improve the equivalence of the Dutch TAM-R to the US original.

STUDY 1

Method

Participants and Procedure

The Dutch TAM-R is completed by the primary caregiver during monthly telephone interviews of families receiving MST. Data was collected from July 2009 to November 2011. Three institutions participated. Questionnaires not administered in Dutch were excluded. The TAM-R was completed by 580 families, providing 1875 completed TAM-R reports over an average of 3.24 administrations per family (*SD* = 1.59). They were served by 63 therapists.

Based on routinely collected data, sample characteristics are discussed below. Dutch youth were on average 15.2 (*SD* = 1.4) years of age, 67% were male, and 25% of the adolescents were from a minority ethnicity. The adolescents showed borderline internalising problem behaviour at the start of MST (mean *T*-score of 60.7, *SD* = 9.8) and clinical externalising problem behaviour (mean *T*-score of 68.0, *SD* = 10.0) based on the Child Behavior Check List 6-18 (CBCL; Achenbach & Rescorla, 2001). Of the primary caregivers, 26% were from a minority population and the majority had attended at least high school (84%) with 5% having had no education at all.

The US data consisted of response category frequency distributions of the items of the US TAM-R, made available by MST Services over the same time period as the Dutch data collection. Data underlying these distributions was simulated and a random

sample of 1875 reports out of over 89000 US TAM-R reports was selected. Published US sample characteristics from the US transportability studies were used to describe the US population, as no background information on this specific sample was available. Because the transportability studies aimed to investigate factors influencing successful dissemination and implementation of MST across community sites (Henggeler, 2011), it can be assumed that these samples are representative of the MST population in the US. In the US, adolescent mean age was 14.0 years, 65% were male and 40% of the adolescents and 35% of the primary caregivers were from a minority ethnicity (Schoenwald, Carter, Chapman, & Sheidow, 2008; Schoenwald, Chapman, Sheidow, & Carter, 2009). The majority (66%) of the US caregivers completed at least high school (Schoenwald et al., 2008). The US adolescents showed borderline internalising problem behaviour at the start of MST (mean *T*-score of 62) and clinical externalising problem behaviour (mean *T*-score of 69) based on the Child Behavior Check List 6-18 (CBCL; Achenbach & Rescorla, 2001).

Measure

The Therapist Adherence Measure-Revised (TAM-R; Henggeler et al., 2006) consists of 28 items, which can be scored on a 5-point Likert scale (1 = *not at all*, 2 = *a little*, 3 = *some,* 4 = *pretty much*, and 5 = *very much*). The items assess the therapist's adherence to the MST clinical process and the nine principles of MST according to the primary caregiver. Based on US Rasch analysis the rating scale was collapsed into two categories (category 1-4 = 0, category 5 = 1) for scoring purposes (MST Institute, 2012). The Dutch TAM-R was introduced in the Netherlands after translation and back-translation by two independent translation offices and after approval by MST Services. The US and Dutch TAM-R can be found on http://www.mstinstitute.org/qa_program/tam_languages.shtml.

Analyses

Rasch analysis. The Dutch TAM-R was fitted to the Andrich Rating Scale Model (RSM), a Rasch model for Likert scale data, which was based on joint maximum likelihood estimation using WINSTEPS (version 3.73, Linacre, 2011a). Procedures by Mallinson (2011) were followed to examine whether multilevel modelling was necessary to account for the nested structure of the data (therapists [upper level] were rated by a number of parents [middle level], who each completed the TAM-R on a monthly basis while treatment was ongoing [lowest level]). This procedure consisted of cross-plotting questionnaire score estimates (in Rasch terminology 'person measures') based on the whole sample against questionnaire score estimates based on one random measurement per family. Care was taken to include an equal proportion of each measurement moment in both samples. Questionnaire score estimate differences smaller than 0.5 logits would indicate that the nested structure of

the data had a negligible effect on the estimates and could therefore be ignored (Linacre, 2011b). This procedure was replicated to test for nesting of families in therapists by taking one random measurement per therapist. Results showed that the largest discrepancy was 0.14 logits when testing for multiple measurements within families and 0.18 logits when testing for multiple measurements within therapists. This means that the nested structure of the data had a negligible effect on the estimates and that multilevel modelling was not deemed necessary.

Unidimensionality, a major assumption of the RSM, was analysed using a Principal Component analysis (PCA). The purpose of the PCA was not a thorough evaluation of the structure underlying the TAM-R, but rather an examination of the appropriateness of the assumption of unidimensionality. Without unidimensionality underlying the questionnaire, the RSM cannot be fitted to the data reliably. For this purpose three criteria were used: 1) the variance explained by the first component should be at least 40%, 2) the first eigenvalue should be at least five times greater than the second eigenvalue, and 3) all items should load .30 or higher on the first component (Carmines & Zeller, 1979; Hambleton, Swaminathan, & Rogers, 1991).

Further, category thresholds should be ordered to allow for reliable estimations of the item measures and questionnaire scores, and the thresholds should be at least 1 logit apart from one another to allow for clear differentiation (Bond & Fox, 2007). Reliability of the scale was assessed with the person reliability index (comparable to Cronbach's alpha, should be > .80) and the person separation coefficient (signal-to-noise ratio, should be > 1.50, Tennant & Conaghan, 2007). Item infit and outfit mean square standardized residuals (MNSQ) were used to test for model fit, as these are relatively independent of sample size (Smith, Rush, Fallowfield, Velikova, & Sharpe, 2008). MNSQ is a chi-square statistic divided by its degrees of freedom and is standardized (mean = 0, standard deviation = 1). Values exceeding 2.00 indicate a distortion of the data (Bond & Fox, 2007; Linacre, 2011b).

Cross-national comparison. First, the response category frequency distributions of both countries were compared to gauge the extent to which differences at item-level might be systematic across the whole questionnaire. For this purpose, 28 (for each individual item) Fisher's exact tests with a *p*-value of .05 (Bonferroni-corrected to a *p*-value of .0018) were conducted. Also, the percentage of reports passing the adherence threshold of .61 set by MST Services was investigated.

Second, Differential Item Functioning (DIF) was examined to investigate whether Dutch items functioned differently from US items. An item shows DIF when subjects from different groups, who are equal in their level on the underlying trait, do not have the same probability of endorsing a test item (Bond & Fox, 2007). Items were considered to show DIF when the difference between the item measures of both countries was both significant (with a Bonferroni-corrected *p*-value of .0018) and larger than 0.5 logits (Linacre, 2011b).

Results

Rasch analysis. Unidimensionality was analysed with a PCA. The explained variance of the first and second factor was 48.9% and 7.1% respectively, with corresponding eigenvalues of 13.7 and 2.0. All items had their highest positive loading on the first factor, which was at least .50 for all items. This means that all criteria for unidimensionality were met and fitting the RSM was deemed appropriate.

The lowest two category thresholds were not ordered, which hampered reliable estimations of the item measures and questionnaire scores (Table 1). Nevertheless, the category measure and the observed average measure (the mean questionnaire score per category) were increasing monotonically, suggesting that the categories themselves were not disordered. Instead the observed disordering is considered the result of the low frequencies of categories 1-3 (Adams, Wu, & Wilson, 2012). Also, the first three thresholds were closer to one another than 1 logit. By collapsing the categories 1-3 the fit of the rating scale was improved and all criteria were met. Therefore, the 3-point rating scale was used for further analyses.

Table 1 RSM category estimates for 5-point and 3-point rating scale study 1

	Observed observations	Category measure	Observed average measure	Category threshold	Infit Mean Square	Outfit Mean Square
			5 categories (12345)			
1	2309 (5%)	-1.97	-0.37	-	1.38	1.95
2	1819 (4%)	-0.92	-0.10	-0.13	1.03	1.27
3	5675 (11%)	-0.19	0.37	-0.97	0.94	1.02
4	14309 (28%)	0.77	1.05	-0.15	0.94	0.81
5	26114 (52%)	2.51	2.32	1.25	0.94	0.96
			3 categories (11145)			
1	9803 (20%)	-1.97	-0.92	-	1.13	1.30
4	14309 (28%)	0.00	0.26	-0.71	0.90	0.84
5	26114 (52%)	1.96	1.98	0.71	0.95	0.98

Reliability and model fit of the TAM-R with 3-point rating scale were found to be good: Person reliability was .88, person separation was 2.86, and all items had their MNSQ values below 2.00. From here on, 11145 will be used to refer to this 3-point rating scale, indicating that the first three categories were recoded to be the same and only category 4 and 5 were maintained as separate categories. The US scoring system of collapsing the categories 1 to

4 and only preserving category 5 will be referred to as 11115, while the full 5-point scale will be referred to as 12345, since all 5 categories are being maintained.

Cross-national comparison. Fisher Exact tests of the 28 items showed that the response category frequency distributions of the US and the Netherlands were significantly different for all items, except item 16 ('My family was sure about the direction of treatment'). US parents choose category 5 ('very much') 75% of the time, which was significantly more often than the Dutch parents (52%). In contrast category 4 ('pretty much') was scored by 28% of the parents in the Netherlands compared to only 17% in the US. The percentage of Dutch reports passing the adherence threshold was 44% when using the US scoring system of 11115 and increased to 65% when using 11145. In comparison, 70% of the MST clients in the US reported adherence above the threshold of .61 (MST Institute, 2010).

DIF analysis showed that seven items (namely item 2, 3, 9, 10, 16, 19, and 20) fulfilled the criterion of being both significant and showing a difference larger than 0.5 logits when using 11145. Interestingly, the number of items with DIF increased to 14 (2, 3, 4, 8, 9, 10, 15, 16, 18, 19, 20, 21, 23, and 26) when using 11115.

Discussion of Study 1

The purpose of study 1 was to evaluate the equivalence of the Dutch TAM-R to the US original. First, several items showed differential item functioning. This means that respondents with a similar questionnaire score (and thus a similar assumed trait level) in the US and the Netherlands do not respond in a similar manner to certain items. As a consequence, these items lead to bias or nonequivalence: The same score on the TAM-R can no longer warrant a similar ability on the underlying trait (therapist adherence).

Secondly, this study showed that US response frequencies are more heavily positively skewed than Dutch frequency distributions for the questionnaire as a whole; all items but one had a significantly different distribution. This lack of a comparable distribution across almost all items could point to problems of nonequivalence at a more general level, affecting how the TAM-R as a whole is functioning across countries, besides bias at the level of specific items.

Interestingly, the discrepancies between the Netherlands and the US increased when applying the US scoring system (11115). Since this scoring system is based on the US frequency distributions and is applied to all items, this increase in discrepancy corroborates the hypothesis of nonequivalence at the level of the questionnaire.

An expert panel (consisting of four experienced interviewers from the call-centre collecting Dutch TAM-R data and six Dutch MST consultants, one of whom was bi-lingual) suggested that several of the items with DIF have a different meaning in the two countries (e.g. Dutch items being formulated more strongly or emphasising different aspects of the

question) and that an adaptation of the translation might improve equivalence. Also, the response scale was thought to be quite unfamiliar to Dutch parents, which may influence the manner in which it is used and thereby influence the questionnaire as a whole. Study 2 was set up to investigate poor item and response category translation as a potential source of nonequivalence. Aim was to attempt improving the equivalence between the Dutch and US TAM-R through adaptations of the translation of the items and the response categories.

STUDY 2

Method

Procedure

For study 2, the translation of the Dutch TAM-R was adapted, such that items showed more similarity in structure, content and intent to the US TAM-R, but also that response categories were more familiar to Dutch participants. In the Appendix, all adaptations to the items are recorded. Item adaptations could broadly be separated into three categories. Since some items included multiple adaptations, items could fall into more than one category. Firstly, some items were adapted in order to be closer to the English wording of the item. This included adaptations such as adding a single word (item 1, 3, and 18) or replacing a word of the item (item 6, 7, 11, 14, 16-18, 22, and 24). Secondly, some items were adapted in order to be easier to understand or provide a grammatically more correct sentence (item 6, 7, 9, and 10). Lastly, Dutch items were adapted to be closer to the content or intent of the US items. In MST, the therapist should help the caregivers to take responsibility; the therapist is a coach but is not taking over. Everything is done in collaboration. Nevertheless, in some of the Dutch items too much emphasis was placed on the therapist forcing instead of helping the family (item 5), taking over responsibility (item 24), or on passiveness from the caregivers regarding the therapy process instead of asking their active collaboration (item 12-14, 21, and 22). Other items proved to be somewhat different in content for different reasons. The original Dutch item 4 revolved around the precision of the recommendation, instead of whether the recommendation was targeted at a specific problem. When hearing the original Dutch item 15 caregivers often felt they had to rate the achievements or success of the therapy, instead of rating whether the session was action-oriented. In the original Dutch item 19 caregivers were asked whether the therapist's recommendations made family members more responsible, instead of asking whether these recommendations helped family members to become more responsible. The remaining items (item 2, 8, 20, 23, and 25-28) did not require any adaptations. Response categories proved to be quite similar in content and structure to the English response categories, however, were quite unfamiliar to Dutch caregivers. Therefore, response categories were changed from a 5-point scale

ranging from 'not at all' to 'very much' into a 5-point scale with the categories 1 = totally disagree, 2 = disagree, 3 = partly agree / partly disagree, 4 = agree, and 5 = totally agree.

The wordings of all items and the response categories of the Dutch TAM-R were discussed with two focus groups (one with MST interviewers from the call-centre, and one with MST consultants). The input from these two focus groups was integrated with the results from study 1. The resulting adaptations to the TAM-R were sent to all members of the two focus groups for feedback. Remaining disagreement was resolved during a second meeting with the members of the MST consultant focus group. This resulted in a pilot version of the questionnaire, which was submitted to five primary caregivers of youth receiving MST during a face-to-face interview. After consulting these parents, no further improvements were deemed necessary.

In study 2, three different versions of the Dutch TAM-R were evaluated: The original translation without any adaptations (version R), a version with only the adapted translation of the items (version A), and a second version in which both the items and the wording of the response categories were adapted (version B).

Participants

Families receiving MST between September 2012 and September 2013 at one of the participating institutions from study 1 were randomly allocated to one of the three versions of the Dutch TAM-R. Families were administered the same questionnaire during the whole duration of therapy. Random allocation was stratified on therapist to control for any therapist effects. This resulted in 85 families being allocated the current translation (version R), 78 families being allocated the version with adapted translation of the items (version A), and 74 families being allocated the version with adapted translation of the items as well as the wording of the response categories (version B). With an average of 4.05 administrations per family (*SD* = 1.63), this provided a total of 292 TAM-R R reports, 258 TAM-R A reports, and 259 TAM-R B reports. The families were served by 32 therapists. As in study 1, questionnaires not administered in Dutch were excluded. The US data consisted of response category frequency distributions of the items of the US TAM-R, made available by MST Services for study 1. Data underlying these distributions was simulated and a random sample of 300 reports out of over 89000 US TAM-R reports was selected.

The Dutch youth were on average 15.4 (*SD* = 1.5) years of age, 69% were male, and 8% of the adolescents were born outside of the Netherlands. The adolescents showed borderline internalising problem behaviour at the start of MST (mean *T*-score of 61.0, *SD* = 9.5) and clinical externalising problem behaviour (mean *T*-score of 67.9, *SD* = 11.1) based on the Child Behavior Check List 6-18 (CBCL; Achenbach & Rescorla, 2001). Of the primary caregivers 30% were born outside of the Netherlands. The majority had attended at least high school (85%) with 4% having had no education at all. The sample characteristics did

not differ significantly for the three versions of the TAM-R.

Analyses

The analyses of study 1 were replicated for the three versions of the Dutch TAM-R (version R, version A, and version B) separately.

Results

Rasch analysis. Unidimensionality was analysed with a PCA for each version of the TAM-R separately. Except for the second criterion for version B (the ratio of eigenvalues of the first and second factor was only 4.8 instead of a minimum of 5), all criteria were met. Since all criteria were met for version R and A, and two out of three criteria were met for version B (with the second criterion only slightly below the required minimum), unidimensionality was deemed sufficient to continue analyses.

As was the case in study 1, the lowest two category thresholds were not ordered. This was true for all three versions of the TAM-R (Table 2). Also, the first three thresholds were closer to one another than 1 logit and the outfit MNSQ was above 2.00 for all versions. Since the category measure and the observed average measure (the mean questionnaire score per category) were increasing monotonically, the categories themselves did not appear to be disordered. Rather, the observed disordering was the result of the low frequencies of categories 1-3 (Adams et al., 2012). By collapsing the categories 1-3 the fit of the rating scale was improved and all criteria were met for all three versions (see Table 3). Therefore, the 3-point rating scale (11145) was used for further analyses.

Reliability and model fit of the TAM-R with a 3-point rating scale were found to be good for all three versions. Person reliability varied between .81 and .86, person separation varied between 2.09 and 2.51, and the MNSQ of all items were below 2.00 for the three versions of the TAM-R.

Cross-national comparison. First, the response category frequency distributions of the items of the US were compared to the distributions of the three Dutch versions using Fisher's Exact tests for all items. Results showed that distributions of the US and the Netherlands were significantly different for the majority of the items. Version R differed from the US on 12 items, version A differed from the US on 18 items, and version B differed on 24 items.

Table 2 RSM category estimates for TAM-R R, TAM-R A and TAM-R B, 5-point scale study 2

	Observed observations	Category measure	Observed average measure	Category threshold	Infit Mean Square	Outfit Mean Square
			TAM-R R			
1	265 (3%)	-1.72	-0.11	-	1.47	2.46
2	159 (2%)	-0.79	0.03	0.35	0.96	1.31
3	614 (8%)	-0.16	0.51	-1.02	0.91	0.90
4	1831 (24%)	0.63	1.06	-0.23	0.90	0.74
5	4912 (63%)	2.21	2.16	0.91	0.94	0.96
			TAM-R A			
1	272 (4%)	-1.85	-0.32	-	1.47	2.32
2	196 (3%)	-0.83	-0.05	0.00	1.09	1.56
3	575 (8%)	-0.15	0.52	-0.82	0.94	1.01
4	1694 (25%)	0.70	1.11	-0.19	0.97	0.72
5	4125 (60%)	2.30	2.44	1.01	0.94	0.95
			TAM-R B			
1	235 (3%)	-1.74	0.02	-	1.42	2.17
2	104 (2%)	-0.88	0.21	0.68	1.11	1.23
3	665 (10%)	-0.26	0.57	-1.54	1.03	1.22
4	2314 (34%)	0.65	0.94	-0.43	1.11	0.77
5	3570 (52%)	2.50	2.25	1.28	0.83	0.91

US parents choose category 5 ('very much') 75% of the time, which was significantly more often than the Dutch parents in all three versions (63%, 60%, and 52% respectively). In contrast category 4 ('pretty much') was scored by 24% of the parents in the Netherlands using version R, 25% of the parents using version A, and 34% of the parents using version B, compared to 18% of the parents in the US. Contrary to our expectations, version B showed more dissimilarity with the US than version R and A (see Table 4).

The percentages of Dutch reports passing the adherence threshold were 68% for version R, 62% for version A, and 57% for version B when using the US scoring system of 11115. These values increased to 86% (TAM-R R), 81% (TAM-R A), and 83% (TAM-R B) when using 11145.

Table 3 RSM category estimates for TAM-R R, TAM-R A and TAM-R B, 3-point scale study 2

	Observed observations	Category measure	Observed average measure	Category threshold	Infit Mean Square	Outfit Mean Square
			TAM-R R			
1	1032 (13%)	-1.86	-0.51	-	1.19	1.38
4	1831 (24%)	0.00	0.46	-0.58	0.86	0.75
5	4895 (63%)	1.86	2.02	0.58	0.95	0.97
			TAM-R A			
1	1035 (15%)	-1.88	-0.68	-	1.19	1.48
4	1680 (25%)	0.00	0.43	-0.61	0.88	0.74
5	4101 (60%)	1.88	2.25	0.61	0.93	0.94
			TAM-R B			
1	1004 (15%)	-2.10	-0.36	-	1.27	1.52
4	2314 (34%)	0.00	0.28	-0.90	0.84	0.75
5	3547 (52%)	2.10	2.06	0.90	0.85	0.88

Lastly, DIF analyses were conducted contrasting the three Dutch versions to the US. When using 11145 the amount of items with DIF was eight for version R and version A, and four for version B. When using 11115 the amount of items with DIF was eight for all three versions.

Discussion Study 2

The aim of study 2 was to investigate whether adaptations to the translation of the items and the response categories of the TAM-R would improve equivalence between the Dutch and the US version. The results showed that the adaptations did not achieve the intended improvements. The number of items showing DIF was comparable between the adapted and the original questionnaire versions. Also the percentage of questionnaires passing the adherence threshold was not influenced by the version of the TAM-R. Comparison of the response distributions showed that version A (adapted items only) was quite similar to the original version, whereas the distribution for version B (which included adaptations to the wording of the response categories) had a slightly different pattern. Interestingly, this pattern showed more dissimilarity with the US than the other two versions, indicating that the adaptations to the response categories had an influence opposite to the expected direction.

Table 4 Total category percentages for each version of the TAM-R

Category	Study 1		Study 2			
	US	Dutch TAM-R	US	TAM-R R	TAM-R A	TAM-R B
1	1	5	1	3	4	3
2	1	4	1	2	3	2
3	5	11	4	8	8	10
4	17	28	18	24	25	34
5	75	52	75	63	60	52

The adaptations of the translation of the items did not improve equivalence. The number of items with DIF, as well as the response category frequency distribution of the items, was very similar for version A (adapted items) and version R (original Dutch TAM-R). Poor item translation did not prove to be the source of bias between the US and Dutch TAM-R. Nevertheless, the wording of the adapted items is closer to the US version and the interviewers from the call-centre stated that the questions were easier to understand for the caregivers completing the TAM-R. Therefore, the adapted translation of the items (version A) might still be preferable to the original Dutch version.

The adapted response categories decreased equivalence with the US, contrary to our hypotheses. As intended, the scale was more familiar to the Dutch respondents, as confirmed by the TAM-R interviewers from the call-centre; parents found these response categories easier to understand and to use. However, using a more familiar scale meant that the translation was not as close to the US version as the original Dutch TAM-R. The current results suggest that the new translation actually has a different meaning, which influences the manner in which the response categories are used. Therefore, the adaptations to the response categories (version B) do not seem able to improve equivalence compared to the original Dutch version.

General discussion

This two-study report investigated the equivalence of the Dutch TAM-R to the US original and looked into one of the most important typical sources of bias in cross-national comparison of scores, namely poor item and response category translation (Van de Vijver & Tanzer, 2004; Van Widenfelt, Treffers, De Beurs, Siebelink, & Koudijs, 2005). The results showed that the TAM-R is nonequivalent between the two countries and that an adapted translation of the items and response categories could not improve the equivalence, even though both MST consultants and TAM-R interviewers agreed upon the improved conceptual and linguistic similarity of the adapted Dutch translation to the US original. This article will conclude by discussing other potential sources of bias, based on the overview of Van de Vijver and Tanzer (2004), in an attempt to discern which source of bias is most likely to underlie the current nonequivalence.

Dissimilarity in the population, administration conditions and instruction, and in the construct or measurement of the construct across countries do not seem likely sources of bias in the current study. Comparison of the US and Dutch MST population showed the severity of youth problem behaviour to be very similar in both samples. Altough some dissimilarities were found regarding youth age, and youth and caregiver ethnicity, these client characteristics have been found not to relate to therapist adherence (Ryan, et al., 2013; Schoenwald, Halliday-Boykins, & Henggeler, 2003). The difference in caregiver educational levels seems most likely due to differences in the educational system of the US and the Netherlands, in which case it would not reflect an actual difference across populations. Bias due to differences in administration conditions and instruction does also not seem likely since both US and Dutch data were collected by telephone interviews according to international guidelines regarding data collection and family instructions (MST Institute, n.d.). Lastly, bias due to differences in the construct or the measurement of the construct seems unlikely for several reasons. In the first place, a recent publication on the international implementation of MST argued that the principles of MST proved to be applicable across international sites (Schoenwald et al., 2008). Moreover, the quality assurance system ensures that the core principles of MST are provided in a similar manner worldwide. Lastly, the involvement of Dutch MST consultants (who have extensive clinical knowledge of the MST model and the Dutch culture) in the adaptations to the translation of the items of the TAM-R in study 2 ensured that all adapted items would be appropriate for the Dutch context. In this way potential small differences in the actual behaviour of the therapist would be accounted for.

The current study did however provide some clues regarding more plausible sources of bias. In the first place, the adherence scores in the Netherlands increased between study 1 (conducted between 2009 and 2011) and study 2 (conducted between 2012 and 2013). Since in both studies the exact same questionnaire was completed by a very similar sample

under similar conditions, it is unlikely that these differences are due to study differences. Instead, it is more likely that these results corroborate previous findings that adherence scores reflect the (lack of) success of the implementation of an intervention (Durlak & DuPre, 2008; Landsverk, 2013). Specifically for MST, a recent study showed that therapist adherence increases with increasing team and organisational years of experience in providing MST (Löfholm et al., 2014). Since MST has only been disseminated outside of the US in the last 10 to 15 years, this implementation effect might (partly) explain why adherence scores in the Netherlands, but also internationally, are lower than in the US (MST Institute, 2010). However, further research would be required, since the relationship between implementation of MST and therapist adherence has only been studied in Sweden so far (Löfholm et al., 2014). Investigating the level of implementation of MST in different countries could enlighten some of the international differences in adherence scores.

Secondly, findings from the current study point to differential response styles as a potential source of bias, as US scores were more heavily positively skewed than Dutch scores. Numerous studies have pointed to the fact that a score is not only influenced by the content of the item, but also by the respondent's response style (Diamantopoulos, Reynolds, & Simintiras, 2006; Harzing, 2006). These response styles vary as a consequence of (familiarity with) the format of the items and response scale, demographic characteristics, and cultural or national groups (Arce-Ferrer, 2006; Diamantopoulos et al., 2006; Harzing, 2006). Relevant for the current study are the findings from a large cross-national survey showing that US participants are more prone to using positive, but not negative, extreme response categories than Dutch respondents (Harzing, 2006). To determine whether response style is indeed a main source of bias in the Dutch TAM-R, it is recommended to investigate the influence of response tendencies, especially international differences in a positive extreme response style. Several analytical approaches are available for this purpose (Morren, Gelissen, & Vermunt, 2011). Importantly, future studies should not only investigate the extent to which current TAM-R scores are distorted by response style, but should also provide descriptives of the response styles accros different countries, to allow controlling for response style in the ongoing use of the TAM-R for quality assurance purposes worldwide.

This study has some limitations. First, the analyses were limited by the lack of full US data. Although the item frequency distributions allowed full DIF analyses and comparison of response category frequency distributions, thereby demonstrating bias at the item level (lack of scalar equivalence), it was not possible to test for other forms of measurement equivalence (such as metric equivalence or structural equivalence). Investigating full measurement invariance would provide a more thorough understanding of what levels of equivalence are and are not achieved. Since Rasch analysis mainly focuses on item functioning, other statistical approaches such as Structural Equation Modelling (SEM) should be considered for this purpose (Byrne, 2012; Van der Schoot, Lugtig, & Hox, 2012). Also, the

current data did not allow further investigation into other sources of the nonequivalence of the TAM-R, such as testing the hypothesis regarding extreme response style. Nevertheless, this study did provide a thorough discussion of the probability of several sources of bias as well as the implications of these findings for future research. Lastly, this study was restricted to a comparison of the Netherlands and the US. As MST is being implemented in multiple countries, a multi-country-comparison would be required to estimate the extent to which international comparison of adherence scores is legitimate.

In conclusion, this study has demonstrated that the US and Dutch TAM-R as yet are not equivalent and that this could not be attributed to poor item and response category translation. Response style differences and differences in (years since) implementation have been identified as plausible in explaining the nonequivalence of the Dutch and US TAM-R and therefore require further research.

Assessing the equivalence of an instrument is only a first step in developing cross-nationally comparable instruments and the identification of the source of bias may not always be straightforwardly solved by improvement of the translation. Thorough evaluation of typical sources of bias can assist in identifying what sources of bias require further investigation. It is hoped that future research will be able to establish the main source of bias in the TAM-R, so that the TAM-R can be used as a reliable and valid measure to monitor and improve MST therapist adherence across nations. Until satisfactory evidence regarding the equivalence of the TAM-R is provided, international TAM-R scores should be interpreted with caution.

Acknowledgement

We would like to thank the mental health institutions de Viersprong, Dimence, and Vincent van Gogh for their cooperation in data collection, MST Services for providing response distributions for the US TAM-R, and the expert panel for sharing their expertise and ideas with us.

References

Achenbach, T., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles.* Burlington, VT: University of Vermong, Research Center for Children, Youth & Families.

Adams, R. J., Wu, M. L., & Wilson, M. (2012). The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement, 72*, 547-573.

Arce-Ferrer, A. J. (2006). An investigation into the factors influencing extreme response style: Improving meaning of translated and culturally adapted rating scales. *Educational and Psychological Measurement, 66,* 374-392.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in human sciences*. Mahwah, NJ: Erlbaum.

Byrne, B. M. (2012). *Structural Equation Modeling with Mplus. Basic concepts, applications, and programming*. NY: Routledge Taylor & Francis Group.

Byrne, B. M., & van de Vijver, F. J. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing, 10,* 107-132.

Carmines, E.G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.

Diamantopoulos, A., Reynolds, N. L., & Simintiras, A. C. (2006). The impact of response styles on the stability of cross-national comparisons. *Journal of Business Research , 59,* 925-935.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology , 41*, 327-350.

Hambleton, R. K., Swaminathan, J., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.

Harzing, A. W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management, 6,* 243-266.

Henggeler, S. W. (2011). Efficacy studies to large-scale transport: The development and validation of Multisystemic Therapy Programs. *Annual Review of Clinical Psychology, 7,* 351-381.

Henggeler, S. W., Borduin, C. M., Schoenwald, S. K., Huey, S. J., & Chapman, J. E. (2006). *Multisystemic Therapy Adherence Scale-Revised (TAM-R).* Unpublished instrument. Charleston, SC: Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina.

Henggeler, S. W., & Schoenwald, S. K. (1999). The role of quality assurance in achieving outcomes in MST programs. *Journal of Juvenile Justice and Detention Services, 14*, 1-17.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (2010). *Multisysteem therapie voor kinderen en jongeren met antisociaal gedrag* [Multisystemic therapy for antisocial behavior in children and adolescents]. Amsterdam: Uitgeverij SWP.

Landsverk, J. (2013). Reflections on treatment integrity in a dissemination and implementation framework. *Clinical Psychology: Science & Practice , 20*, 114-119.

Linacre, J. M. (2011a). Winsteps® Rasch measurement computer program. Beaverton, Oregon: Winsteps.com

Linacre, J. M. (2011b). Winsteps® Rasch measurement computer program User's Guide. Beaverton, Oregon: Winsteps.com

Löfholm, C. A., Eichas, K., & Sundell, K. (2014). The Swedish implementation of Mutisystemic Therapy for adolescents: Does treatment experience predict treatment adherence? *Journal of Clinical Child & Adolescent Psychology, 43,* 643-655.

Mallinson, T. (2011). Rasch analysis of repeated measures. *Rasch Measurement Transactions, 25,* 1317.

Morren, M., Gelissen, J. P., & Vermunt, J. K. (2011). Dealing with extreme response style in cross-cultural research: A restricted latent class factor analysis approach. *Sociological Methodology, 41,* 13-47.

MST Institute (n.d.). *Adherence data: Tips and strategies.* Retrieved from www.mstinstitute.org/qa_program/pdfs/TAMCollectionTips.pdf

MST Institute. (n.d.). *MST Therapist Adherence-Revised (TAM-R): Guidelines for administration.* Retrieved from www.mstinstitute.org/qa_program/pdfs/TAMAdministration.pdf

MST Institute (2010). *2010 MST Data Report*. Retrieved from http://www.mstinstitute.org

MST Institute (2012). *MST Therapist Adherence Measure – Revised (TAM-R): Guidelines for Interpretation*. Retrieved from http://www.mstinstitute.org/qa_program/tam.shtml

Perepletchikova, F. (2011). On the topic of treatment integrity. *Clinical Psychology, Science and Practice, 18,* 148-153.

Ryan, S. R., Cunningham, P. B., Foster, S. L., Brennan, P. A., Brock, R. L., & Whitmore, E. (2013). Predictors of therapist adherence and emotional bond in Multisystemic Therapy: Testing ethnicity as a moderator. *Journal of Child and Family Studies, 22,* 122-136.

Schoenwald, S. K. (2008). Towards evidence-based transport of evidence-based treatments: MST as an example. *Journal of Child and Adolescent Substance Abuse, 17,* 69-91.

Schoenwald, S. K. (2011). It's a bird, it's a plane, it's fidelity measurement in the real world. *Clinical Psychology, Science and Practice, 18,* 142-147.

Schoenwald, S. K., Carter, R. E., Chapman, J. E., & Sheidow, A. J. (2008). Therapist adherence and organizational effects on change in youth behavior problems one year after Multisystemic Therapy. *Administration and Policy in Mental Health and Mental Health Services Research, 35,* 379-394.

Schoenwald, S. K., Chapman, J. E., Sheidow, A. J., & Carter, R. E. (2009). Long-term youth criminal outcomes in MST transport: The impact of therapist adherence and organizational climate and structure. *Journal of Clinical Child & Adolescent Psychology, 38,* 91-105.

Schoenwald, S. K., Halliday-Boykins, C. A., & Henggeler, S. W. (2003). Client-level predictors of adherence to MST in community service settings. *Family Process, 42,* 345-359.

Schoenwald, S. K., Heiblum, N., Saldana, L., & Henggeler, S. W. (2008). The international implementation of Multisystemic Therapy. *Evaluation & the Health Professions, 31,* 211-225.

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *Medical Research Methodology, 8*:33. doi:10.1186/1471-2288-8-33

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why is it? When should it be applied and what should one look for in a Rasch paper? *Arthritis & Rheumatism (Arthritis Care & Research), 57,* 1358-1362.

Van der Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9,* 1-7.

Van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliqué, 54,* 119-135.

Van Widenfelt, B. M., Treffers, P. D., De Beurs, E., Siebelink, B. M., & Koudijs, E. (2005). Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clinical Child and Family Psychology Review, 8,* 135-147.

Chapter 3

It's not just the therapist: Therapist and country-wide experience predict therapist adherence and adolescent outcome

Published as **Lange, A.M.C.**, van der Rijken, R.E.A., Busschbach, J.J.V., Delsing, M.J.M.H., & Scholte, R.H.J. (2017). It's not just the therapist: Therapist and country-wide experience predict therapist adherence and adolescent outcome. *Child and Youth Care Forum*, 46, 455-471

Abstract

Objective:
Therapist adherence is a quality indicator in routine clinical care when evaluating the success of the implementation of an intervention. The current study investigated whether therapist adherence mediates the association between therapist, team, and country-wide experience (i.e. number of years since implementation in the country) on the one hand, and treatment outcome on the other hand. We replicated and extended a study by Löfholm and colleagues (2014).

Method:
Data over a ten-year period were obtained from 4290 adolescents (12-17 years) with antisocial or delinquent problem behavior, who were treated with Multisystemic Therapy (MST) by 222 therapists, working in 27 different teams in the Netherlands. Multilevel structural equation modeling was used to assess the associations between experience, therapist adherence, and post-treatment outcomes.

Results:
Treatment outcomes were directly predicted by therapist experience, country-wide experience, and therapist adherence, but not by team experience. Moreover, therapist adherence mediated the association between therapist and country-wide experience, and treatment outcomes. The association between therapist experience and therapist adherence was not affected by the number of years of team experience or country-wide experience.

Conclusion: The effect of country-wide experience on outcome may reflect increasing experience of training and supporting the therapists. It suggests that nation-wide quality control may relate to better therapist adherence and treatment outcome for adolescents treated with systemic therapy.

Evidence-based psychotherapies are being favored by many for their superior effectiveness over treatment as usual. However, the evidence supporting the effectiveness of these psychotherapies is often achieved in highly-controlled research settings (i.e. efficacy studies) that may not easily generalize to clinical practice (Henggeler, 2011; Weisz et al., 2013). Indeed, studies conducted within everyday practice tend to achieve smaller effects than efficacy trials (Henggeler, 2011). One of the reasons for this difference may be that it is harder for practitioners in everyday practice to deliver an evidence-based intervention as intended, i.e. to deliver the treatment with high adherence (Durlak & DuPre, 2008; Shirk & Peterson, 2013).

Therapist adherence is defined as the extent to which a therapist adheres to the treatment protocol or manual (McLeod et al., 2013; Perepletchikova & Kazdin, 2005). Therapist adherence has been found to be related to positive treatment outcomes (Forgatch et al., 2005; Mihalic, 2004; Schoenwald, 2008). In addition, therapist adherence is a salient indicator of successful implementation (Durlak & DuPre, 2008; Fixsen et al., 2005; McLeod et al., 2013; Schoenwald & Garland, 2013). This is clearly represented in the implementation framework developed by Fixsen and colleagues (2005; see Fig. 1), in which therapist adherence is conceived as a part of treatment fidelity. The framework consists of several elements: the *source* represents the core components of the evidence-based intervention; the *destination* represents the practitioner who delivers the intervention; the *communication link* consists of practitioner training and coaching in order to maintain adherence to the core components. Since adherence is associated with treatment outcomes, the implementation framework also includes a *feedback* loop, which represents fidelity measures to monitor adherence at the level of the practitioner, the manager, and the organization. The implementation of an intervention may be influenced by factors other than just the practitioner, manager, and organization, as all interventions operate within a dynamic and demanding environment. This environment is visualized in the framework as *influence* and consists of factors such as funding, regulation, licensing, community relations, and agency collaboration (Durlak & DuPre, 2008; Fixsen et al., 2005; Shirk & Peterson, 2013).

Sustaining adequate adherence in clinical practice within this dynamic environment may be challenging (Durlak & DuPre, 2008; Shirk & Peterson, 2013). Several scholars have thus stressed the need to include quality-control methods (i.e. training, coaching, and monitoring instruments) to achieve and sustain adequate adherence and desirable outcomes (Garland & Schoenwald, 2013; Henggeler & Sheidow, 2012; Southam-Gerow & McLeod, 2013). Without such quality-control methods, an intervention may quickly start to drift, resulting in lower adherence, reduced effectiveness, and inclusion of clients who do not meet the treatment's inclusion criteria (Henggeler et al., 2008; Smith-Boydston et al., 2014).

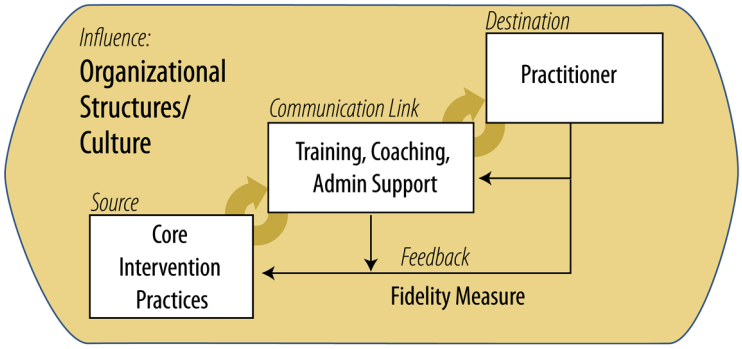

Fig. 1 Implementation framework (Fixsen et al., 2005, p. 28)

An example of such an intervention with quality-control methods is Multisystemic Therapy (MST), an intensive home and community-based intervention addressing the multidetermined nature of the antisocial or delinquent behavioral problems of adolescents 12-18 years old (Henggeler et al., 2009). The MST quality-assurance system consists of data-driven and qualitative feedback loops to sustain adherence (Henggeler & Schoenwald, 1999; Schoenwald, 2008). It supports therapists and supervisors all over the world through the following: An initial training and quarterly booster sessions; weekly supervision and consultation on all cases; and data-driven monitoring of treatment outcomes and treatment adherence. Therapist adherence is monitored using the Therapist Adherence Measure (TAM; Henggeler & Borduin, 1992). Organizational support is provided through manuals, initial meetings to prepare for the implementation of MST at the site, monitoring of program performance (such as therapist caseload and treatment duration), and continued expert consultation on program drift (internal and external factors affecting treatment fidelity; Henggeler & Schoenwald, 1999; Schoenwald, 2008). Research has shown that higher therapist adherence is associated with better treatment outcomes, such as favorable long-term criminal outcomes, fewer out-of-home placements, and better family functioning (e.g., Huey et al., 2000; Löfholm et al., 2014; Schoenwald et al., 2009a).

Nevertheless, even in the presence of a quality-assurance system, variance in therapist adherence occurs, a phenomenon which has also been observed within MST (MST Institute 2010, 2014). To date, little is known about the factors that influence therapist adherence. Identifying which factors affect therapist adherence is an essential step towards increasing the likelihood of successful implementation of interventions, leading to better treatment outcomes.

A recent study by Löfholm and colleagues (2014) suggested that therapist adherence may relate to the amount of experience of delivering a treatment. They studied therapist experience, team experience, and country-wide experience. Therapist experience did not

predict therapist adherence, however team experience did predict adherence: Therapists achieved higher adherence scores when surrounded and supported by a team with more years of experience. Country-wide experience also predicted therapist adherence: Therapists who received training once the treatment had been running in the country for more than 2 years achieved higher adherence scores than 'first-batch' therapists, who started at the initial implementation of the treatment in the country. Trying to explain the absent relation between therapist experience and therapist adherence, Löfholm and colleagues hypothesized that therapist experience might start making a distinctive contribution only after the teams and organizations have acquired sufficient experience and stability to succesfully support therapist adherence. The current study aimed to test this hypothesized moderating effect.

Because clinical implementation and treatment outcomes may be influenced by factors at different levels (i.e. youth and family, clinician, organization, and service system; Schoenwald, 2008), cross-national replication of implementation research is important to assess the generalizability of previous findings. In fact, as health care systems, funding, regulation etc. may vary between countries, findings from one study are not necessarily applicable to other countries and settings. Therefore, we replicated the analytical model by Löfholm and colleagues (2014) using a longer follow-up period of Dutch MST data and extended the model by adding their hypothesized moderating effect. The present study tested 1) whether therapist and team experience with MST, and country-wide experience (i.e. time since implementation of MST) predicted therapist adherence; 2) whether therapist adherence predicted post-treatment outcomes (adolescent living at home, having had no police contact, and going to school or work); 3) whether the associations between therapist, team, and country-wide experience, and post-treatment outcomes were mediated by therapist adherence; and 4) whether the association between therapist experience and therapist adherence was moderated by team and country-wide experience. As such, we aimed to evaluate how experience with the treatment model at different levels (therapist, team, and country-wide) related to therapist adherence and treatment outcome.

Method

Participants and Procedures

Adolescents. A total of 5435 adolescents and their families completed MST between September 2004 and October 2014 in the Netherlands. Of these, 1145 families were excluded as they did not have any valid adherence assessments, resulting in a final sample of 4290 clients (79% of the total sample). No adolescents needed to be excluded due to missing data on the post-treatment outcomes, since completion of this information was a prerequisite for case closure.

The group excluded from analyses was compared with the sample included in the study (see Table 1). Adolescents excluded from analyses did not differ significantly from the study sample on sex, therapist experience, or team experience. However, the families excluded from the study were more likely to close their treatment due to lack of engagement or placement of the adolescent in a restrictive setting, and were more likely to have negative post-treatment outcomes. Adolescents excluded from the study were also older than adolescents in the study sample, and therapists treating these excluded adolescents started earlier in the implementation process than therapists treating included families. No information was available on ethnicity or socioeconomic status of the families.

Therapists and Teams. Initially, 4 Dutch MST teams, divided over 2 organizations, started providing MST. During the period under study, the number of teams grew to 27 divided over 8 organizations. Two teams switched to another organization during the course of this study. These teams were considered as the same team over the whole period of data collection.

In total, 222 therapists and 48 supervisors, supervised by 13 consultants, provided MST over the course of this study. Although no information was available about their level of education, therapists should have completed higher education in a relevant domain to qualify for MST therapist in the Netherlands.

Table 1 Descriptive statistics for families included in and excluded from the study sample

	Families included in study	**Families excluded from study**	***p***
N	*4290*	*1145*	
Dichotomous variables	*%*	*%*	
Gender (male)	70%	69%	.80
Interview language (Dutch)	87%	-	-
Home	91%	80%	.00
School/Work	83%	75%	.00
No new arrests	86%	82%	.00
Continuous variables	*M (SD)*	*M (SD)*	*p*
Age	15.62 (1.38)	15.78 (1.34)	.00
Therapist experience	16.37 (13.78)	16.23 (15.35)	.78
Team experience	3.22 (2.23)	3.10 (2.70)	.17
Country-wide experience	3.71 (2.20)	3.11 (2.25)	.00
Therapist adherence	4.32 (0.53)	-	-

Procedures. Families were referred to MST due to severe externalizing behavioral problems of the adolescent, such as delinquency, problems at school, or risk of out-of-home placement. Families had to meet the MST inclusion criteria, which have been specified by MST Services, the international licensor for the dissemination of MST (MST Services, 2014). Case enrollment and discharge information, as well as the TAM / TAM-R were entered into the MST Institute website (MSTi; www.MSTInstitute.org) by a staff person working in the organization that housed the MST team. The collected information from MSTi consisted of start and end date of the treatment, TAM interview language, gender and date of birth of the adolescent, and therapist-reported post-treatment outcomes. As this study used data retrospectively, formal consent was not required.

Measures

All measures in the current study were defined in the same way as in the original study by Löfholm and colleagues (2014). In cases where this was not possible, this is explicitly mentioned.

Therapist Adherence. The original Therapist Adherence Measure (TAM; Henggeler & Borduin, 1992) was developed to monitor therapist adherence to the MST model and consisted of 26 items rated on a 5-point Likert scale (1 = not at all, 2 = a little, 3 = some, 4 = pretty much, and 5 = very much). The TAM was scored by the primary caregiver who was called on a monthly basis by an agency staff other than the family's therapist. Items assessed therapist adherence to the MST clinical process and the treatment principles of MST, such as 'The therapist tried to understand how my family's problems all fit together' and 'The therapist's recommendations required family members to work on our problems almost every day'.

Later improvements of the TAM led to inclusion of an additional 12 items assessing whether the treatment focused on important aspects of the adolescent's school, peer, and neighborhood, consistent with the MST model. Psychometric analyses of this new set of items resulted in a revised instrument, the Therapist Adherence Measure-Revised (TAM-R; Henggeler et al., 2006a), consisting of 19 of the original 26 items and 9 of the additional 12 items (Schoenwald et al., 2008a). These 28 items were rated on the same 5-point Likert scale as the original TAM.

Predictive validity and reliability of the TAM and TAM-R were assessed during initial randomized clinical trials of MST, as well as later transportability studies (Henggeler et al., 1997, 2002; Schoenwald et al., 2003, 2008;). Discriminant validity was supported by findings that the TAM discriminated between MST and treatment as usual (Henggeler et al., 2006b).

The Dutch TAM was introduced in the Netherlands after translation and back-translation by two independent translation offices and after approval by MST Services at the end of 2004. The revised TAM-R was introduced in 2007. The English and Dutch TAM-R can be

found on http://www.mstinstitute.org/qa_program/tam_languages.shtml. Since the TAM-R was introduced halfway the period under study, the analyses were conducted on the 19 items overlapping in both the TAM and TAM-R. In the current study, the average therapist-adherence score for each family based on these 19 items correlated highly ($r = .99$) with the average therapist-adherence score per family based on all 28 items.

Only valid assessments (assessments with a maximum of three missing items, and where face-to-face contact between the family and the therapist had occurred in the last 2 weeks prior to administration of the TAM or TAM-R) were included for analyses. The mean number of adherence-reports provided by the caregivers was 3.5 ($SD = 1.6$). Since TAM ratings have been found to be stable within a family's treatment episode (Schoenwald, 2008) and the internal consistency of the 19 items was high ($\alpha = .92$), a mean adherence-score could be calculated for each family. This score represented the mean of all completed reports by the family during an MST treatment episode and represented the mean level of therapist adherence as experienced by this family.

Country-Wide Experience. Country-wide experience was defined as the number of years since MST had been implemented in the Netherlands at the time the therapist first started providing MST. This variable was represented at the cohort level (see data analysis strategy below for a definition of cohort). It was comparable to the dichotomous variable 'wave' in the Swedish study by Löfholm and colleagues (2014), the only difference being that our variable was continuous instead of binary. Since the implementation of MST in the Netherlands was not characterized by two consecutive waves, as was the case in Sweden, the current continuous operationalization was chosen. As a 10-year period was used in the present study, country-wide experience ranged from 0 to 9 years.

Team Experience. Team experience was defined as the number of years a team had been active in providing MST at the time the family began treatment. Team experience ranged from 0 to 9 years.

Therapist Experience. Therapist experience was defined as the number of previous families to whom the therapist had provided MST. This score ranged from 0 to 81 families.

TAM Interview Language. MST was provided to families with divergent ethnic backgrounds, including families for whom Dutch was not their first language. In these cases the TAM/TAM-R could be administered in another than the Dutch language. Language was dichotomized in Dutch or a different language. If interview language had varied within a family over the course of treatment, a family was categorized as Dutch only if all of the TAM/TAM-R administrations had been in Dutch.

Post-Treatment Outcomes. Post-treatment outcomes were reported by the therapist at the completion of treatment and consisted of three dichotomous outcomes, namely whether the adolescent (a) lived at home (i.e. all stable home situations, including, but not restricted to, living with grand parents or foster parents), (b) was engaged in school (no

truancy) or work (at least 20 h a week), and (c) had not been arrested. The first two outcomes represented the situation at the end of treatment. However, if the adolescent was placed out of home during treatment, MST was stopped, leading to a negative treatment outcome for this MST episode. The third outcome referred to all arrests during the MST treatment episode. These three outcomes are being used by MST as ultimate outcomes and should be attained by all adolescents at the end of treatment (MST Institute, 2016). All three outcomes have been operationalized and standardized by MST Services to ensure that these outcomes are being scored in the same way by all therapists. As MST is a community-based treatment, MST therapists have connections with all relevant youth services in their working area, and can easily request information to validate their scores.

Data Analyses Strategy

Analyses were performed in Mplus 7.3 (Muthén & Muthén, 1998-2012) for multilevel structural equation modeling. The amount of missing data was minimal (gender: 5%; language: 5%) and were taken into account using robust full maximum likelihood (MLR). Also any deviates from normality were addressed with MLR, as MLR is a robust estimator for non-normal and dependent data using all available data.

The model consisted of three levels: Families (level 1; N=4290) were nested within 'cohorts' (all the families seen by the same therapist in the same 'team-experience' year; level 2; N=816), and cohorts were nested within therapists (level 3; N=222). Cohorts were included as a level as it was assumed that families treated by the same therapist in the same year would be more similar than families treated by the same therapist a couple of years later, when therapist and team experience would have increased. To account for the non-independence in the data the TYPE = COMPLEX TWOLEVEL command in Mplus was used (Asparouhov & Muthén, 2006; Muthén & Muthén, 1998-2012). The COMPLEX command was used to adjust standard errors for non-independence within therapists. As such, non-independence within therapists was accounted for but not explicitly modeled. In contrast, non-independence within cohorts was explicitly included into the model using the TWOLEVEL command.

Analytical Model. The first aim of this study was to replicate the analytical model of Löfholm and colleagues (2014) and investigate the associations of therapist, team, and country-wide experience with therapist adherence and post-treatment outcomes. The model was specified as follows (see Fig. 2). At the family level (level 1), therapist experience, adolescent gender, and TAM-interview language were included as predictors of therapist-adherence scores (the mean therapist adherence level achieved in a family). Therapist adherence in turn was included as a predictor for the three post-treatment outcomes (adolescent living at home, engaged in school/work, and no new arrests). As the post-treatment outcomes were categorical, the resulting model was a logistic regression model.

A direct path was specified from adolescent gender to 'no new arrests', and from therapist experience to all three post-treatment outcomes. Further, the model allowed the intercept of therapist adherence to vary across cohort clusters. At the cohort level (level 2), team experience and country-wide experience were included as predictors of the intercept of therapist adherence. The adherence intercept represented the therapist's overall adherence score for all the families seen by that therapist in a single year of team experience. An additional 39 cohorts were created to account for cross-classified therapists (i.e. therapists that were members of two teams in that year).We also included direct paths from team experience and country-wide experience to the three treatment outcomes, and used the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for model selection. Rules of thumb suggest that an increase between four and seven for each additional parameter on the AIC and an increase between two and six for each additional parameter on the BIC may be positive evidence for the alternative model (Burnham & Anderson, 2004; Raftery, 1999).

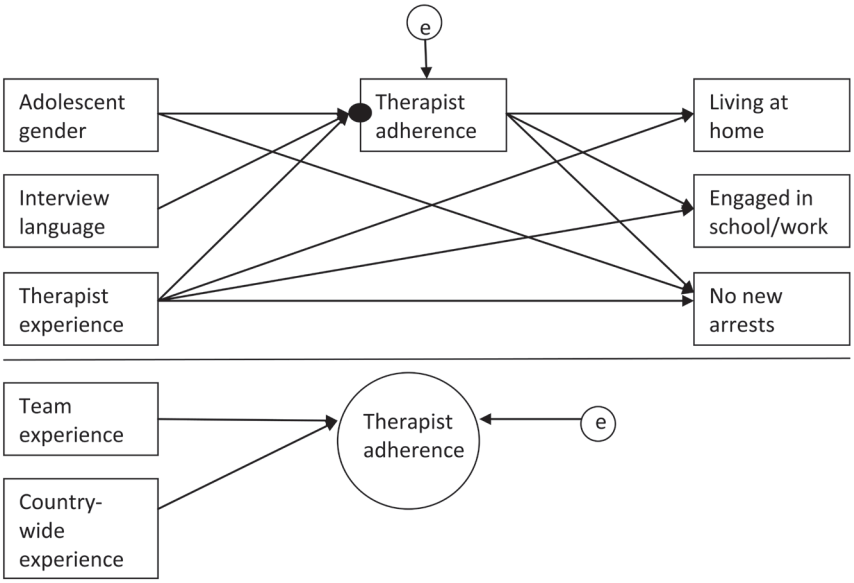

Fig. 2 Visual representation of original model

Note: Variables above the line were measured at the family level, whereas years of team experience was measured at the cohort level. The small filled circle represents the random intercept of therapist adherence, which was allowed to vary across cohorts. At the cohort level, therapist's yearly TAM/TAM-R is the TAM intercept-as-outcome. Circles with the letter 'e' are error terms that represent unexplained variance. The cohort-level error term for therapist adherence represents unexplained variance in its random intercept.

Mediation Analyses. We subsequently tested for indirect effects of therapist, team, and country-wide experience on post-treatment outcomes via therapist adherence. These analyses were only conducted if the individual paths from experience to adherence, and from adherence to outcome were significant. Indirect effects were identified using the joint significance test (MacKinnon et al., 2002). This test consists of adding the joint paths of experience-to-adherence and adherence-to-outcome to the model, and evaluates whether the combination of both these paths is significant. If these joint paths are significant, the association between experience and outcome is mediated through therapist adherence.

Moderator Analyses. Lastly, we extended the model with two cross-level interactions to investigate whether team or country-wide experience moderate the association between therapist experience and therapist adherence. These cross-level interactions were added to the model independently of one another, by adding a path from team and country-wide experience to the slope from therapist experience to therapist adherence. The AIC and BIC were used for model selection (Burnham & Anderson, 2004; Raftery, 1999).

Results

Table 2 provides an overview of the development of the therapist-adherence scores and post-treatment outcomes over the course of the study. Adherence scores and post-treatment outcomes increased in the first few years and then appeared to stabilize.

Preliminary Family-Level Model

First, we evaluated a single-level model, which consisted of only the family-level paths, to gain an initial idea about the Chi square-based model fit at the family level prior to building a multilevel model. Model fit was evaluated using a robust weighted least-squares estimator (weighted least squares mean and variance adjusted; Muthén & Muthén, 1998–2012). Although the Chi square value was significant ($\chi 2 = 14.75$, $p = .04$), indicating bad model fit, all other model fit indices met the criteria for good fit (CFI = .99, TLI = .96, RMSEA = .02). As the Chi square value is sensitive to large sample-sizes, and all other model fit indices were good, we assumed good model fit for the single-level model.

Table 2 Development of adherence scores and post-treatment outcomes over time

Year of implementation	Year 1&2	Year 3&4	Year 5&6	Year 7&8	Year 9&10
N	*182*	*704*	*1075*	*1270*	*1059*
Mean and *SD* of therapist adherence (19 items)	3.93 (0.47)	4.35 (0.49)	4.36 (0.51)	4.31 (0.55)	4.34 (0.53)
Home	84%	89%	91%	92%	92%
School/work	77%	81%	83%	84%	84%
No new arrests	75%	83%	84%	86%	92%

Analytical Model

Predictors of Therapist Adherence. Table 3 presents the parameter estimates for family- and cohort-level paths specified in the model. At the family-level, therapist adherence was predicted by therapist experience and language, but not the gender of the adolescent. Therapist adherence was higher when the therapist had more experience and the TAM-R was assessed in another than the Dutch language. On average, therapist-adherence scores increased with 1.25% (an increase of 0.05 on a scale of 1 to 5) for an additional ten families treated (approximately one additional year of experience). At the cohort-level, country-wide experience predicted therapist adherence. A therapist starting 1 year later in the implementation process acquired therapist-adherence scores that were 0.5% higher (an increase of 0.02) than therapists starting 1 year earlier in the implementation process.

Predictors of Post-Treatment Outcomes. All three post-treatment outcomes were predicted by therapist adherence (see Table 3). This indicates that higher therapist adherence increased the odds that the adolescent was living at home at the end of the treatment (*OR* = 1.49, 95% CI [1.27,1.74]), that the adolescent was engaged in school or work (*OR* = 1.78, 95% CI [1.53, 2.06]), and that the adolescent had not been arrested during the course of the treatment (*OR* = 1.33, 95% CI [1.12, 1.58]). As causality could be reversed, with behavioral change during treatment leading to higher therapist adherence instead of adherence leading to better treatment outcomes, we conducted a sensitivity analysis to see whether early treatment therapist adherence similarly predicted treatment outcome. For this purpose, we replaced the mean therapist adherence score in the model by adherence as assessed during the first month of MST. This did not change any of the model results. Therapist adherence increased the odds that the adolescent was living at home at the end of treatment (OR = 1.69, 95% CI [1.46, 1.94]), that the adolescent was engaged in school or work (OR = 1.79, 95% CI [1.55, 2.06]), and that the adolescent had not been arrested during the course of the treatment (OR = 1.28, 95% CI [1.10, 1.49]).

Table 3 Parameter estimates original model for 19-item adherence scores

Outcome	Predictor	Estimate	95% CI	*p*	Standardized effect
			Level 1		
Therapist adherence	Therapist experience	0.005	[0.002, 0.007]	.00	0.14
	Female	0.03	[-0.01, 0.06]	.16	0.02
	Dutch language	-0.07	[-0.12, -0.02]	.00	-0.05
Home	Therapist adherence	0.40	[0.24, 0.56]	.00	0.11
	Therapist experience	0.003	[-0.005, 0.01]	.48	0.02
School/work	Therapist adherence	0.58	[0.42, 0.73]	.00	0.15
	Therapist experience	0.006	[-0.001, 0.01]	.10	0.04
No new arrests	Therapist adherence	0.29	[0.11, 0.46]	.00	0.07
	Female	0.97	[0.70, 1.24]	.00	0.24
	Therapist experience	0.01	[0.002, 0.02]	.02	0.08
			Level 2		
Mean yearly therapist adherence	Team experience	-0.02	[-0.03, 0.001]	.06	-0.18
	Country-wide experience	0.02	[0.004, 0.03]	.02	0.20

Note: Parameter estimates for therapist adherence are linear regression coefficients. Estimates for post-treatment outcomes (home, school/work, no new arrests) are logistic regression coefficients. Standardized effects for continuous predictors are path coefficients standardized with respect to both predictor and outcome. Standardized effects for dichotomous predictors are path coefficients standardized with respect to the outcome only.

Higher therapist experience only increased the odds for not having been arrested (OR = 1.01, 95%CI [1.00, 1.02]). An odds ratio of 1.01 for 'no new arrests' indicated that a 1-unit increase in therapist experience (one additional family treated) was associated with a 1% increase in the odds that the adolescent had not been arrested during the course of the treatment. Thus, after approximately 1 year of additional therapist experience (ten additional families treated) the odds of 'no new arrests' increased with 11%. Moreover, girls had higher odds of not having been arrested during the course of the treatment than boys (OR = 2.63, 95% CI [2.01, 3.44]).

We also tested the direct effect of team experience and country-wide experience on post-treatment outcomes. AIC and BIC substantially improved when including those direct paths (Δ AIC = 92, Δ BIC = 34, Δ df = 11). Results are presented in Table 4. Team experience only predicted 'engaged in school/work' (B = -0.07, p < .01), indicating that one additional

year of team experience was associated with a 7% decrease in the odds that the adolescent was engaged in school or work. Country-wide experience significantly predicted 'engaged in school/work' ($B = 0.09$, $p < .001$) and 'no new arrests' ($B = 0.16$, $p < .01$). One additional year of country-wide experience was associated with a 9% increase in the odds that the adolescent was engaged in school/work, and a 16% increase in the odds that the adolescent had no new arrests.

Mediation Analyses

Indirect effects of experience on post-treatment outcomes through therapist adherence were tested for therapist experience and country-wide experience. We did not include team experience in these analyses as team experience did not significantly predict therapist adherence. Monte Carlo confidence intervals were computed using the web utility of Selig and Preacher (2008), as these non-symmetric intervals are more appropriate when the assumption of a normal distribution for the indirect effects may not hold (Preacher & Selig, 2012). Indirect effects of therapist experience were significant for all three post-treatment outcomes ($B = .002$, 95% CI [0.000, 0.003], $p < .05$ for 'living at home'; $B = .003$, 95% CI [0.001, 0.004], $p < .01$ for 'engaged in school/work'; $B = .001$, 95% CI [0.000, 0.002], $p < .05$ for 'no new arrests'). Indirect effects of country-wide experience were only significant for 'living at home' ($B = .008$, 95% CI [0.001, 0.016], $p < .05$) and 'engaged in school / work' ($B = .01$, 95% CI [0.002, 0.022], $p < .05$), but not for 'no new arrest' ($B = .005$, 95% CI [0.001, 0.012], $p = .07$).

Moderator Analyses

Team experience and country-wide experience were independently included as a moderator of the slope from therapist experience to therapist adherence. AIC and BIC values were higher when including the moderator paths, indicating worse fit. Inspection of these models showed that residual variance of the slope was zero, indicating that there was no random effect of therapist experience on therapist adherence. Therefore we concluded that the effect of therapist experience on therapist adherence was equal across the whole range of team experience and country-wide experience. Thus, we did not find a moderator effect of team or country-wide experience on the association between therapist experience and adherence.

Table 4 Parameter estimates expanded model, including direct effects on outcome at Level 2

Outcome	Predictor	Estimate	95% CI	*p*	Standardized effect
			Level 1		
Therapist adherence	Therapist experience	0.004	[0.002, 0.007]	.00	0.13
	Female	0.02	[-0.01, 0.06]	.21	0.02
	Dutch language	-0.07	[-0.12, -0.03]	.00	-0.05
Home	Therapist adherence	0.40	[0.23, 0.56]	.00	0.11
	Therapist experience	0.01	[-0.003, 0.02]	.09	0.08
School/work	Therapist adherence	0.58	[0.42, 0.73]	.00	0.15
	Therapist experience	0.02	[0.01, 0.03]	.00	0.12
No new arrests	Therapist adherence	0.26	[0.08, 0.44]	.01	0.07
	Female	1.02	[0.75, 1.33]	.00	0.25
	Therapist experience	0.02	[0.01, 0.03]	.00	0.13
			Level 2		
Therapist adherence	Team experience	-0.01	[-0.03, -0.002]	.08	-0.16
	Country-wide experience	0.02	[0.004, 0.03]	.02	0.20
Home	Team experience	-0.05	[-0.11, 0.02]	.15	-0.21
	Country-wide experience	0.08	[-0.02, 0.18]	.10	0.33
School/work	Team experience	-0.07	[-0.12, -.02]	.00	-0.38
	Country-wide experience	0.09	[0.05, 0.14]	.00	0.47
No new arrests	Team experience	0.03	[-0.03, 0.08]	.36	0.08
	Country-wide experience	0.16	[0.06, 0.26]	.00	0.46

Note: Parameter estimates for therapist adherence are linear regression coefficients. Estimates for post-treatment outcomes (home, school/work, no new arrests) are logistic regression coefficients. Outcomes at level 2 were mean yearly scores. Standardized effects for continuous predictors are path coefficients standardized with respect to both predictor and outcome. Standardized effects for dichotomous predictors are path coefficients standardized with respect to the outcome only.

Discussion

The current study replicated the mediation model of Löfholm and colleagues (2014), examining whether (1) therapist, team, and country-wide experience predicted therapist adherence, (2) therapist adherence predicted post-treatment outcomes (adolescent living at

home, having had no police contact, and going to school or work), and (3) the associations between therapist, team, and country-wide experience, and post-treatment outcomes were mediated by therapist adherence. Moreover, we extended the model with a moderator path, resulting in a fourth question: We investigated whether the association between therapist experience and therapist adherence was moderated by team and country-wide experience. Our findings indicate that (1) therapist and country-wide experience, but not team-experience, predict therapist adherence; (2) therapist adherence predicts all three post-treatment outcomes; (3) therapist adherence also mediates the associations between therapist and country-wide experience on the one hand, and treatment outcome on the other hand, and (4) the association between therapist experience and therapist adherence is not moderated by team or country-wide experience. Contrary to the study by Löfholm and colleagues (2014), we found therapist experience instead of team experience to predict therapist adherence and treatment outcome. The role of country-wide experience in our study was similar to its role in the Swedish study: More country-wide experience is associated with higher therapist adherence, and therapist adherence mediates the effects of country-wide experience on post-treatment outcomes.

The finding that country-wide experience is a significant predictor of therapist adherence and treatment outcome corresponds with previous studies underscoring the relevance of outer contextual factors for implementation of evidence-based interventions. Such factors include interorganizational networks, connections with evidence-based intervention developers (Fixsen et al., 2005; Novins et al., 2013), but also the use of quality-control methods, such as the MST quality-assurance system (Henggeler et al., 2008; Holth et al., 2011; Smith-Boydston et al., 2014). Methods such as adherence monitoring, training, and supervision may be especially relevant, as these are directly targeted at sustaining adequate therapist adherence and positive treatment outcomes over time (Bond et al., 2014; Fixsen et al., 2005; Garland & Schoenwald, 2013; Novins et al., 2013). Moreover, these methods may be particularly sensitive to country-wide experience. The director of MST-the Netherlands, the Dutch Network Partner licensed to implement MST, has suggested that, over the years, MST-the Netherlands has gained a better understanding of the core components of MST and has improved its ability to deploy the quality-assurance system (personal communication with Wim van Geffen, June 2015).

This study suggests that differences in therapist-adherence scores between different countries, as have been found for MST adherence scores (Lange et al., 2015; MST Institute, 2010), may partly be attributable to differences in country-wide experience of supporting adherence. Nevertheless, one must bear in mind that many factors may affect therapist adherence scores. For example, cultural differences in response style have been proposed as an alternative explanation for the differences observed between Dutch and American adherence scores (Lange et al., 2015). As the current study was not designed to provide

evidence for the origin of these differences, it does not allow drawing any definite conclusions on the role of country-wide experience in this.

Unlike the Löfholm et al. study (2014) report, we found that therapist experience was a significant predictor of therapist adherence. Our findings are similar to recent analyses of 2298 MST therapists worldwide, showing that therapists with more MST experience achieved higher adherence scores than therapists with less experience (MST Institute, 2014). Nevertheless, the effect of therapist experience on adherence was small in our study. This may be a consequence of the overall high adherence scores, an effect also reported in the study by Löfholm and colleagues (2014). This left little room for these scores to improve any further, which may explain the absence of a significant effect in the Löfholm report.

The association between therapist experience and therapist adherence was not moderated by country-wide experience. This may be due to a similar ceiling effect of therapist-adherence scores as described above. If so, studies reporting a wider range of adherence scores, for example due to implementation difficulties, may be better capable of illuminating these associations.

Contrary to the study by Löfholm and colleagues (2014), team experience did not predict therapist adherence in our study. Although, so far, we remain uncertain as to the cause of this difference, various explanations are possible. For example, turnover of therapists, supervisors, or consultants may have been greater in the Dutch teams than in the Swedish ones. Thus, it may not be the experience of the team, but rather the experience of the supervisor which is predictive of therapist adherence. Previous research has shown how supervisor adherence is related to therapist adherence, suggesting that supervisor behavior is related to therapist behavior (Schoenwald et al., 2009b). Another unexpected finding was the negative association between team experience and the engagement of the adolescent in school or work. As this was the only significant association of team experience, this may be a Type I error. However, the association between team experience and therapist adherence was also negative, albeit non-significant (p=.06). Anecdotal evidence suggests that, at least in some teams, more team experience may lead supervisors and their teams to pay less attention to the adherence scores, which may result in lower therapist adherence and subsequently lower treatment outcomes.

Limitations

We have suggested that the effect of country-wide experience on therapist adherence and treatment outcome may reflect increasing experience of training and supporting MST therapists, indicating successful implementation of MST. Nevertheless, therapist adherence is only one aspect of treatment fidelity. Assessing a broader range of treatment fidelity measures may provide a more complete picture of how experience is related to successful implementation. For example, Brunk and colleagues (2014) have developed an index of

treatment fidelity for MST, including not only measures of therapist adherence, but also items on critical program practices, such as organizational support and essential clinical operations. Future research may benefit from using such an index.

The families excluded from our study differed with regard to some key variables from the families included for study. Most notably, excluded families had poorer post-treatment outcomes than the study sample, and MST had been implemented for a shorter period of time when their therapists were engaged as an MST therapist. Therefore, we hypothesize that excluded families reported lower therapist adherence scores than included families. This may have restricted the range of adherence scores in our study and, as such, hampered the likelihood of finding strong effects. However, we have no reason to believe it influenced the direction of our effects.

As we did not have information regarding behavioral change during treatment, we cannot rule out the possibility that initial behavioral change led to higher therapist adherence scores instead of the adherence predicting the positive treatment outcomes. Nevertheless, we have several indications to assume that the direction of the effects is as discussed. We felt safe using average family adherence scores because these scores have been found to be stable within families (Schoenwald, 2008a). The sensitivity analysis showed that early treatment adherence similarly predicted treatment outcomes. Moreover, the results of this study are in line with numerous previous studies that demonstrated the association between adherence and treatment outcome within MST (e.g. Huey et al., 2000; Löfholm et al., 2014; Schoenwald et al., 2009a).

The aim of the current study was to replicate previous findings to investigate whether these findings are robust across different countries that are characterized by different health care settings, funding agencies etc.. Schoenwald colleagues (2008b and) have described how MST needs to be adapted at several levels to be suitable for implementation in other countries. As a consequence, findings from one country cannot be assumed to apply to other countries. Replicating studies across countries can help researchers and practitioners understand what factors affect implementation success and, more specifically, therapist adherence. This study confirmed previous findings on the relevance of country-wide experience, but also calls for more research into the role of team experience, as the findings of team experience in relation to therapist adherence and outcome were opposite to previous research.

Conclusions

This study showed that therapist experience as well as country-wide experience matters for sustaining therapist adherence and achieving favorable treatment outcomes. Implementing an intervention with high adherence is not an easy task; stakeholders at different levels need

to acquire experience in delivering the intervention with high adherence or in supporting therapists to do so. Using a quality-assurance system may be essential to sustaining therapist adherence and warranting good treatment outcomes.

Acknowledgement

We would like to thank MST-the Netherlands as well as all Dutch mental health institutions providing MST for their cooperation in data collection, and the MST Institute for making the data available. We are also grateful to Dr. Eichas and Dr. Löfholm for answering all our questions and providing us with the required information for replication of their study.

Compliance with Ethical Standards

Ethical approval: All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.
Informed consent: As this study used data retrospectively, formal consent was not required.

References

Asparouhov, T., & Muthén, B. (2006). Multilevel modeling of complex survey data. *Proceedings of the joint statistical meeting in seatle, August 2006. ASA Section on Survey Research Methods*. (pp. 2718-2726).

Bond, G. R., Drake, R. E., McHugo, G. J., Peterson, A. E., Jones, A. M., & Williams, J. (2014). Long-term sustainability of evidence-based practices in community mental health agencies. *Administration and Policy In Mental Health*, 41, 228-236.

Brunk, M. A., Chapman, J. E., & Schoenwald, S. K. (2014). Defining and evaluating fidelity at the program level in psychosocial treatments: A preliminary investigation. *Zeitschrift für Psychologie, 222*, 22-29.

Burnham, K. P., & Anderson, D. R. (2004). Multimodal inference. Understanding AIC and BIC in model selection. *Sociological Methods and Research, 33*, 261-304.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327-350.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Forgatch, M. S., Patterson, G. R., & DeGarmo, D. S. (2005). Evaluating fidelity: Predictive validity for a measure of competent adherence to the Oregon model of parent management training. *Behavior Therapy, 36*, 3-13.

Garland, A., & Schoenwald, S. K. (2013). Use of effective and efficient quality control methods to implement psychosocial interventions. *Clinical Psychology: Science and Practice, 20*, 33-43.

Henggeler, S. (2011). Efficacy studies to large-scale transport: The development and validation of Multisystemic Therapy programs. *Annual Review of Clinical Psychology, 7*, 351-381.

Henggeler, S. W., & Borduin, C. M. (1992). *Multisystemic therapy adherence scale*. Unpublished instrument. Charleston, SC: Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina.

Henggeler, S. W., & Schoenwald, S. K. (1999). The role of quality assurance in achieving outcomes in MST programs. *Journal of Juvenile Justice and Detention Services, 14(2)*, 1-17.

Henggeler, S. W., & Sheidow, A. J. (2012). Empirically supported family-based treatments for conduct disorder and delinquency in adolescents. *Journal of Marital and Family Therapy, 38*, 30-58.

Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic Therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65*, 821-833.

Henggeler, S. W., Schoenwald, S. K., Liao, J. G., Letourneau, E. J., & Edwards, D. L. (2002). Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs. *Journal of Clinical Child Psychology, 31*, 155-167.

Henggeler, S. W., Borduin, C. M., Schoenwald, S. K., Huey, S. J., & Chapman, J. E. (2006a). *MST therapist adherence measure-revised (TAM-R)*. Charleston: Medical University of South Carolina, Family Services Research Center.

Henggeler, S. W., Halliday-Boykins, C. A., Cunningham, P. B., Randall, J., Shapiro, S. B., & Chapman, J. E. (2006b). Juvenile drug court: Enhancing outcomes by integrating evidence-based treatments. *Journal of Consulting and Clinical Psychology*, *74*, 42-54.

Henggeler, S. W., Sheidow, A. J., Cunningham, P. B., Donohue, B. C., & Ford, J. D. (2008). Promoting the implementation of an evidence-based intervention for adolescent marijuana abuse in community settings: Testing the use of intensive quality assurance. *Journal of Clinical Child and Adolescent Psychology, 37*, 682-689.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (2009). *Multisystemic therapy for antisocial behavior in children and adolescents*. New York: Guilford.

Holth, P., Torsheim, T., Sheidow, A. J., Ogden, T., & Henggeler, S. (2011). Intensive quality assurance of therapist adherence to behavioral interventions for adolescent substance use problems. *Journal of Child and Adolescent Substance Abuse, 20*, 289-313.

Huey, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in Multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology, 68*, 451-467.

Lange, A. M., Scholte, R. H., van Geffen, W., Timman, R., Busschbach, J. J., & van der Rijken, R. E. A. (2015). The lack of cross-national equivalence of a therapist adherence measure (TAM-R) in Multisystemic Therapy (MST). *European Journal of Psychological Assessment*. doi: 10.1027/1015-5759/a000262.

Löfholm, C. A., Eichas, K., & Sundell, K. (2014). The Swedish implementation of Multisystemic Therapy for adolescents: Does treatment experience predict treatment adherence? *Journal of Clinical Child and Adolescent Psychology, 43*, 643-655.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83-104.

McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodriguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. *Clinical Psychology: Science and Practice, 20*, 14-32.

Mihalic, S. (2004). The importance of implementation fidelity. *Report on Emotional and Behavioral Disorders in Youth, 4*, 83-105.

MST Institute (2010). *2010 MST Data Report*. Retrieved July 2015 from http://www.mstinstitute.org.

MST Institute (2014). *2014 MST Data overview report*. Retrieved July 2015 from http://www.mstinstitute.org.

MST Institute (2016). *Frequently Asked Questions*. *MST Institute Enhanced Website*. Retrieved May 2016 from https://www.msti.org/documents/EW_FAQs.pdf.

MST Services (2014*). Multisystemic Therapy® (MST®) Organizational Manual*. Charleston, SC: Author.

Muthén, L., & Muthén, B. O. (1998-2012). *Mplus user's guide*. *Seventh Edition*. Los Angeles, CA: Author.

Novins, D. K., Green, A. E., Legha, R. K., & Aarons, G. A. (2013). Dissemination and implementation of evidence-based practices for child and adolescent mental health: A systematic review. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*, 1009-1025.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change. Issues and research recommendations. *Clinical Psychology: Science and Practice, 12*, 365-383.

Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures, 6*, 77-98.

Raftery, A. E. (1999). Bayes factors and BIC. *Sociological Methods and Research, 27*, 411-427.

Schoenwald, S. K. (2008). Toward evidence-based transport of evidence-based treatments: MST as an example. *Journal of Child and Adolescent Substance Abuse, 17*, 69-91.

Schoenwald, S. K., & Garland, A. F. (2013). A review of treatment adherence measurement methods. *Psychological Assessment, 25*, 146-156.

Schoenwald, S. K., Sheidow, A. J., Letourneau, E. J., & Liao, J. G. (2003). Transportability of Multisystemic Therapy: Evidence for multilevel influences. *Mental Health Services Research, 5*, 223-239.

Schoenwald, S. K., Carter, R. E., Chapman, J. E., & Sheidow, A. J. (2008a). Therapist adherence and organizational effects on change in youth behavior problems one year after Multisystemic Therapy. *Administration and Policy in Mental Health and Mental Health Services Research, 35*, 379-394.

Schoenwald, S. K., Heiblum, N., Saldana. L., & Henggeler, S. W. (2008b). The international implementation of Multisystemic Therapy. *Evaluation & the Health Professions, 31*, 211-225.

Schoenwald, S. K., Chapman, J. E., Sheidow, A. J., & Carter, R. E. (2009a). Long-term youth criminal outcomes in MST transport: The impact of therapist adherence and organizational climate and structure. *Journal of Clinical Child and Adolescent Psychology, 38*, 91-105.

Schoenwald, S. K., Sheidow, A. J., & Chapman, J. E. (2009b). Clinical supervision in treatment transport: Effects on adherence and outcomes. *Journal of Consulting and Clinical Psychology, 77*, 410-421.

Selig, J. P., & Preacher, K. J. (2008). Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects [Computer software]. Available from http://quantpsy.org/.

Shirk, S. R., & Peterson, E. (2013). Gaps, bridges, and the bumpy road to improving clinic-based therapy for youth. *Clinical Psychology: Science and Practice, 20*, 107-113.

Smith-Boydston, J. M., Holtzman, R. J., & Roberts, M. C. (2014). Transportability of Multisystemic Therapy to community settings: Can a program sustain outcomes without MST Services oversight? *Child and Youth Care Forum, 43*, 596-605.

Southam-Gerow, M. A., & McLeod, B. D. (2013). Advances in applying treatment integrity research for dissemination and implementation science: Introduction to special issue. *Clinical Psychology: Science and Practice, 20*, 1-13.

Weisz, J. R., Ugueto, A. M., Cheron, D. M., & Herren, J. (2013). Evidence-based youth

psychotherapy in the mental health ecosystem. *Journal of Clinical Child and Adolescent Psychology, 42*, 274-286.

II

People or protocol:

The unique and joint contribution of adherence and alliance to optimal treatment outcomes

Chapter 4

Factorial structure of the Therapist Adherence Measure-Revised (TAM-R) within Multisystemic Therapy

Submitted as **Lange, A. M. C.**, Delsing, M.J.M.H., Scholte, R.H.J., & van der Rijken, R.E.A. Factorial structure of the Therapist Adherence Measure-Revised (TAM-R) within Multisystemic therapy. *Submitted*

Abstract

The Therapist Adherence Measure (TAM-R) is a central assessment within the quality-assurance system of Multisystemic Therapy (MST). MST is an intensive, home- and community-based treatment for adolescents (12-18 years old) showing externalizing behavioral problems or delinquency. Although the validity and reliability of the TAM have been examined in several US studies, these studies found varying numbers of latent factors underlying the scores on the TAM. Therefore, the current study aimed to re-examine its factor structure in the Netherlands. For this purpose, we used two independent samples of families participating in MST in the Netherlands. The factor structure was explored using a Principal Component Analysis (PCA) in Sample 1 (N=580). As the TAM-R was completed multiple times by most families, we could replicate our PCA-analyses across two random subsamples of questionnaires from Sample 1, each consisting of 1 TAM-R assessment per family. A two-component solution showed to be the most appropriate. The two components were labeled therapist adherence and client-therapist alliance, respectively. Cross-loading items were dropped to create two well-differentiating components. Internal consistency of the reduced components was good. This two-factor model also showed good model fit in a subsequent Confirmatory Factor Analysis (CFA) in Sample 2 (N=723). The current finding of an alliance-component corroborates previous studies and fits with the focus of the MST-treatment model on creating engagement. Although only a subset of the items was retained for the final component-solution, MST clinicians are advised to continue to rely on the full TAM-R for feedback and quality assurance.

MST is an intensive, home- and community-based treatment for adolescents (12-18 years old) showing externalizing behavioral problems or delinquency (Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 2009). MST improves family functioning, decreases delinquency, and decreases the number of out-of-home placements (see Asscher et al., 2014 and Deković, Asscher, Manders, Prins, & Van der Laan, 2012 for results of MST in the Netherlands, and Henggeler, 2011 for an international overview of MST treatment and results).

To ensure successful dissemination and implementation of MST at sites all over the world, an elaborate quality-assurance system was developed, consisting, among other things, of training, supervision and consultation, and continuous data-monitoring (Henggeler & Schoenwald, 1999; Schoenwald, 2008). The Therapist Adherence Measure (TAM, later TAM-R) is a central assessment within this quality-assurance system, as the therapist is the primary linkage with the family and, therefore, critical in achieving desired outcomes for youths and families (Henggeler & Schoenwald, 1999). The TAM-R assesses adherence to the MST clinical process and the nine principles of MST.

The (factorial) validity and reliability of the TAM have been examined in several US studies. These studies have found varying numbers of latent factors underlying the scores on the TAM. As sample size increased, the number of factors decreased from six (Henggeler, Melton, Brondino, Scherer, & Hanley, 1997) to three (Henggeler, Schoenwald, Liao, Letourneau, & Edwards, 2002) to one (Letourneau, Sheidow, & Schoenwald, 2002). This single-factor solution has, however, been criticized and an alternative two-factor solution has been proposed (Ellis, Weiss, Han, & Gallop, 2010). Later studies have used a revised version of the TAM (i.e., the TAM-R, consisting of 19 of the original items combined with 9 new items) and referred to Rasch-based analyses supporting a single factor, but without providing the technical details or results. Facing this uncertainty regarding the structure of the TAM-R, the current study sought to re-analyze its factor structure in the Netherlands.

Methods

Participants and Measures

Two samples were used for this study. In both samples, we only included TAM-R data that was collected through an independent call center to ensure all questionnaires were administered in the same way.

The first sample consisted of families who finished MST between July 2009 and November 2011. Families who did not complete the TAM-R in Dutch were excluded. This resulted in a sample of 580 families who had been treated by 63 different therapists in the Netherlands.

The second sample consisted of families having started an MST treatment in 2012 or 2013. This resulted in a sample of 723 families who had been treated by 88 different therapists in the Netherlands. As the data collection was part of clinical practice and the data was provided anonymously to the researchers (i.e., it concerned retrospective file data), no informed consent was required.

The *Therapist Adherence Measure-Revised* (TAM-R; Henggeler, Borduin, Schoenwald, Huey, & Chapman, 2006) consists of 28 items, which can be scored on a 5-point Likert scale (ranging from 1 – 'not at all', to 5 – 'very much'). The TAM-R is completed on a monthly basis during treatment by the primary caregiver through a telephone interview by an independent call center. As treatment duration is 3-5 months families could have up to 5 TAM-Rs. Some families only had one TAM-R assessment, for example due to early drop-out or non-response.

Analyses

Principal Component Analysis. The first sample, consisting of 580 families, was used to explore the underlying factor structure of the TAM-R using a Principal Component Analysis (PCA). The fact that the TAM-R was completed multiple times by most families enabled us to replicate our PCA-analyses across two random subsamples of questionnaires. These subsamples consisted of 1 TAM-R assessment per family. We made sure no TAM-R was included twice. A PCA with varimax rotation was conducted on each subsample. The following criteria were used to define the number of components (Floyd & Widaman, 1995): 1) The scree plot of eigenvalues, 2) the parallel analysis criterion, and 3) theoretical interpretability. The scree plot plots the eigenvalues on a coordinate plane. The number of components was determined based on the point where the slope approaches zero. A parallel analysis is a Monte Carlo simulation generating random eigenvalues. This was conducted using computer software by Watkins (2000). Components with an eigenvalue higher than the randomly generated eigenvalues were deemed to be true factors. Lastly, the items defining the components should form a coherent and theoretically relevant aspect of adherence to the MST model. The final solutions of both PCA's were compared and merged to arrive at one optimal solution.

Confirmatory Factor Analysis. A confirmatory factor analysis (CFA) was conducted on the second sample, consisting of 723 families using Mplus 7.4 (Muthén & Muthén, 1998-2015). For each family, one questionnaire was randomly selected to be included in the CFA. Nesting of families in therapists was taken into account with the COMPLEX module in Mplus.

Results

Principal Component Analysis. The first random subsample of TAM-Rs to be included in the PCA consisted of 580 questionnaires. Both the scree plot and the parallel analysis suggested a two-component solution. Theoretical interpretability of the two components was good.

The second random subsample of TAM-Rs consisted of 470 questionnaires. The scree plot suggested that two to three components might be the most meaningful. The parallel analysis suggested a three-component solution, although the third component just barely surpassed the third randomly generated eigenvalue. Theoretically, the third component did not seem to form a meaningful and coherent aspect of adherence.

Combining the two solutions suggested that a two-component solution was the most appropriate. The two components were readily interpretable and were labeled *therapist adherence* and *client-therapist alliance*, respectively. *Therapist adherence* referred to specific behaviors of the therapist to tackle specific problems. *Client-therapist alliance* referred to shared decision making and setting goals, as well as the bond between the client and the therapist. Table 1 presents the rotated component matrix of both two-component PCA's.

Several items loaded highly on both components and did not seem to differentiate well. To be able to clearly differentiate between both components for research purposes, items loading highly on both components were dropped. The following criteria were used for this (Floyd & Widaman, 1995): Items were identified as cross-loading when 1) loadings on both components were higher than .40, or 2) if the difference between the loadings on both components was no larger than .20, or 3) if the item loaded on both components on theoretical grounds. The following items were dropped as they loaded highly on both components according to criteria 1 or 2 in at least one of the subsamples: Item 7, 11, 13, 14, 15, 17, 22, and 12. Item 1, 6, and 18 were dropped according to the third criterion, as they loaded on alliance, but also clearly referred to one of the principles of MST, meaning that alliance and adherence were both represented in these items.

Table 1 Rotated component matrix of a Principal Component Analysis with varimax rotation

1	The therapist tried to understand how my family's problems all fit together.
2	**My family and the therapist worked together effectively.**
3	**My family knew exactly which problems we were working on.**
4	**The therapist recommended that family members do specific things to solve our problems.**
5	**The therapist's recommendations required family members to work on our problems almost every day.**
6	The therapist understood what is good about our family.
7	My family and the therapist had similar ideas about ways to solve problems.
8	**The therapist tried to change some ways that family members interact with each other.**
9	**The therapist tried to change some ways that family members interact with people outside the family.**
10	**My family and the therapist were honest and straightforward with each other.**
11	The therapist's recommendations should help the children to mature.
12	**Family members and the therapist agreed upon the goals of the session.**
13	My family talked with the therapist about how well we followed her/his recommendations from the previous session.
14	My family talked with the therapist about the success (or lack of success) of her/his recommendations from the previous session.
15	We got much accomplished during the therapy session.
16	**My family was sure about the direction of treatment.**
17	The therapist's recommendations made good use of our family's strengths.
18	My family accepted that part of the therapist's job is to help us change certain things about our family.
19	**The therapist's recommendations should help family members to become more responsible.**
20	**The therapist talked to family members in a way we could understand.**
21	**Our family agreed with the therapist about the goals of treatment.**
22	The therapist checked to see whether homework was completed from the last session.
23	The therapist did whatever it took to help our family with tough situations.
24	**The therapist helped us to enforce rules for the child.**
25	**The therapist helped family members talk with each other to solve problems.**
26	**The therapist helped us keep our child from hanging around with troublesome friends.**
27	**The therapist helped us improve our child's behavior at school.**
28	**The therapist helped us get our child to stay in school every day.**

Note. Items in bold were retained in the final factor solution

First random sample N = 580		Second random sample N = 470		
Adherence	*Alliance*	*Adherence*	*Alliance*	
0.25	0.68	0.15	0.70	Dropped, criterion 3
0.33	**0.64**	0.30	**0.78**	Alliance
0.20	**0.62**	0.24	**0.75**	Alliance
0.71	0.38	**0.70**	0.40	Adherence
0.67	0.30	**0.66**	0.36	Adherence
0.22	0.75	0.34	0.67	Dropped, criterion 3
0.35	0.60	0.44	0.64	Dropped, criterion 1 & 2
0.57	0.35	**0.57**	0.35	Adherence
0.65	0.10	**0.62**	0.12	Adherence
0.04	**0.77**	0.12	**0.60**	Alliance
0.58	0.39	0.56	0.49	Dropped, criterion 1 & 2
0.24	**0.77**	0.35	**0.74**	Alliance
0.60	0.40	0.67	0.39	Dropped, criterion 2
0.48	0.47	0.66	0.27	Dropped, criterion 1 & 2
0.73	0.35	0.57	0.48	Dropped, criterion 1 & 2
0.22	**0.69**	0.27	**0.65**	Alliance
0.55	0.53	0.58	0.47	Dropped, criterion 1 & 2
0.33	0.61	0.39	0.70	Dropped, criterion 3
0.71	0.22	**0.77**	0.29	Adherence
0.21	**0.71**	0.31	**0.75**	Alliance
0.36	**0.60**	0.27	**0.77**	Alliance
0.66	0.24	0.34	0.45	Dropped, criterion 2
0.51	0.40	0.50	0.57	Dropped, criterion 1 & 2
0.74	0.32	**0.73**	0.34	Adherence
0.68	0.32	**0.78**	0.19	Adherence
0.77	0.17	**0.76**	0.22	Adherence
0.81	0.13	**0.75**	0.23	Adherence
0.82	0.17	**0.68**	0.25	Adherence

The resulting *adherence* component consisted of 10 items ($M_{sample1}$ = 3.90, $SD_{sample1}$ = 1.57; $M_{sample2}$ = 4.08, $SD_{sample2}$ = 1.27). The internal consistency of the scale was excellent based on Cronbach's alpha ($\alpha_{sample1}$ = .92, $\alpha_{sample2}$ = .91). Inter-item correlations ranged from .38 to .84 in sample 1 and from .36 to .78 in sample 2. The resulting *alliance* component consisted of 7 items ($M_{sample1}$ = 4.51, $SD_{sample1}$ = 0.62; $M_{sample2}$ = 4.54, $SD_{sample2}$ = 0.54). The internal consistency of the scale was good based on Cronbach's alpha ($\alpha_{sample1}$ = .86, $\alpha_{sample2}$ = .86). Inter-item correlations ranged from .30 to .66 in sample 1 and from .38 to .61 in sample 2.

Confirmatory Factor Analysis. These components were tested in a confirmatory factor analysis in the second sample, consisting of 723 families. Model fit of the two-factor solution with an adherence-factor and an alliance-factor was acceptable (RMSEA = .06, CFI = .89, TLI = .87, SRMR = .06). All items loaded significantly on the corresponding factor and the correlation between both factors was .33. Modification indices suggested that the items 8 and 9 (M.I. = 31), and the items 27 and 28 (M.I. = 93) were highly correlated with one another. Indeed, these items were very similar in phrasing and content. Therefore, these items were allowed to correlate in a subsequent model. AIC and BIC improved substantially (Δ AIC = 240, Δ BIC = 231, Δ df = 2). Model fit of this modified model was good (RMSEA = .04, CFI = .95, TLI = .94, SRMR = .05).

Conclusion and Discussion

This study analyzed the structure of the Dutch TAM-R. We found that the TAM-R consisted of two underlying components, namely *therapist adherence* and *client-therapist alliance*. Cross-loading items were dropped to arrive at two components which could clearly differentiate between adherence and alliance. These components proved to have a high internal consistency and could be replicated in a CFA on an independent sample.

The current finding of an alliance-component within the MST adherence measure is not new. Previous studies have found the TAM (the adherence measure preceding the TAM-R) to consist of an alliance factor (Ellis et al., 2010; Henggeler et al., 2002), or to have a high correlation with other alliance-measures (Granic et al., 2012; Manders, Deković, Asscher, Van der Laan, & Prins, 2011). The current study corroborated these findings for the TAM-R. The presence of an alliance component corresponds to the MST treatment model which states that creating engagement and forming a strong alliance is an important aspect of MST (Henggeler et al., 2009).

The components obtained in the current study retained only 17 of the 28 items (61%) of the original TAM-R. This means that these components may not fully represent all aspects of adherence to the MST treatment model. Rather, they provide a basic understanding of two central elements of MST. Dropping cross-loading items ensured these two components to be clearly differentiated from one another, allowing them to be used for research purposes.

This may provide future studies with the opportunity to explore in more detail how the different aspects of the MST treatment model, namely adherence (specific behaviors to tackle specific problems) and alliance (agreement about goals and tasks, as well as the client-therapist bond) play a role within the MST treatment. We would, however, strongly urge MST therapists, supervisors, and consultants to continue to rely on the full TAM-R, consisting of 28 items, for feedback and quality assurance, as the complete TAM-R may provide a more comprehensive understanding of the delivery of the MST treatment model than the 17 items retained in the current study.

References

Asscher, J. J., Deković, M., Manders, W., van der Laan, P. H., Prins, P. J. M., van Arum, S., & Dutch MST Cost-Effectiveness Study Group (2014). Sustainability of the effects of multisystemic therapy for juvenile delinquents in the Netherlands: Effects on delinquency and recidivism. *Journal of Experimental Criminology, 10*, 227-243.

Deković, M., Asscher, J. J., Manders, W. A., Prins, P. J., & van der Laan, P. (2012). Within-intervention change: mediators of intervention effects during multisystemic therapy. *Journal of Consulting and Clinical Psychology, 80*, 574-587.

Ellis, M. L., Weiss, B., Han, S., & Gallop, R. (2010). The influence of parental factors on therapist adherence in multi-systemic therapy. *Journal of Abnormal Child Psychology, 38*, 857-868.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. Psychological Assessment, 7, 286-299.

Granic, I., Otten, R., Blokland, K., Solomon, T., Engels. R. C. M. E., & Ferguson, B. (2012). Maternal depression mediates the link between therapeutic alliance and improvements in adolescent externalizing behavior. *Journal of Family Psychology, 26*, 880-885.

Henggeler, S. (2011). Efficacy studies to large-scale transport: The development and validation of Multisystemic Therapy programs. *Annual Review of Clinical Psychology, 7*, 351-381.

Henggeler, S. W., Borduin, C. M., Schoenwald, S. K., Huey, S. J., & Chapman, J. E. (2006). *Multisystemic Therapy Adherence Scale-Revised (TAM-R)*. Unpublished instrument. Charleston, SC: Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina.

Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic Therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65*, 821-833.

Henggeler, S. W., & Schoenwald, S. K. (1999). The role of quality assurance in achieving outcomes in MST programs. *Journal of Juvenile Justice and Detention Services*, 14, 1-17.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (2009). Multisystemic therapy for antisocial behavior in children and adolescents. New York: Guilford.

Henggeler, S. W., Schoenwald, S. K., Liao, J. G., Letourneau, E. J., & Edwards, D. L. (2002). Transporting efficacious treatments to field settings: the link between supervisory practices and therapist fidelity in MST programs. Journal of Clinical Child and Adolescent Psychology, 31, 155-167.

Lange, A. M. C., van Busschbach, J., van Geffen, W., Timman, R., Scholte, R. H. J., & van der Rijken, R. E. A. (2015). The lack of cross-national equivalence of a therapist adherence measure (TAM-R) in Multisystemic Therapy (MST). *European Journal of Psychological Assessment, 32*, 312-325.

Letourneau, E. J., Sheidow, A. J., & Schoenwald, S. K. (2002). *Structure and reliability of the MST Therapist Adherence Measure scale in a large community sample* (Tech. Rep.). Charleston, SC: Medical University of South Carolina, Families Services Research Center.

Manders, W. A., Deković, M., Asscher, J. J., van der Laan, P. H., & Prins, P. J. M. (2011). De implementatie van Multisysteem Therapie in Nederland: De invloed van behandelintegriteit en nonspecifieke factoren op behandeluitkomsten [The implementation of Multisystemic Therapy in the Netherlands: The effect of adherence and nonspecific factors on treatment outcomes] *Gedragstherapie, 44*, 327-340.

Muthén, L. K., & Muthén, B. O. (1998-2015). Mplus user's guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Schoenwald, S. K. (2008). Toward evidence-based transport of evidence-based treatments: MST as an example. Journal of Child and Adolescent Substance Abuse, 17, 69-91.

Watkins, M. W. (2000). *Monte Carlo PCA for Parallel Analysis* [computer software]. State College, PA: Ed & Psych Associates.

Chapter 5

Alliance and adherence in a systematic therapy

Published as **Lange, A.M.C.**, van der Rijken, R.E.A., Delsing, M.J.M.H., Busschbach, J.J.V., van Horn, J.E., & Scholte, R.H.J. (2016). Alliance and adherence in a systemic therapy. *Child and Adolescent Mental Health*, *22*, 148-154.

Abstract

Background: The client-therapist working alliance is a key contributor to effective adult psychotherapy. However, little is known about its role in family and systemic therapy. Moreover, few studies have assessed alliance longitudinally or have investigated how it interrelates with other process variables, such as therapist adherence (i.e. the extent to which the therapist adheres to the treatment protocol or manual). We hypothesised that alliance and adherence interrelate over the course of the therapy.

Method: This study investigated the bidirectional associations between alliance and therapist adherence using cross-lagged panel analyses for a sample of 1,970 adolescents and their families participating in Multisystemic Therapy (MST). A number of client characteristics were included as moderators, namely demographic characteristics, type and severity of adolescent problem behaviour, and whether or not the MST treatment was court-ordered. Alliance and adherence were scored by the primary caregiver through telephone interviews at monthly intervals during treatment.

Results: Alliance in one month predicted therapist adherence in a subsequent month. Adherence only predicted subsequent alliance during the middle part of the treatment process. The results were not moderated by any of the client factors.

Conclusions: The results suggest that alliance and therapist adherence may reinforce one another during therapy. Whereas alliance may facilitate the development of therapist adherence, adherence may subsequently deepen and consolidate the client-therapist alliance. These results are independent of client characteristics.

Key Practitioner Message

- Working alliance between client and therapist, and adherence of the therapist to the treatment protocol both contribute to effective family and systemic therapy, but the bidirectional associations between them over time is unknown.
- Our results show that alliance and therapist adherence reinforce one another.
- It seems important to build a strong alliance at the start of therapy. Adherence to the therapy protocol helps to deepen and consolidate the working alliance.

Working alliance is a key contributor to effective psychotherapy and can be defined as the affective and collaborative aspects of the client-therapist relationship. It is usually conceptualised as personal alliance (the affective bond) and task-related alliance (addressing the goals of the treatment and the tasks required to achieve those goals) (Bordin, 1979; Hougaard, 1994). The association of a strong alliance with positive treatment outcomes is well established in individual adult psychotherapy (Martin, Garske, & Davis, 2000; Norcross & Wampold, 2011). Recent meta-analyses have suggested that alliance can also be important for effective family and systemic therapy (Friedlander et al., 2011; Karver et al., 2006). In family and systemic therapy a therapist often has to deal with multiple alliances (Robbins et al., 2003). The current study will focus on caregiver-reported alliance within a systemic therapy in which sessions primarily take place with the primary caregiver. Previous research regarding systemic therapy has suggested that parent-therapist alliance may be a better predictor of child outcomes than child- or adolescent-therapist alliance (Hogue et al., 2006; McLeod, 2011).

So far, the process through which alliance plays a role in therapy remains unknown for a number of reasons. Firstly, most studies have measured alliance on only one occasion, thus failing to take into account the longitudinal and developmental nature of alliance during therapy (Berkel et al., 2011; Crits-Christoph et al., 2011). Alliance may fluctuate over time and different developmental patterns of alliance may be associated with different treatment outcomes (Stiles et al., 2004). Secondly, few studies have assessed how alliance relates with other process variables, such as therapist adherence or client involvement (McLeod, 2011; McLeod et al., 2013). Yet, theoretical models posit that these process variables work together to initiate and facilitate therapeutic change (Goldfried & Davila, 2005; Hill, 2005; Karver et al., 2005).

Therapist adherence is the extent to which the therapist adheres to a treatment protocol or manual (McLeod et al., 2013; Perepletchikova & Kazdin, 2005). Therapist adherence is crucial in the dissemination and implementation of evidence-based interventions as it ensures that the key components of the intervention are being delivered as intended (Mihalic, 2004). Since evidence-based interventions have demonstrated their effectiveness in empirical studies, ensuring therapist adherence is a means to warrant continued positive treatment outcomes (Mihalic, 2004; Schoenwald, 2008).

Studies investigating the conjoint role of alliance and adherence have usually tested specific hypotheses regarding the mediating or moderating role of alliance on outcome, assessing alliance and adherence at a single point in time (e.g. Castonguay et al., 1996; The Multisite Violence Prevention Project, 2014; Tschuschke et al., 2015; Webb et al., 2012). Yet, theoretical models argue that alliance and therapist adherence are essential in each phase of therapy and are interrelated all along (Goldfried & Davila, 2005; Hill, 2005). Alliance and adherence both enable the therapist and the client to create engagement and confidence in

the therapy, to explore the problems and underlying causes, to accomplish the therapeutic tasks, and to facilitate successful termination of treatment. Alliance contributes to this through its affective and collaborative bond, which motivates and encourages the client. Adherence, on the other hand, may create confidence in the therapists' skills and provide the actual tools and techniques to foster therapeutic change. On top of their individual contribution, alliance and adherence may enhance one another: whereas a strong alliance may be a precondition for the adherent implementation of the intervention techniques, adherence may foster confidence in the therapist's skills and thereby deepen the client-therapist alliance (Goldfried & Davila, 2005; Hill, 2005).

Failing to take the developmental nature of alliance and adherence into account, by using only a single score for each construct, may prevent studies from discovering the true processes through which alliance and adherence jointly influence treatment outcomes. This may have led to contradictory findings in the past. Some studies did not find any association between alliance and adherence (The Multisite Violence Prevention Project, 2014), whereas others showed alliance to predict adherence (Tschuschke et al., 2015), or to mediate the association between adherence and outcome (Weck et al., 2015). In the absence of a strong alliance, a rigid focus on adherence may either lead to further deterioration of the alliance and interfere with therapeutic change (Barber et al., 2006; Castonguay et al., 1996), or may 'save' a treatment with low alliance, leading to positive treatment outcomes (Webb et al., 2012). Yet, as far as we know, studies that did investigate alliance and adherence longitudinally have not yet been able to support the bidirectional associations between alliance and adherence hypothesised in the theoretical models. Hukkelberg and Ogden (2013) did not find any significant associations between alliance and adherence in parent management training (PMTO, a family-focused method for children with externalising problem behaviour). Weck and colleagues (2015), who, besides the analyses on single alliance-scores discussed above, also conducted longitudinal analyses, showed that alliance during the first, but not the second, session predicted adherence in the following session. Adherence did not predict subsequent alliance. Of the studies described above, only two assessed family interventions (Hukkelberg & Ogden, 2013; The Multisite Violence Prevention Project, 2014). Both interventions worked primarily with the primary caregivers and targeted children and adolescents with emerging behavioural problems. These studies did not find any significant associations between alliance and adherence.

In our study, we hypothesised that alliance and therapist adherence would influence one another over the course of therapy, which is in accordance with the theoretical models of Goldfried and Davila (2005) and Hill (2005). For this purpose, we used routinely collected data from Multisystemic Therapy (MST), an evidence-based, and intensive home- and community based intervention for adolescents with antisocial and/or delinquent behavioural problems (12-18 years old; Henggeler et al., 2009). Sessions mainly take place

with caregivers, as, according to the MST theory of change, reductions in the adolescent externalising behavioural problems can be achieved through an increase of the parental sense of competence and the use of positive discipline (Deković et al., 2012). Within MST, caregiver-therapist alliance and therapist adherence are both related to reductions in antisocial and delinquent behavioural problems (Granic et al., 2012; Schoenwald, 2008; Schoenwald et al., 2009), yet their bidirectional association has not been assessed before.

Thus, this study aimed to investigate whether alliance in one month would influence therapist adherence in a subsequent month and vice versa. Therefore, we assessed both variables conjointly at five monthly intervals. We also tested these associations across subsamples of our client population to investigate whether the functioning of alliance and adherence within treatment would be stable across client characteristics (i.e. demographic characteristics, type of problem behaviour, severity of problem behaviour, and whether or not the family participated in MST on court-order; Barnhoorn et al., 2013).

Method

Participants

Adolescents. Families were referred to MST due to severe externalizing behavioural problems of the adolescent. Families had to meet the MST inclusion criteria, which have been specified by MST Services, the international licensor for the dissemination of MST (MST Services, 2014). A total of 2,393 MST trajectories started at one of the four participating treatment centres between July 2008 and January 2015. If a family started MST twice during the research period (N=11), only the first treatment episode was included for analyses, because inclusion of both treatment episodes would lead to dependency in the data. Another 412 families were excluded as they did not have any valid alliance or adherence assessments, resulting in a final sample of 1,970 clients (82% of the total sample).

The mean age of the 1,970 participating adolescents was 15 (SD = 1.41), 69% were male, 24% of the adolescents were of non-western origin, 47% lived in a single-parent household, and 57% participated in MST on a court-order. Most adolescents experienced externalising behavioural problems in the clinical range (75%) or the borderline range (9%) based on the Child Behavior Check List 6-18 (CBCL; Achenbach & Rescorla, 2001). Moreover, 56% of the adolescents also experienced internalising behavioural problems in the borderline or clinical range.

Therapists. MST was provided by 130 therapists working across 22 teams in four treatment centres. As part of the routine quality assurance and improvement system of MST, aimed at upholding adherence to the MST treatment model, all therapists followed an initial 5-day orientation training, participated in weekly supervision and expert consultation meetings, and attended quarterly booster sessions.

Measures

Alliance and Adherence. Client-therapist alliance and therapist adherence were measured using the Therapist Adherence Measure Revised (TAM-R; Henggeler et al., 2006; see also http://www.mstinstitute.org/qa_program/tam_languages.shtml). This questionnaire consists of 28 items rated on a 5-point Likert scale (1 'not at all', 2 'a little', 3 'some', 4 'pretty much', and 5 'very much'). On a monthly basis, employees from an independent call centre completed the TAM-R by interviewing the primary caregiver, which was most often the mother (82%), followed by the father (15%).

Although the TAM-R was originally developed to monitor therapist adherence to the MST model, several previous studies have found the questionnaire to also contain an alliance-factor (Ellis et al., 2010; Henggeler et al., 2002). A recent Dutch study confirmed two factors: 'client-therapist alliance' and 'therapist adherence' (Lange & van der Rijken, 2014). In the current study, only items clearly distinguishing between both factors were retained, dropping items loading on both factors. Reliability of the resulting two factors was good (Cronbach's $\alpha = .86$ for 'client-therapist alliance' and $\alpha = .91$ for 'therapist adherence'). 'Client-therapist alliance' consisted of seven items and measured the personal alliance (e.g. 'My family and the therapist were honest and straightforward with each other') as well as the task-related alliance (e.g. 'Our family agreed with the therapist about the goals of treatment'). 'Therapist adherence' consisted of ten items assessing therapist adherence to the MST clinical process and the treatment principles of MST (e.g. 'The therapist's recommendations required family members to work on our problems almost every day'). Three of these items targeted specific behavioural problems (e.g. 'The therapist helped us keep our child from hanging around with troublesome friends').

Only valid assessments (assessments by the primary caregiver, with a maximum of four missing items, and where face-to-face contact between the family and the therapist had occurred in the last 2 weeks prior to administration of the TAM-R; MST Institute, n.d.a; MST Institute, n.d.b) were included for analyses. Families provided on average 3.41 valid TAM-R administrations ($SD = 1.36$). Scores for alliance and therapist adherence could only be computed if all items on the specific factor had been scored.

Client Characteristics. The primary caregiver completed the Child Behavior Checklist for children aged 6 to 18 years (CBCL; Achenbach & Rescorla, 2001) to assess type and severity of the behavioural problems, and a questionnaire on demographic characteristics. Both questionnaires were completed on paper or online, depending on the routine practices of the treatment centre. These client characteristics were included as moderators in the analyses.

Procedures

All Dutch treatment centres collecting their alliance and therapist adherence scores through an independent call centre (N=4) were requested to share their data, which they had collected as part of their routine practices. All four centres agreed. Clients were informed that completing the questionnaires was part of the treatment and that the data could also be used for research purposes. The study was approved by the Committee Scientific Research Participation of the Vincent van Gogh Institute and complied to the American Psychological Association's ethical principles regarding research with human participants.

Strategy for Analysis

To investigate the bidirectional associations between alliance and adherence, we conducted cross-lagged panel analyses in Mplus 7.3 (Muthén & Muthén, 1998-2012). We specified a model with a fixed number of five time points (T1 through T5). Since alliance and adherence were collected on a monthly basis, our model included all assessments collected in the first five months of the MST treatment, because the length of an MST treatment should, in general, not exceed five months. Missing alliance or adherence scores were taken into account using a FIML estimator with robust standard errors, implemented as MLR in Mplus, to make use of all the available data and provide better estimations of standard errors when normality assumptions are violated.

The basic model (see Figure 1) included the initial covariance between alliance and adherence at T1, as well as the disturbance covariances between alliance and adherence at T2 to T5 (the latter are not shown for reasons of clarity). Furthermore, the model contained the stability paths between adjacent measurements, as well as the cross-lagged effects of alliance at one point in time on adherence at the next point in time and vice versa. We performed a series of multi-group analyses to test whether the observed cross-lagged associations were moderated by gender, age (based on median split: < 16, ≥ 16), ethnicity (western origin, non-western origin), type of household (single-parent or multiple-parent household), type of problem behaviour (no problems, externalising problems, comorbid externalising and internalising problems based on a T-score in the borderline or clinical range), severity of externalising problem behaviour (non-clinical, borderline, clinical T-score), and referral reason (court-ordered or not). The Satorra and Bentler (2001) scaled chi-squared difference test was used to compare the fit of the unconstrained model (no constraints on all covariances, stability paths, and cross-lagged paths) with a constrained model in which all covariances, stability paths, and cross-lagged paths were constrained to be equal across groups. The COMPLEX module implemented in Mplus was used to account for non-independence of observations due to cluster sampling (therapists treating more than one family). The goodness of fit of the models was assessed using the chi-square and

p values, the Comparative Fit Index (CFI: Bentler, 1990), and the Root Mean Square Error of Approximation (RMSEA: Steiger, 1990). CFI values above 0.90 indicate an acceptable fit, and values above 0.95 indicate an excellent fit to the data. RMSEA values below 0.08 suggest an acceptable fit, and values below 0.05 indicate a good fit (Hu & Bentler, 1999).

Results

Descriptive Statistics

Table 1 lists the correlations between the alliance and adherence scores across the five measurements, as well as the means and standard deviations. Paired samples *t*-tests revealed mean T1-alliance scores to be significantly ($p < .01$) lower than mean alliance scores at subsequent measurements, which did not differ significantly from one another. Mean adherence scores were found to increase significantly ($p < .01$) across all successive measurement intervals.

Table 1 Pearson correlations among alliance and adherence, and means and standard deviations

Measure	1	2	3	4	5	6	7	8	9	10
1. Alliance T1										
2. Alliance T2	.57									
3. Alliance T3	.51	.65								
4. Alliance T4	.41	.56	.67							
5. Alliance T5	.43	.54	.64	.74						
6. Adherence T1	.47	.24	.26	.19	.18					
7. Adherence T2	.37	.53	.39	.31	.30	.59				
8. Adherence T3	.28	.43	.58	.44	.38	.47	.68			
9. Adherence T4	.31	.40	.50	.63	.47	.44	.60	.75		
10. Adherence T5	.29	.39	.48	.55	.61	.39	.57	.70	.77	
Mean	4.51	4.58	4.61	4.63	4.66	3.36	3.86	4.03	4.11	4.20
SD	.56	.52	.50	.51	.49	1.06	.83	.77	.75	.71

Note. All correlations are significant at the 0.01 level (two-tailed)

Cross-Lagged Panel Models

Although for one of the multi-group analyses the difference in model fit between the constrained and unconstrained model reached significance, all of the modification indices for the paths of interest were small in the constrained model (M.I. < 10). This indicates that model fit would not improve much when allowing the parameters of interest to differ across groups. Therefore, we retained a single-group model. For reasons of parsimony, we constrained the cross-lagged paths from alliance to adherence to be equal across measurement intervals. This did not result in a significant deterioration of model fit ($p > .05$). However, when constraining the cross-lagged paths from adherence to alliance to be equal across measurement intervals, the model fit did significantly deteriorate ($p < .05$). Therefore we left these paths unconstrained in the final model. In a further attempt to specify the most parsimonious model, we constrained the disturbance covariances, as well as the stability paths of alliance and adherence to be equal across measurement points. These three actions also led to a significant deterioration in model fit, so, in the final model, these paths were also left unconstrained.

Figure 1 summarizes the standardized results of the final cross lagged model estimating the over-time associations between alliance and adherence. This model fitted the data well (χ^2 [27] = 270.40; CFI = .94; RMSEA = .07). The CFI and RMSEA values were close to the recommended cut-off values of .95 and .05 indicating good fit. Firstly, a significant positive association ($r = .47$, $p < .01$) was found at T1 between alliance and adherence, indicating that at this stage higher levels of alliance go together with higher levels of adherence. Secondly, stability coefficients of alliance and adherence were all significant ($p < .01$) and appeared to increase across subsequent measurement intervals, suggesting that alliance and adherence become increasingly stable over time. That is, as the intervention progresses, earlier relative levels of alliance and adherence become increasingly predictive of later relative levels of alliance and adherence, respectively. Stabilities of alliance and adherence appear to be about equally strong. Finally, and foremost, significant cross-lagged effects were found in both directions. Alliance had a positive stable effect on subsequent adherence across all measurement intervals (standardized beta = .11; $p < .001$). Adherence only had an effect on subsequent alliance from T2 to T3 and from T3 to T4 (standardized betas of .10, $p < .001$, and .06, $p < .01$), whereas it did not relate to subsequent alliance at the start and end of treatment (standardized betas ranging from .01 to .03, $p > .05$). These findings indicate that higher levels of alliance at one measurement point predicted an increase in adherence at the next measurement point (i.e. one month later). Similarly, higher levels of adherence predicted an increase in alliance one month later, but only during the middle part of the treatment process.

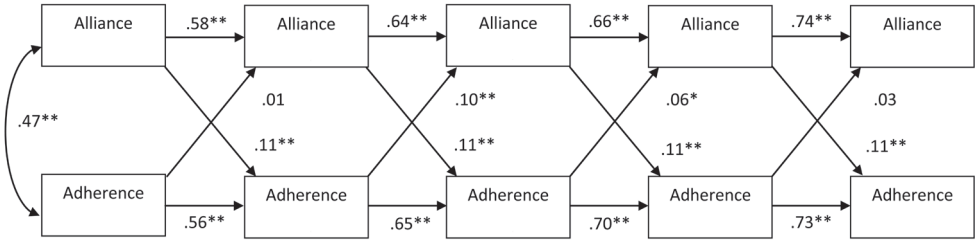

Fig. 1 Standardized coefficients for the associations between alliance and adherence from T1 to T5. *Note.* $*p < .01$; $**p < .001$

Discussion

The current study investigated the bidirectional associations of caregiver-therapist alliance and therapist adherence over time. We found that alliance in one month consistently predicted therapist adherence in a subsequent month. Therapist adherence only predicted subsequent alliance during the middle phase of treatment. These results were stable across a range of client characteristics, namely adolescent age, gender, and ethnicity, single-parent versus multiple-parent households, type and severity of adolescent problem behaviour, and whether or not the treatment was court-ordered. This means that the bidirectional associations between alliance and adherence were not moderated by client characteristics and can be generalised to the whole MST population.

The results suggest that alliance may function as a catalyst for adherence. Building a good working relationship and setting common goals may facilitate adherence to the treatment protocol. These findings are similar to findings from previous studies on individual adult psychotherapy showing that alliance predicted adherence (Tschuschke et al., 2015; Weck et al., 2015). Moreover, the results are in accordance with the MST treatment manual, according to which an MST therapist should start by creating engagement and a positive working relationship, and formulating common treatment goals. After this initial phase, the therapist can use specific interventions to address the identified problems (Henggeler et al., 2009).

Adherence predicted subsequent alliance, but only during the middle phase of treatment. Providing MST according to the treatment model may further deepen and consolidate the alliance between the therapist and the client. A client's confidence in the therapist may improve if a clear strategy is apparent. Also, if the therapist is delivering the treatment in an adherent manner, it may be easier to identify common goals and associated tasks (Goldfried & Davila, 2005; Hill, 2005). It is surprising that the effect of adherence on alliance only emerged halfway treatment, as adherence to the MST treatment model also requires focussing on topics such as client motivation and engagement, which are important at the

start of treatment. However, for the purposes of the current study, we only included items clearly distinguishing between alliance and adherence. The resulting adherence-measure mainly consisted of items reflecting problem-solving techniques, which are expected to be most apparent in the middle of therapy, where the therapist and client jointly work on the client's problems (Henggeler et al., 2009). Moreover, the current findings are comparable to the results of Hukkelberg and Ogden (2013) regarding parent management training (PMTO). They assessed alliance and adherence at the 3rd (T1), 12th (T2), and 20th (T3) session (with a mean number of 24 sessions). The effect of alliance on adherence was strongest at the start of therapy (from T1 to T2), whereas the effect from adherence on alliance was strongest halfway the treatment process (from T2 to T3). Although none of these associations were significant, the standardized effects were larger than corresponding associations in our study.

Our cross-lagged effects did not vary across client demographic characteristics nor across type and severity of adolescent behavioural problems. The conclusion that alliance and adherence may reinforce one another does therefore seem to hold for a varied MST population. Nevertheless, closer inspection of the data revealed that adherence scores did vary across client characteristics. For example, non-western families and families participating in MST on court-order provided higher adherence scores than western families or families participating in MST without a court-order. Adherence at the start of treatment was lower for adolescents with externalising behavioural problems in the clinical range than for adolescents without externalising behavioural problems. This is in line with previous research suggesting that problem severity may hamper adherent implementation of MST (Schoenwald, Letourneau, & Halliday-Boykins, 2005). Alliance did not appear to vary much across client characteristics, although this may be a consequence of the small variance of alliance in the current study.

Alliance only increased between T1 and T2, after which it stabilised. This is consistent with the MST treatment model, stating that alliance should be established in the initial phase, after which it should remain relatively stable (notwithstanding that this may require considerable work on the part of the therapist) (Henggeler et al., 2009). Previous research has shown that alliance may be characterized by short rupture-and-repair sequences (Stiles et al., 2004). As these ruptures can be repaired in just one or two sessions, identifying such ruptures would require session-to-session assessments, instead of our monthly assessments. It would be interesting to investigate whether such temporary ruptures of alliance would also impact therapist adherence.

Contrary to alliance, adherence increased during the whole treatment period. This may indicate that adherence becomes easier as treatment progresses. However, so far, little is known about such underlying processes. An alternative hypothesis for the increasing adherence scores may be that, as treatment progresses, parents develop a better

understanding of what the therapist is doing, and, therefore, are better able to identify adherent therapist behaviour. It is also possible that families become more positive about their therapist when positive treatment outcomes are being achieved, and hence give higher scores on the adherence items.

We should note that, although our results were significant, the effects were relatively small. It is likely that other factors, such as parental engagement or therapist experience, influence the development of alliance and therapist adherence during treatment as well. Besides, MST is a treatment with an elaborate quality assurance and improvement system, aimed at supporting therapists providing MST. This is reflected in the high mean scores on alliance and therapist adherence and the small standard deviation. With such restricted ranges, it may be harder to detect effects. Nevertheless, significant cross-effects were found. An important strength of the cross-lagged panel design employed is its control for the initial correlation between alliance and adherence, and for their stabilities over time. Given these controls, the size and consistency of the cross-lagged effects suggest meaningful relationships of sufficient strength to warrant attention.

Several caveats should be kept in mind when interpreting our results. Firstly, the TAM-R was developed to measure adherence to the MST treatment model, and was not designed as an alliance measure. In addition, both alliance and adherence were scored by the same informant, meaning that the results might have been inflated due to shared-method variance. Nevertheless, since the correlation between both scales at the start of the treatment was only medium, the factors can be assumed to measure two distinct processes. The TAM-R scales for alliance and adherence also had high internal consistencies and achieved high stability within an MST treatment episode. Thus, the reliability and validity of the two factors of the TAM-R seem adequate.

Secondly, some scholars have argued that caregivers may be less accurate adherence-informants than therapists or trained raters (Chapman et al., 2013). Being untrained in the treatment, caregivers may not be able to detect changes in adherence. Due to their loyalty towards their therapist, they may further be unwilling to rate the therapist poorly (Chapman et al., 2013; Schoenwald et al., 2011). Nevertheless, the TAM-R is a validated and reliable adherence-measure, and predicts a range of short- and long-term treatment outcomes (Huey et al., 2000; Schoenwald et al., 2009). The increasing adherence-scores in our study further suggest that families may be capable of detecting changes in adherence.

It would be interesting to replicate the current findings using other informants for alliance and adherence. Previous studies on client-therapist alliance and therapist adherence within MST suggest that adolescents tend to report somewhat lower levels of alliance and adherence than caregivers (Chapman et al., 2013; Ryan et al., 2013). Adolescent-therapist alliance may be more difficult to achieve and maintain, since these adolescents usually do not experience their life as problematic and may feel frustrated in their freedom

as parental interventions are being implemented (personal communication with a panel of clinicians). In a study on alliance in family therapy (multidimensional family therapy; MDFT), adolescent-reported alliance was associated with decreases in externalising behavioural problems, but only if the initial alliance was weak and subsequently improved (Hogue et al., 2006). It would be interesting to investigate whether such lower adolescent-therapist alliance is compensated for by an increase in adherence.

Although our study was restricted to MST, we have no reason to believe that the reinforcing patterns of alliance and therapist adherence would be different in other family therapies whereby sessions primarily take place with the caregiver. However, since the mean alliance and adherence scores in our sample were high, we do not know whether alliance and adherence would equally reinforce one another if one of them is very low. Indeed, previous research has suggested that a rigid focus on adherence in the absence of alliance may hamper treatment (Barber et al., 2006; Castonguay et al., 1996).Notwithstanding the caveats mentioned above, the present study represents one of the first attempts to model the bidirectional associations of alliance and therapist adherence over five assessments for almost 2,000 families. The large number of participating clients further allowed us to conduct moderator analyses, to investigate whether the results would differ for different subgroups of clients. Our findings were not dependent on client characteristics.

Conclusion

Taken together, we have put theoretical models on the associations of alliance and adherence to the test. Our results support the importance of building a strong alliance at the start of the treatment, as this may facilitate adherent implementation of intervention techniques. Adherence may be important to maintain a strong working alliance.

Acknowledgements

Funding for this work was provided by ZonMw (729101006). The authors thank the participating treatment centres (de Viersprong, de Waag, St. Ottho Gerhard Heldring & Ambulatorium, Vincent van Gogh Instituut), Praktikon, whose call centre collected the alliance and adherence-assessments, and the panel of clinicians, which assisted in interpretation of the results.
The authors declare that they have no competing or potential conflicts of interest.
The first and third authors take full responsibility for the integrity of the data and the accuracy of the data analyses.

References

Achenbach, T., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington, VT: University of Vermong, Research Center for Children, Youth & Families.

Barber, J. P., Gallop, R., Crits-Christoph, P., Frank, A., Thase, M. E., Weiss, R. D., & Gibbons, M. B. C. (2006). The role of therapist adherence, therapist competence, and alliance in predictingÐutcome of individual drug counseling: Results from the National Institute Drug Abuse Collaborative Cocaine Treatment Study. *Psychotherapy Research*, 16, 229-240.

Barnhoorn, J., Broeren, S., Distelbrink, M., de Greef, M., van Grieken, A., Jansen, W., et al. (2013). *Cliënt-, professional- en alliantiefactoren: Hun relatie met het effect van zorg voor jeugd* [Cliënt-, professional-, and alliancefactors: Their association with the effect of youth care]. Nijmegen, Rotterdam, Utrecht, the Netherlands: HAN, Erasmus MC, Gemeente Rotterdam, Verwey-Jonker Instituut.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238.

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science, 12*, 23-33.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research, and Practice, 16*, 252-260.

Castonguay, L. G., Goldfried, M. R., Wiser, S., Raue, P. J., & Hayes, A. M. (1996). Predicting the effect of cognitive therapy for depression: A study of unique and common factors. *Journal of Consulting and Clinical Psychology, 64*, 497-504.

Chapman, J. E., McCart, M. R., Letourneau, E. J., & Sheidow, A. J. (2013). Comparison of youth, caregiver, therapist, trained, and treatment expert raters of therapist adherence to a substance abuse treatment protocol. *Journal of Consulting and Clinical Psychology, 81*, 674-680.

Crits-Christoph, P., Gibbons, M. B., Hamilton, J., Ring-Kurtz, S., & Gallop, R. (2011). The dependability of alliance assessments: The alliance-outcome correlation is larger than you might think. *Journal of Consulting and Clinical Psychology, 79*, 267-278.

Deković, M., Asscher, J. J., Manders, W. A., Prins, P. J. M., & van der Laan, P. (2012). Within-intervention change: Mediators of intervention effects during Multisystemic Therapy. *Journal of Consulting and Clinical Psychology, 80*, 574-587.

Ellis, M. L., Weiss, B., Han, S., & Gallop, R. (2010). The influence of parental factors on therapist adherence in Multi-systemic therapy. *Journal of Abnormal Child Psychology, 38*, 857-868.

Friedlander, M. L., Escudero, V., Heatherington, L., & Diamond, G. M. (2011). Alliance in couple and family therapy. *Psychotherapy, 48*, 25-33.

Goldfried, M. R., & Davila, J. (2005). The role of relationship and technique in therapeutic change. *Psychotherapy: Theory, Research, Practice, Training, 42*, 421-430.

Granic, I., Otten, R., Blokland, K., Solomon, T., Engels, R. C., & Ferguson, B. (2012). Maternal depression mediates the link between therapeutic alliance and improvements in adolescent externalizing behavior. *Journal of Family Psychology, 26*, 880-885.

Henggeler, S. W., Borduin, C. M., Schoenwald, S. K., Huey, S. J., & Champan, J. E. (2006). *Multisystemic Therapy Adherence Scale-Revised (TAM-R).* Unpublished instrument. Charleston, SC: Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (2009). *Multisystemic therapy for antisocial behavior in children and adolescents.* New York: Guilford.

Henggeler, S. W., Schoenwald, S. K., Liao, J. G., Letourneau, E. J., & Edwards, D. L. (2002). Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs. *Journal of Clinical Child Psychology, 31*, 155-167.

Hill, C. E. (2005). Therapist techniques, client involvement, and the therapeutic relationship: Inextricably intertwined in the therapy process. *Psychotherapy: Theory, Research, Practice, Training, 42*, 431-442.

Hogue, A., Dauber, S., Stambaugh, L. F., & Cecero, J. J. (2006). Early therapeutic alliance and treatment outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology, 74*, 121-129.

Hougaard, E. (1994). The therapeutic alliance - A conceptual analysis. *Scandinavian Journal of Psychology, 35*, 67-85.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1-55.

Huey, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in Multisystemic Therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology, 68*, 451-467.

Hukkelberg, S. S., & Ogden, T. (2013). Working alliance and treatment fidelity as predictors of externalizing problem behaviors in parent management training. *Journal of Consulting and Clinical Psychology, 81*, 1010-1020.

Karver, M. S., Handelsman, J. B., Fields, S., & Bickman, L. (2005). A theoretical model of common process factors in youth and family therapy. *Mental Health Services Research, 7*, 35-51.

Karver, M. S., Handelsman, J. B., Fields, S., & Bickman, L. (2006). Meta-analysis of therapeutic relationship variables in youth and family therapy: The evidence for different relationship variables in the child and adolescent treatment outcome literature. *Clinical Psychology Review, 26*, 50-65.

Lange, A. M. C., & van der Rijken, R. E. A. (2014). *Structuur van de 'Therapist Adherence Measure-Revised' (TAM-R) binnen Multisysteem Therapie (MST)* [Structure of the 'Therapist Adherence Measure-Revised' (TAM-R) within Multisystemic Therapy (MST)]. Halsteren: the Viersprong Institute for Studies on Personality Disorders.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68*, 438-450.

McLeod, B. D. (2011). Relation of the alliance with outcomes in youth psychotherapy: A meta-analysis. *Clinical Psychology Review, 31*, 603-616.

McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodríguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health. *Clinical Psychology: Science and Practice, 20*, 14-32.

Mihalic, S. (2004). The importance of implementation fidelity. *Report on Emotional & Behavioral Disorders in Youth, 4*, 83-105.

MST Institute (n.d.a). *MST Therapist Adherence-Revised (TAM-R): Guidelines for administration*. Retrieved on 3rd of June 2016 from ://www.mstinstitute.org/qa_program/tam.shtml

MST Institute (n.d.b). *MST Therapist Adherence Measure – Revised (TAM-R): Guidelines for interpretation*. Retrieved on 3rd of June 2016 from ://www.mstinstitute.org/qa_program/tam.shtml

MST Services (2014). *Multisystemic Therapy® (MST®) Organizational Manual*. Charleston, SC: Author.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Norcross, J. C., & Wampold, B. E. (2011). Evidence-based therapy relationships: Research conclusions and clinical practices. *Psychotherapy, 48*, 98-102.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change. Issues and research recommendations. *Clinical Psychology: Science and Practice, 12*, 365-383.

Robbins, M. S., Turner, C. W., Alexander, J. F., & Perez, G. A. (2003). Alliance and dropout in family therapy for adolescents with behaviour problems: Individual and systemic effects. *Journal of Family Psychology, 17*, 534-544.

Ryan, S. R., Cunningham, P. B., Foster, S. L., Brennan, P. A., Brock, R. L., & Whitmore, E. (2013). Predictors of therapist adherence and emotional bond in Multisystemic Therapy: Testing ethnicity as a moderator. *Journal of Child and Family Studies, 22*, 122-136.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507-514.

Schoenwald, S. K. (2008). Toward evidence-based transport of evidence-based treatments: MST as an example. *Journal of Child and Adolescent Substance Abuse, 17*, 69-91.

Schoenwald, S. K., Chapman, J. E., Sheidow, A. J., & Carter, R. E. (2009). Long-term youth criminal outcomes in MST transport: The impact of therapist adherence and organizational climate and structure. *Journal of Clinical Child & Adolescent Psychology, 38*, 91-105.

Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research, 38*, 32-43.

Schoenwald, S. K., Letourneau, E. J., & Halliday-Boykins, C. (2005). Predicting therapist adherence to a transported family-based treatment for youth. *Journal of Clinical Child and Adolescent Psychology, 34*, 658-670.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behaviour Research, 25,* 173-180.

Stiles, W. B., Glick, M. J., Osatuke, K., Hardy, G. E., Shapiro, D. A., Agnew-Davies, R., et al. (2004). Patterns of alliance development and the rupture-repair hypothesis: Are productive relationships U-shaped or V-shaped. *Journal of Counselling Psychology, 51*, 81-92.

The Multisite Violence Prevention Project (2014). Implementation and process effects on prevention outcomes for middle school students. *Journal of Clinical Child & Adolescent Psychology, 43*, 473-485.

Tschuschke, V., Crameri, A., Koehler, M., Berglar, J., Muth, K., Staczan, P., et al. (2015). The role of therapists' treatment adherence, professional experience, therapeutic alliance, and clients' severity of psychological problems: Prediction of treatment outcome in eight different psychotherapy approaches. Preliminary results of a naturalistic study. *Psychotherapy Research, 25*, 420-434.

Webb, C. A., DeRubeis, R. J., Dimidjian, S., Hollon, S. D., Amsterdam, J. D., & Shelton, R. C. (2012). Predictors of patient cognitive therapy skills and symptom change in two randomized clinical trials: The role of therapist adherence and the therapeutic alliance. *Journal of Consulting and Clinical Psychology, 80*, 373-381.

Weck, F., Grikscheit, F., Jakob, M., Höfling, V., & Stangier, U. (2015). Treatment failure in cognitive-behavioural therapy: Therapeutic alliance as a precondition for an adherent and competent implementation of techniques. *British Journal of Clinical Psychology, 54*, 91-108.

Chapter 6

Development of alliance and therapist adherence in relation to treatment outcomes of adolescents with behavioral problems

Submitted as **Lange, A.M.C.**, van der Rijken, R.E.A., Busschbach, J.J.V., Delsing, M.J.M.H., & Scholte, R.H.J. Development of alliance and therapist adherence in relation to treatment outcomes of adolescents with behavioral problems. *Submitted*.

Abstract

Objective: The client-therapist working alliance and therapist adherence to the treatment manual are both associated with treatment outcomes. We investigated how alliance and adherence develop during treatment and how this development is uniquely and jointly related to treatment outcomes up to 18-months post-treatment using a variable-centered (latent growth curve modeling) and a person-centered approach (latent class growth analysis).

Method: We used routinely collected data from 848 adolescents (66% male and 76% western, mean age = 15.25 years) and their caregivers participating in Multisystemic Therapy (MST), a family- and community-based intervention for antisocial adolescents. Alliance and adherence were measured monthly through phone interviews with the caregivers using the Therapist Adherence Measure-Revised. Outcomes were assessed at the end of the treatment and at 18-months post-treatment using the scale 'rule-breaking behavior' of the Child Behavior Checklist and two MST Ultimate Outcomes (i.e., police contact and out-of-home placement).

Results: On average, alliance and adherence showed an increasing and then flattening slope. We identified two trajectory classes for alliance and three classes for adherence, which were mainly characterized by different initial levels of alliance and adherence, respectively. Both alliance and adherence predicted treatment outcomes at the end of treatment, but not at 18 months post-treatment. The effects of alliance could not be replicated using the person-centered approach.

Conclusions: We advocate the need to consider the dynamic nature of adherence in research as well as clinical practice. The lack of a robust alliance-outcome relationship may be a consequence of the small variance observed in alliance.

Understanding what processes cause favorable short- and long-term treatment outcomes is essential for the development and dissemination of treatments (Kazdin, 2005; Kazdin & Nock, 2003). Two often-studied processes are the working alliance between the client and the therapist, and therapist adherence. Alliance consists of personal alliance (the affective bond) and task-related alliance (addressing the goals of the treatment and the tasks required to achieve those goals; Bordin, 1979; Hougaard, 1994). Therapist adherence is the degree to which the therapist delivers the specified components of a specific intervention, i.e., the degree to which the delivery of the intervention is consistent with the proscribed procedures in the treatment protocol or manual (McLeod, Southam-Gerow, Tully, Rodríquez, & Smith, 2013; Perepletchikova & Kazdin, 2005).

A good alliance has been identified as a robust predictor of positive treatment outcomes in individual psychotherapy for adults (Martin, Garske, & Davis, 2000; Norcross & Wampold, 2011) and is increasingly being studied in other forms of psychotherapy. For instance, meta-analyses and systematic reviews suggest that alliance can also have positive effects on outcome in family and systemic therapy (De Greef, Pijnenburg, van Hattum, McLeod, & Scholte, 2016; Friedlander, Escudero, Heatherington, & Diamond, 2011; Karver, Handelsman, Fields, & Bickman, 2006). A good working relationship between the client and the therapist can improve engagement and motivation of the client and can provide a reliable and warm context in which the therapist can use his therapeutic skills to tackle the psychological problems of the client (Hill, 2005). Alliance is not a static characteristic of a client-therapist dyad, but may change and develop during treatment (Chu, Skriner, & Zandberg, 2014; Hudson et al., 2014; Kendall et al., 2009; Prince, Connors, Maisto, & Dearing, 2016; Stiles et al., 2004). Alliance is often reported to increase during the course of treatment (Chu et al., 2014; Kendall et al., 2009), although the speed of alliance increase may reduce as treatment progresses (Weiss, Kivity, & Huppert, 2014). Some even find a decrease of alliance during treatment (Hudson et al., 2014). In fact, there may exist different developmental trajectories of alliance for different client-therapist dyads (Stiles et al., 2004; Weiss et al., 2014). Also, alliance may be characterized by non-linear changes, such as ruptures, resulting in a temporary or lasting decrease of alliance (Escudero, Boogmans, Loots, & Friedlander, 2012; Stiles et al., 2004), or sudden gains, resulting in a lasting increase of alliance (Weiss et al., 2014).

Owen, Miller, Seidel, and Chow (2016) showed that the growth of alliance over time predicted more of the outcome variance than alliance measures that did not take the dynamic nature of alliance into account, such as a single-assessment of alliance, or an average alliance score aggregating all assessments of alliance during treatment. Several studies on the effects of alliance on outcomes of family and systemic therapy have shown that growth in alliance is a good predictor of treatment outcome (Hogue, Dauber, Stambaugh, Cecero, & Liddle 2006; Keeley, Geffken, Ricketts, McNamara, & Storch, 2011; Lerner, Mikami, & McLeod,

2011; Marker, Comer, Abrammova, & Kendall, 2013). Most of these analyses, however, were based on small samples, ranging from 25 to 86 participants.

In addition to alliance, therapist adherence may be related to treatment outcome. The key ingredients of an evidence-based intervention should be delivered as intended in order to achieve positive treatment outcomes (Mihalic, 2004). Studies have shown that failure to adequately implement an intervention can have detrimental effects on treatment outcomes (Durlak & Dupre, 2008; Sexton & Turner, 2010). Research findings on therapist adherence, however, are mixed. A frequently-cited meta-analysis on the relation between adherence and outcome focused on individual adult psychotherapy (Webb, DeRubeis, & Barber, 2010). The authors did not find a significant association between adherence and outcome, although they reported significant heterogeneity, indicating that adherence did relate to better outcomes in some studies, but not in others. A recent meta-analysis that focused on evidence-based interventions for juveniles with antisocial behavior found that studies with high treatment integrity (of which therapist adherence is an important aspect) yielded significantly larger effect sizes than studies with moderate or low treatment integrity, even after controlling for therapy duration and intervention type (Goense, Assink, Stams, Boendermaker, & Hoeve, 2016). Contrary to Webb and colleagues (2010), who only included studies that used expert raters to score adherence, Goense and colleagues (2016) also included ratings by caregivers or therapists.

In contrast to alliance, which has often been viewed as varying across each unique client-therapist dyad, therapist adherence was initially conceived as a static characteristic of the therapist, team, or site. As it was mainly studied in efficacy, effectiveness, or implementation studies, it was used as an indicator of the degree to which the intervention under study was well implemented (Goense, Boendermaker, van Yperen, Stams, & van Laar, 2014; Schoenwald, 2011). Yet, recent studies have proposed that a static representation of adherence as either absent or present may not reflect reality. In fact, therapist adherence may vary within a therapist: Adherence has been shown to increase as therapists acquire more experience (Lange, van der Rijken, Busschbach, Delsing, & Scholte, 2017) and may be harder to achieve when clients display higher levels of problem behavior (Boswell et al., 2013; Lebensohn-Chialvo, Hasler, Rohrbaugh, & Shoham, 2016). Adherence may even vary within a single treatment of a client (Robbins et al., 2011; Tschuschke et al., 2015). Nevertheless, adherence trajectories within a single treatment have scarcely been studied, especially in comparison to the number of studies on alliance trajectories. Chiapa and colleagues (2015) found adherence to deteriorate over the course of treatment, whereas Lange and colleagues (2016) found an increase in adherence during the whole treatment period. Robbins and colleagues (2011) decomposed adherence into several techniques and showed that some of these techniques were used more frequently as treatment progresses (such as techniques addressing the problem behaviors), whereas the use of other techniques decreased with time (such as techniques to create engagement).

The dynamic character of adherence has not often been acknowledged when studying its relationship with outcome. Studies that did take this into account showed that positive changes in adherence were related to better treatment outcomes (Chiapa et al., 2015; Robbins et al., 2011).

As alliance and therapist adherence both unfold during treatment, they are inevitably related to one another and may also interact in their association with treatment outcome. Theoretically, alliance and therapist adherence could enhance one another (Goldfried & Davila, 2005; Hill, 2005), and recent research has found support for this hypothesis (Lange et al., 2016). Findings regarding their relation to treatment outcome, however, are mixed. Barber and colleagues (2006) reported that a moderate level of adherence was associated with the best outcome when alliance was weak, but that the level of adherence was irrelevant when alliance was high. In contrast, Weck, Grikscheit, Jakob, Höfling, & Stangier (2015) found that the association between adherence and outcome was stronger when alliance was stronger.

Thus, there is some evidence that alliance and adherence may develop over time and that positive change in alliance and adherence may relate to better treatment outcomes. Moreover, alliance and therapist adherence may interact in their association with outcome, but findings so far are mixed. The current study, therefore, investigated the development of alliance and therapist adherence during treatment and how this development was uniquely and jointly associated with treatment outcomes up to 18 months post-treatment.

To answer these research questions, we used routinely collected data from Multisystemic Therapy (MST; Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 2009). MST is an evidence-based intervention that pays much attention to alliance and adherence and routinely collects data on these two aspects. MST focuses on addressing all environmental systems (home, school, friends, and neighborhood) that impact 12-18 year-old adolescents with antisocial and/or delinquent behavioral problems. It is an intensive family- and community-based intervention that works intensively with parents and caregivers to reduce behavioral problems of the adolescent (Henggeler et al., 2009). As treatment mainly takes place with the caregiver, alliance and adherence were scored by the primary caregiver. Previous research has shown parental alliance to be a significant predictor of child outcomes (Hogue et al., 2006; McLeod, 2011).

MST has an elaborate quality-assurance system to monitor and uphold adherent delivery of MST (Henggeler & Schoenwald, 1999; Schoenwald, 2008). This quality-assurance system centrally revolves around adherence of the therapist, as the therapist is the primary linkage with the family and, therefore, critical in achieving desired outcomes for youths and families (Henggeler & Schoenwald, 1999). There is a large body of evidence showing that higher therapist adherence to the MST treatment model is related to better treatment outcomes (see for example Huey, Henggeler, Brondino, & Pickrel, 2000; Schoenwald, Chapman, Sheidow, & Carter, 2009; Schoenwald, Sheidow, & Chapman, 2009). These studies, however, have all assessed adherence as a static factor.

To develop a more comprehensive of the joint and unique role of alliance and adherence within treatment with regard to outcomes, two different analytical approaches were used, namely a variable-centered approach and a person-centered approach to growth modeling. A variable-centered approach studies the associations between variables, and estimates an average for all participants on the variables under study. When modeling development over time (growth curve modeling), both the change over time as well as the initial level of alliance and adherence are estimated and can be used to predict treatment outcomes. As such, we can test whether the initial level of alliance and adherence, or rather change over time in alliance and adherence best predicts outcome. Moreover, it allows to control for the effect of alliance when estimating the effect of adherence on outcome, and the other way around. The person-centered approach assumes that there is not one average curve that best describes all participants but rather that participants can be categorized according to distinct growth trajectories (latent class growth analysis). These trajectory classes may differ regarding their treatment outcomes. This approach provides the opportunity to investigate the co occurrence of different developmental trajectories of alliance and adherence and how they jointly contribute to treatment outcome. We formulated the following hypotheses: **H1:** The mean trajectories for alliance and adherence are characterized by growth over time during treatment. **H2**: Different subgroups can be identified, characterized by different trajectories. We expected to find an increasing and a decreasing group regarding alliance, as well as an increasing and a decreasing group regarding adherence. **H3**: Positive changes in alliance and adherence during treatment are associated with better treatment outcomes at the end of treatment and 18-months post-treatment. No specific hypothesis was formulated regarding the joint role of alliance and adherence (i.e., how the combination of specific alliance and adherence trajectories relate to treatment outcome), as the findings hereof are mixed so far.

Method

Participants and Procedures

Procedures. All Dutch mental health care agencies that made use of an independent call center to collect the alliance and adherence data as well as the 18 months follow-up data were requested to participate. All three agencies agreed. Data collection was part of routine outcome monitoring and consisted of collecting data on client characteristics and the level of problem behaviors at the start of treatment and assessment of treatment outcomes at the end of treatment and 18 months post treatment. Alliance and adherence data were collected on a monthly basis during treatment. The call center is specialized in collecting data for MST. These data are only available at an aggregated level to therapists, thereby ensuring confidentiality. This is also emphasized in the communication towards the caregivers. Phone interviews take approximately ten minutes.

Clients were informed that all data would be used for quality control and research purposes. As the data collection was part of clinical practice and the data was provided anonymously to the researchers (i.e., it concerned retrospective file data), no informed consent was required. The study was approved by the Committee Scientific Research Participation of one of the participating mental health care agencies and complied to the American Psychological Association's ethical principles regarding research with human participants.

Families. Families were referred to MST due to severe externalizing behavioral problems of the adolescent. Families had to meet the MST inclusion criteria, which have been specified by MST Services, the international licensor for the dissemination of MST (MST Services, 2014). A total of 969 MST trajectories started at one of the three participating treatment centers between July 2008 and November 2012. If a family started MST twice during the research period (n=2), only the first treatment episode was included for analyses, because inclusion of both treatment episodes would lead to dependency in the data. Another 12 families were excluded as they did not have any valid alliance or adherence assessments (see description of TAM-R for a definition of 'valid'). Finally, 17 families were excluded because MST was terminated for a reason which was not related to the treatment itself (such as a family moving away from the treatment area of the MST team), and 90 families were excluded because it was not known whether or not MST was terminated for a treatment-related reason. This resulted in a final sample of 848 clients (88% of the total sample; see Table 1 for the client characteristics of the clients included in, and excluded from, the final sample).

The response rate was > 90% at the end of treatment, except for rule-breaking behavior (caregiver-report), were the response rate was 50%. The response rate was lower for the latter as this was not an obligatory outcome measure of the MST quality assurance system. Nevertheless, 18% of these non-responding families did participate in the 18 months follow-up interview. The group of non-responding families at end of treatment reported lower alliance on T2 and T5 and lower adherence on T5, they showed higher drop-out rates and out-of-home placements and consisted of more families with a non-western origin (see Table 1).

At 18-months post-treatment, the response rate was 46%. The most important reasons for missing data at 18-months were: The call center was unable to get in touch with the family (27%), the caregivers refused to participate (12%), and the caregivers were unable to participate due to a language barrier (3%). For 10% of the families the reason for not completing the follow-up interview had not been noted by the call center. Non-responding families had similar or even better treatment outcomes at the end of treatment than responding families, and they reported comparable levels of alliance and adherence (see Table 1 for a full comparison of the groups with and without missing data). In total, 99.9% of our sample reported on at least one of the three outcome measures at the end of the treatment or at follow-up, and all families reported on at least one occasion on alliance or adherence.

Table 1. Characteristics of the samples included and excluded from the study and of the samples with and without missing data

	Included study	Excluded study	χ^2 (df)	18-months FU present
N	*848*	*119*		*390*
PC start: yes	**53%**	**65%**	**4.01 (1)***	53%
Adolescent gender: male	66%	65%	0.02 (1)	67%
Single-parent family: yes	52%	59%	1.06 (1)	56%
Origin: non-western	24%	25%	0.02 (1)	**17%**
Treatment court-ordered: yes	66%	58%	2.11 (1)	**61%**
Treatment-related drop-out: yes	**8%**	**78%**	**F****	**5%**
OOH end: yes	4%	12%	F	4%
PC end: yes	25%	38%	1.89 (1)	24%
	M (SD)	*M (SD)*	*t (df)*	*M (SD)*
Age adolescent	15.25 (1.37)	15.20 (1.49)	0.36 (884)	15.27 (1.34)
RB start	**69.46 (8.94)**	**72.36 (7.47)**	**-2.68 (828)****	**70.47 (8.66)**
RB end	64.37 (8.90)	66.10 (9.12)	-1.04(496)	**65.20 (8.56)**
Alliance T1	4.49 (0.55)	n/a		4.51 (0.51)
Alliance T2	4.57 (0.51)	n/a		4.58 (0.46)
Alliance T3	4.59 (0.50)	n/a		4.61 (0.47)
Alliance T4	4.60 (0.53)	n/a		4.64 (0.44)
Alliance T5	4.65 (0.50)	n/a		4.67 (0.44)
Adherence T1	3.31 (1.05)	n/a		3.22 (1.05)
Adherence T2	3.83 (0.84)	n/a		3.83 (0.79)
Adherence T3	3.95 (0.80)	n/a		3.98 (0.74)
Adherence T4	4.08 (0.78)	n/a		4.12 (0.72)
Adherence T5	4.14 (0.73)	n/a		4.18 (0.67)

Note. Fisher's Exact test was used when the assumptions for chi-square were not met. As this test does not provide a test-value, a F is denoted in the Table in these situations. For rule-breaking behavior T-scores are presented. Significant findings are in bold. RB = rule-breaking behavior; PC = police contact; OOH = out-of-home placement.
* $p < .05$; ** $p < .01$

18-months FU absent	χ^2 (df)	RB at end present	RB at end absent	χ^2 (df)
458		*422*	*426*	
54%	0.07 (1)	51%	56%	1.61 (1)
65%	0.25 (1)	65%	67%	0.46 (1)
49%	2.99 (1)	52%	52%	0.00 (1)
30%	**17.21 (1)****	**17%**	**31%**	**18.01 (1)****
71%	**9.56 (1)****	63%	70%	6.63 (1)
11%	**7.47 (1)****	**2%**	**14%**	**34.84 (1)****
4%	0.02 (1)	2%	7%	5.46 (1)*
26%	0.22 (1)	23%	27%	1.14 (1)
M (SD)	*t (df)*	*M (SD)*	*M (SD)*	*t (df)*
15.24 (1.40)	0.28 (802)	15.22 (1.37)	15.29 (1.37)	-0.77 (802)
68.71 (9.26)	**2.56 (680)***	69.66 (8.70)	69.42 (9.43)	0.35 (680)
63.34 (9.43)	**2.10 (368.89)***	n/a	n/a	n/a
4.48 (0.59)	0.70 (579.40)	4.51 (0.54)	4.47 (0.57)	0.85 (578.92)
4.56 (0.56)	0.52 (620.93)	**4.61 (0.46)**	**4.53 (0.56)**	**2.11 (598.06)***
4.57 (0.53)	0.91 (612.94)	4.60 (0.48)	4.57 (0.53)	0.73 (559.74)
4.57 (0.61)	1.56 (513.30)	4.63 (0.50)	4.56 (0.57)	1.66 (500.99)
4.62 (0.57)	1.28 (441.34)	**4.69 (0.44)**	**4.59 (0.57)**	**2.01 (386.81)***
3.39 (1.04)	-1.96 (559)	3.25 (1.06)	3.37 (1.03)	-1.42 (559)
3.83 (0.88)	-0.03 (630)	3.89 (0.78)	3.76 (0.89)	1.87 (598.14)
3.93 (0.84)	0.72 (597.96)	3.97 (0.77)	3.93 (0.83)	0.63 (603)
4.05 (0.83)	1.04 (564)	4.11 (0.76)	4.05 (0.80)	0.93 (564)
4.09 (0.80)	1.31 (455.02)	**4.21 (0.66)**	**4.05 (0.82)**	**2.32 (393.42)***

Therapists. MST was provided by 63 therapists working across 13 teams in three treatment centers. The mean number of families treated per therapist was 13, varying from 1 to 37. Therapists all completed higher education in a relevant domain to qualify for MST therapist. After higher education, they all completed the 5-day MST training, participated in weekly supervision and expert consultation meetings, and attended quarterly booster sessions (Henggeler et al., 2009). The training and supervision are part of the routine quality assurance and improvement system of MST, aimed at upholding adherence to the MST treatment model.

Measures

Alliance and Adherence. Client-therapist alliance and therapist adherence were reported by the primary caregiver (80% mothers, 17% fathers) using the Therapist Adherence Measure Revised (TAM-R; Henggeler, Borduin, Schoenwald, Huey, & Chapman, 2006; see also https://msti.org/mstinstitute/qa_program/tam_languages.html). This questionnaire consists of 28 items rated on a 5-point Likert scale (1 'not at all', 2 'a little', 3 'some', 4 'pretty much', and 5 'very much').

Although the TAM-R was originally developed to monitor therapist adherence to the MST model, several studies have shown TAM-R scores to correlate moderately to strongly with scores on alliance measures (Granic et al., 2012; Manders, Deković, Asscher, van der Laan, & Prins, 2011). Others suggested that the questionnaire actually contains an alliance-component (Ellis, Weiss, Han, & Gallop, 2010; Henggeler, Schoenwald, Liao, Letourneau, & Edwards, 2002). As the current study used the revised TAM-R, consisting of 28 items instead of the older TAM, consisting of 26 items (of which 19 overlap with the TAM-R), we used the components extracted from the TAM-R in a recent Dutch study (Lange & van der Rijken, 2014). Using a principal component analysis with varimax rotation, this study extracted two components based on data from 580 families reporting on the TAM-R: 'client-therapist alliance' and 'therapist adherence' (Lange & van der Rijken, 2014). For research purposes, only items clearly distinguishing between both components were retained. The resulting alliance component consisted of 7 items (Cronbach's $\alpha = .86$), whereas adherence consisted of 10 items ($\alpha = .91$). Model fit for these components was found to be good in a subsequent confirmatory factor analysis on an independent sample consisting of 723 families.

'Client-therapist alliance' measured the personal alliance as well as the task-related alliance. Example items are "My family and the therapist were honest and straightforward with each other" and "Our family agreed with the therapist about the goals of treatment". 'Therapist adherence' assessed therapist adherence to the MST clinical process and the treatment principles of MST. For example, "The therapist's recommendations required family members to work on our problems almost every day" and "The therapist's recommendations should help family members to become more responsible".

Only valid assessments (assessments by the primary caregiver, with a maximum of four missing items, and where face-to-face contact between the family and the therapist had occurred in the last 2 weeks prior to administration of the TAM-R; MST Institute, n.d.a; MST Institute, n.d.b) were included for analyses. Families provided on average 3.63 valid TAM-R administrations (*SD* = 1.26). Scores for alliance and therapist adherence were only computed if all items on the specific factor had been completed.

Outcome Measures. Rule-breaking behavior was scored by the primary caregiver using a subscale of the Child Behavior Checklist 6-18 (CBCL; Achenbach & Rescorla, 2001; Dutch version by Verhulst & Van der Ende, 2001). Assessments took place at the start and end of the treatment and at 18 months post-treatment. This scale consists of 17 items with 3-point Likert-scales (ranging from 0 = never, to 2 = often). Cronbach's alpha was .85 according to the manual. For the current study, raw scores were transformed to T-scores, with a higher T-score indicating more problems.

Beside rule-breaking behavior, two other outcome measures were assessed at the same time points: 1) police contact: Whether or not the adolescent had been in contact with the police during the past six months (excluding contact with the police as a victim, but including contact as a suspect or witness), 2) out-of-home placement: Whether or not the youth was currently placed out of home (e.g., a stable home situation with grandparents or foster parents was scored as 'living at home'). If the adolescent was placed out of home during treatment, MST was stopped, leading to a negative treatment outcome for this MST episode. These outcomes are being used by MST as ultimate outcomes and have been operationalized and standardized by MST Services to ensure that these outcomes are being scored in the same way (MST Institute, 2016). The ultimate outcomes were scored by the therapist (after consultation with the caregivers) at the start and end of the treatment, and by the primary caregiver at 18-months post-treatment.

Strategy for Analysis

All analyses were conducted in Mplus 7.4 (Muthén & Muthén, 1998-2015). Missing data was handled using robust full maximum likelihood estimation (MLR). This approach makes use of all the available data and provides better estimations of standard errors when normality assumptions are violated. We accounted for the non-independence of the data due to therapists treating more than one family by adjusting the standard-errors using the COMPLEX module in Mplus. We used both a variable-centered approach and a person-centered approach to evaluate how alliance and adherence develop during treatment and how this development is related to treatment outcome. Both approaches consisted of three steps: 1) fitting overall or group-based trajectory growth models for alliance and adherence, 2) adding treatment outcomes to the best-fitting models of alliance and adherence, and 3) adding treatment outcomes to a joint model of alliance and adherence.

Variable-Centered Approach. As a first step, we specified separate growth models for alliance and adherence during the treatment (**hypothesis 1**), containing a fixed number of five time points (T1 through T5). Since alliance and adherence were collected on a monthly basis, our model included all assessments collected in the first five months of the MST treatment. The length of an MST treatment should, in general, not exceed five months. Both models included an intercept and a linear slope. Subsequently, a quadratic slope was added and compared to the more restrictive model. We used the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for model comparison. Rules of thumb suggest that an increase between four and seven points for each additional parameter on the AIC and an increase between two and six points for each additional parameter on the BIC may be positive evidence for the alternative model (Burnham & Anderson, 2004; Raftery, 1999).
The second step was to investigate the predictive value of alliance and adherence for treatment outcome (**hypothesis 3**). For this purpose, the outcome variables were added to the best fitting growth curve models of alliance and adherence. Separate models were specified for each outcome variable and for each time-point (i.e., rule-breaking behavior, police contact, and out-of-home placement at the end of the treatment and at follow-up). This resulted in twelve models (2 process variables x 3 outcome variables x 2 post-treatment time-points). We fitted linear regression models for the continuous outcomes (rule-breaking behavior) and logistic regression models for the binary outcomes (police contact and out-of-home placement). In each model, we controlled for the outcome as assessed at the start of the treatment, except for the outcome measure out-of-home placement, since all adolescents had to be living at home to start the treatment.

As alliance and adherence are expected to co vary, we also investigated the unique contribution of alliance and adherence on outcome by controlling both constructs for one another. To achieve this, we combined the univariate growth models of alliance and adherence in a bivariate model and used the growth factors to predict treatment outcome in the same way as was done in the separate alliance and adherence models, leading to six models (3 outcome measures x 2 time points). We included the covariances of all alliance and adherence growth factors as well as the covariances between the error terms of alliance and adherence within each time point.

Person-Centered Approach. First, latent class growth analyses (LCGA) were conducted to identify latent trajectory classes of alliance and adherence (Muthén & Muthén, 2000; Nagin & Trambley, 1999) (**hypothesis 2**). LCGA assumes homogenous growth within classes, meaning that the variances of the growth parameters (i.e., intercept and slope factors) were restricted to zero. The growth parameters to be included in the LCGA was based on the final growth models identified in the variable-centered approach. We evaluated models with up to four different classes. Five criteria were used to decide on the number of classes. The BIC and the Lo-Mendell-Rubin ratio likelihood test (LMR-LRT; Lo, Mendell, & Rubin, 2001) were

used as an index of model fit. A p-value < .05 for the LMR–LRT indicates that the k trajectory model is a better fit to the data compared to the k – 1 trajectory model. The entropy and the average latent class probabilities provided an index of classification accuracy, with values closer to 1 indicating greater precision (range 0-1). Lastly, classes had to be substantially different from one another.

Second, families were assigned to their most likely trajectory class based on their highest posterior probability and these class assignments were used to predict treatment outcome (**hypothesis 3**). As with the variable-centered approach, we used linear and logistic regression models and controlled for the baseline assessment if appropriate.

Lastly, the frequencies of the different combinations of the trajectory classes of alliance and adherence were calculated (Nagin & Tremblay, 2001). These joint classes of trajectories were subsequently used to predict outcome in a similar manner as was the case for the separate trajectory classes (no specific hypothesis formulated).

Results

Descriptive Statistics

Table 2 lists the correlations between the alliance and adherence scores across the five measurements and the outcome measures. Means and standard deviations are also provided.

Variable-Centered Approach: Latent Growth Curve Analysis

Development of Alliance and Adherence (H1). To investigate the development of alliance and adherence, growth models with an intercept and linear slope were fitted. Subsequently, a quadratic slope was added and model fit was compared. For alliance, the AIC and BIC values suggested that the model with a quadratic slope showed the best model fit (Δ AIC = 98, Δ BIC = 79, Δ df = 4). Model fit with a quadratic slope was excellent (CFI: 1.00, TLI: 1.01, RMSEA: 0.00, SRMR: 0.02, $\chi 2$ = 2.92, p = .82). On average, alliance showed a relatively high intercept (M (s.e.) = 4.51 (0.02), p < .001) and a slight increase during treatment (M (s.e.) = 0.04 (0.02), p < .05). The mean of the quadratic slope was not significant (M (s.e.) = -0.003 (0.004), p = .54). The variances of the intercept, linear and quadratic slope were all significant, indicating individual differences in the growth parameters.

For adherence, the model with a quadratic slope also showed the best model fit (Δ AIC = 98, Δ BIC = 79, Δ df = 4). Model fit with a quadratic slope was good (CFI: 0.96, TLI: 0.94, RMSEA: 0.08, SRMR: 0.05, χ^2 = 37.20, p < .001). On average, adherence had a somewhat lower intercept than alliance (M (s.e.) = 3.40 (0.05), p < .001), increased during treatment (M (s.e.) = 0.39 (0.03), p < .001), and had a negative quadratic slope, pointing to a flattening slope (M

(s.e.) = -0.05 (0.01), $p < .001$). The variances of the intercept, linear and quadratic slope were also significant, indicating individual differences. The mean growth curves for alliance and adherence are plotted in Figure 1.

Prediction of Treatment Outcome by Alliance and Adherence (H3). We subsequently analyzed whether the growth processes of alliance and adherence predicted treatment outcome. Due to collinearity between the quadratic and linear slopes of alliance (r = -.86) and adherence (r = -.87), we used growth models consisting of only an intercept and linear slope to predict outcome. As the variance of the quadratic slopes in alliance and adherence was very small (i.e., < .01), exclusion of this growth parameter was deemed appropriate.

For alliance, a steeper increase in the slope of alliance was associated with less rule-breaking behavior and lower odds of police contact at the end of the treatment (see Table 3). Alliance did not predict the odds of out-of-home placement, nor any of the outcomes at 18 months post-treatment. When controlling for the growth process of adherence in the bivariate growth curve model, alliance no longer predicted any of the treatment outcomes.

For adherence, a higher intercept was associated with lower odds of police contact and out-of-home placements at the end of treatment (see Table 4). These effects were still present when controlling for the growth process of alliance in the bivariate growth curve model (police contact: B (s.e.) = -0.56 (0.28), 95% CI [-1.12, -0.01], $p < .05$; out-of-home placement: B (s.e.) = -1.01 (0.46), 95% CI [-1.92, -0.10], $p < .05$). A steeper increase in the slope of adherence was associated with lower odds of out-of-home placements at the end of treatment. This effect was no longer significant when controlling for the growth process of alliance. Adherence did not predict rule-breaking behavior nor any of the outcomes at 18 months post-treatment.

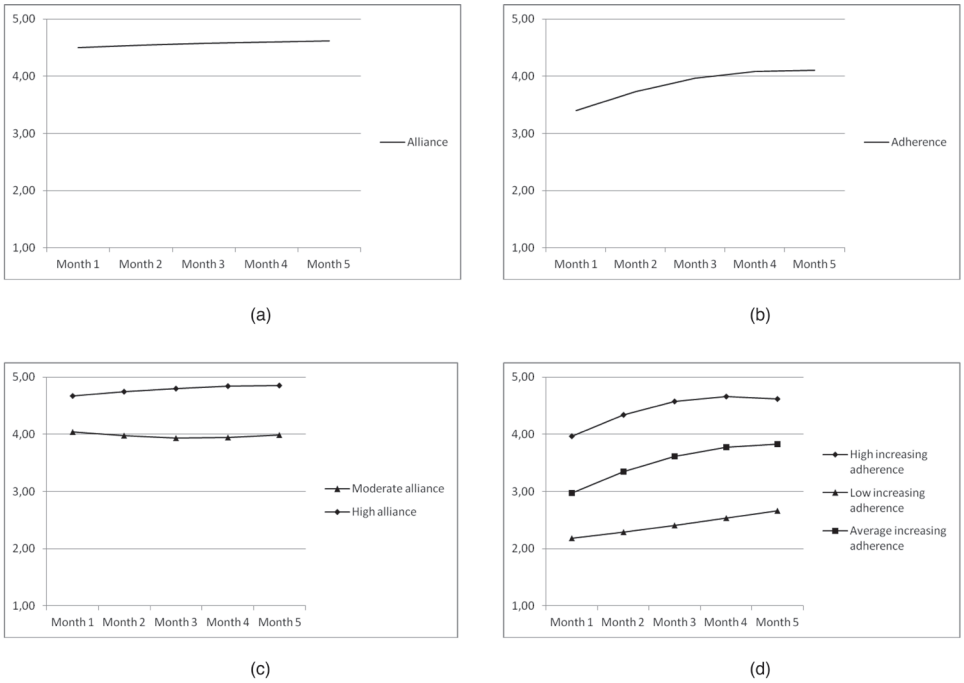

Fig. 1 a) Growth curve for alliance, **b)** growth curve for adherence, **c)** latent classes for alliance, **d)** latent classes for adherence

Table 2. Correlation and descriptive statistics for alliance, adherence, and treatment outcome

	1	2	3	4	5	6	7
1. All T1	-						
2. All T2	.53**	-					
3. All T3	.48**	.68**	-				
4. All T4	.43**	.56**	.74**	-			
5. All T5	.43**	.51**	.63**	.75**	-		
6. Adh T1	**.46****	.23**	.25**	.20**	.19**	-	
7. Adh T2	.35**	**.51****	.41**	.33**	.27**	.62**	-
8. Adh T3	.19**	.40**	**.57****	.46**	.36**	.46**	.70**
9. Adh T4	.34**	.37**	.57**	**.63****	.47**	.43**	.63**
10. Adh T5	.32**	.35**	.51**	.52**	.61**	.42**	.56**
11. RB start	-.03	-.07	-.01	-.01	-.03	-.13**	-.11*
12. RB end	-.07	-.11	-.03	-.11*	-.17**	-.17**	-.13*
13. RB 18m	.03	.03	-.00	-.03	-.06	-.12	-.07
14. PC start	.03	.06	.05	.04	.05	.09*	.06
15. PC end	-.01	.00	-.04	-.07	-.05	-.00	-.07
16. PC 18m	-.01	.10	.05	.03	-.05	.01	.10
17. OOH end	.02	-.03	-.03	-.03	-.03	.03	-.04
18. OOH 18m	.01	-.03	-.01	-.02	-.09	.06	-.00
Mean (*SD*)	4.49 (0.55)	4.57 (0.51)	4.59 (0.50)	4.60 (0.53)	4.65 (0.50)	3.31 (1.05)	3.83 (0.84)
% of zero (no PC / no OOH)	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Note: For all bivariate associations Pearson's correlation was calculated. Variables assessed at the same point in time are bold. All = alliance; Adh = adherence; RB = rule-breaking behavior; PC = police contact; OOH = out-of-home placement.
* $p < .05$. ** $p < .01$

8	9	10	11	12	13	14	15	16	17	18
-										
.79**	-									
.70**	.78**	-								
-.07	-.01	-.07	-							
-.11*	-.07	-.18**	.62**	-						
-.04	-.07	-.07	.33**	.40**	-					
.09*	.10*	.10*	**.11****	.08	-.03	-				
-.07	-.06	-.05	.13**	**.32****	.09	.22**	-			
.06	.07	.00	.07	.11	**.36****	.05	.07	-		
-.05	-.08	-.06	.09*	**.28****	.03	.02	.22**	-.05	-	
.04	.04	-.03	.03	.12	**.26****	.05	.09	.11*	.21**	-
3.95 (0.80)	4.08 (0.78)	4.14 (0.73)	69.55 (9.02)	64.40 (8.99)	62.33 (8.25)	n/a	n/a	n/a	n/a	n/a
n/a	n/a	n/a	n/a	n/a	n/a	47%	75%	82%	96%	90%

Table 3. Regression of treatment outcome on growth parameters of alliance

	Estimate (s.e.)	95% Confidence interval	Standardized effect	Estimate (s.e.)	95% Confidence interval	Standardized effect
	Intercept Alliance			*Slope Alliance*		
Rule-breaking behavior end of treatment	-0.74 (1.01)	[-2.71, 1.24]	-0.03	**-14.50 (6.51)***	**[-27.26, -1.74]**	**-0.16**
Rule-breaking behavior 18 months post-treatment	0.21 (1.37)	[-2.48, 2.90]	0.01	-14.65 (8.18)	[-30.69, 1.39]	-0.18
Police contact end of treatment: yes	-0.19 (0.22)	[-0.62, 0.24]	-0.04	**-2.83 (1.41)***	**[-5.60, -0.07]**	**-0.15**
Police contact 18 months post-treatment: yes	0.50 (0.43)	[-0.35, 1.35]	0.12	-1.36 (2.71)	[-6.68, 3.96]	-0.07
Placed out-of-home end of treatment: yes	-0.30 (0.40)	[-1.07, 0.48]	-0.07	-2.82 (2.03)	[-6.79, 1.16]	-0.15
Placed out-of-home 18 months post-treatment: yes	-0.09 (0.54)	[-1.14, 0.96]	-0.02	-4.29 (5.89)	[-15.83, 7.25]	-0.18

Note: Parameter estimates for rule-breaking behavior are linear regression coefficients. All other outcomes are logistic regression coefficients. Standardized effects are standardized with respect to both predictor and outcome. Results in bold are those results that are significant after application of the correction factor.
*$p < .05$; **$p < .01$

Person-Centered Approach: Latent Class Growth Analysis

Latent Trajectories of Alliance and Adherence (H2). For alliance a two-class model was chosen as this was the only model that showed a significantly better model fit than the k - 1 model ($p < .05$). The BIC values also pointed to the two-class solution (Δ1class - 2class BIC = 1083, Δ2class - 3class BIC = 337). Entropy was good (.85), the average latent class probabilities were close to 1 (namely .92 and .97), and the latent trajectories were substantially meaningful. The classes in the final model can be described as a 'moderate stable alliance' group (26%, n = 220) and a 'high stable alliance' group (74%, n = 626) (see Figure 1).

For adherence a three-class model was chosen. The LMR-LRT regarding the comparison between the three- and two-class models was significant ($p < .05$), entropy was good (.79), the average latent class probabilities were good (.88 and .91), and the latent trajectories were substantially meaningful. Although these values were comparable in the two-class solution (entropy = .76, latent class probabilities = .92 to .94) and BIC values substantially decreased in all models ($\Delta_{1class - 2class}$ BIC = 866, $\Delta_{2class - 3class}$ BIC = 337, $\Delta_{3class - 4class}$ BIC = 159), we

Table 4. Regression of treatment outcome on growth parameters of adherence

	Estimate (s.e.)	95% Confidence interval	Standardized effect	Estimate (s.e.)	95% Confidence interval	Standardized effect
	Intercept Adherence			*Slope Adherence*		
Rule-breaking behavior end of treatment	-1.01 (0.52)	[-2.02, 0.01]	-0.09	-7.02 (4.28)	[-15.41, 1.38]	-0.13
Rule-breaking behavior 18 months post-treatment	-1.11 (0.76)	[-2.61, 0.39]	-0.10	-5.38 (4.80)	[-15.23, 3.57]	-0.12
Police contact end of treatment: yes	**-0.43 (0.15)****	**[-0.72, -0.14]**	**-0.17**	-2.03 (1.18)	[-4.34, 0.28]	-0.17
Police contact 18 months post-treatment: yes	0.32 (0.24)	[-0.15, 0.80]	0.14	0.38 (1.61)	[-2.78, 3.54]	0.03
Placed out-of-home end of treatment: yes	**-0.48 (0.23)***	**[-0.92, -0.03]**	**-0.19**	**-3.79 (1.74)***	**[-7.20, -0.39]**	**-0.32**
Placed out-of-home 18 months post-treatment: yes	0.13 (0.24)	[-0.35, 0.61]	0.06	-1.18 (2.22)	[-5.53, 3.18]	-0.10

Note: Parameter estimates for rule-breaking behavior are linear regression coefficients. All other outcomes are logistic regression coefficients. Standardized effects are standardized with respect to both predictor and outcome. Results in bold are significant.
$^{*}p < .05$; $^{**}p < .01$

chose the three-class solution as model fit was significantly better than the two-class model, and because entropy was slightly higher. As the latent trajectory classes were subsequently used as predictors of treatment outcome, accurate classification was deemed essential. The classes in the final model could be described as 'high increasing adherence' (48%, $n = 400$), 'average increasing adherence' (45%, $n = 375$), and 'low increasing adherence' (7%, $n = 66$) (see Figure 1).

Prediction of Treatment Outcome by Latent Trajectories (H3). Table 5 presents the associations between trajectory class and outcome. The trajectory classes of alliance did not predict any of the treatment outcomes. The trajectory classes of adherence predicted rule-breaking behavior and police contact at the end of the treatment. Families in the high increasing group reported less rule-breaking behavior at the end of the treatment than families in the average increasing group. Families in the high increasing group also reported less police contact at the end of the treatment than the families in the low increasing group.

Joint Trajectories and Outcome. Almost half of all families (44%, n = 371) reported high increasing levels of adherence in combination with high levels of alliance. The second largest group of families reported average increasing levels of adherence in combination

Table 5. Regression of treatment outcome on latent classes

	Estimate (s.e.)	95% C.I.	Standardized effect
		Alliance groups (1 – moderate alliance; 2 – high alliance)	
RB end	-1.81 (0.92)	[-3.63, -0.00]	-0.20
RB 18m	-1.16 (0.91)	[-2.95, 0.62]	-0.14
PC end: yes	-0.12 (0.19)	[-0.49, 0.26]	-0.06
PC 18m: yes	0.03 (0.31)	[-0.57, 0.63]	0.02
OOH end: yes	-0.17 (0.47)	[-1.09, 0.76]	-0.09
OOH 18m: yes	-0.51 (0.40)	[-1.30, 0.28]	-0.28

Note: Parameter estimates for rule-breaking behavior are linear regression coefficients. All other outcomes are logistic regression coefficients. Standardized effects are standardized with respect to outcome only as the predictors are binary. RB = rule-breaking behavior; PC = police contact; OOH = out-of-home placement.

*$p < .05$; **$p < .01$

with high alliance (27%, n = 225). Families reporting moderate alliance most frequently reported average increasing levels of adherence (19%, n = 156). Only few people reported high increasing adherence in combination with moderate alliance (3%, n = 27), low increasing adherence in combination with moderate levels of alliance (4%, n = 35), and low increasing adherence in combination with high alliance (3%, n = 25).

These joint trajectory classes were subsequently used to predict outcome. Classes representing less than 5% of the families were not included as predictor in the regression analyses, resulting in three groups: 'high alliance with high increasing adherence' (reference group), 'moderate alliance with average increasing adherence', and 'high alliance with average increasing adherence'. Families experiencing moderate levels of alliance in combination with average increasing adherence reported more rule-breaking behavior at the end of the treatment than families experiencing high levels of alliance in combination with high increasing adherence (B (s.e.) = 2.15 (0.96), 95% CI [0.27, 4.03], $p < .05$). No other differences were observed between the groups.

Discussion

The current study investigated how alliance and therapist adherence develop during treatment and how this development is related to treatment outcomes at the end of treatment and at 18 months post-treatment, using two different analytical approaches. As

Estimate (s.e.)	95% C.I.	Standardized effect	Estimate (s.e.)	95% C.I.	Standardized effect
Dummy adherence 1 (0 – high increasing; 1 – low increasing)			*Dummy adherence 2* (0 – high increasing; 1 – average increasing)		
1.47 (1.10)	[-0.68, 3.61]	0.16	**1.70 (0.61)****	**[0.49, 2.90]**	**0.19**
2.79 (1.83)	[-0.80, 6.38]	0.34	0.27 (1.01)	[-1.70, 2.25]	0.03
0.72 (0.34)*	**[0.05, 1.38]**	**0.38**	0.37 (0.20)	[-0.02, 0.75]	0.19
-0.18 (0.45)	[-1.07, 0.71]	-0.10	-0.47 (0.31)	[-1.07, 0.13]	-0.26
0.24 (0.80)	[-1.32, 1.81]	0.13	0.63 (0.41)	[-0.17, 1.43]	0.34
-0.87 (1.09)	[-3.01, 1.27]	-0.48	0.23 (0.35)	[-0.46, 0.91]	0.12

hypothesized in H1, we found that both alliance and therapist adherence increased during treatment, although the increase of alliance was minimal as it was high throughout the treatment. Accordingly, when estimating trajectory classes, only two classes were identified for alliance, characterized by high stable and moderate stable alliance. Three trajectory classes could be distinguished for adherence, which were characterized by low increasing, average increasing, and high increasing adherence. This is contrary to our hypothesis (H2) that we would also be able to distinguish a decreasing group for alliance and adherence. Both alliance and adherence predicted some, but not all of the outcome measures at the end of the treatment, and none of the 18-months outcomes. Thus, our third hypothesis (H3) was only partially confirmed. The effects of adherence on outcome seemed more robust than the effects of alliance, as the effects of alliance disappeared after controlling for adherence and could not be reproduced using the person-centered approach.

The developmental patterns of alliance and therapist adherence were both characterized as increasing during treatment. Alliance was very high from the start of the treatment. As the first alliance assessment usually takes place in the second half of the first month of treatment, it appears that, within MST, alliance is generally well established after the first few weeks, and may continue to improve somewhat as treatment progresses. This is consistent with the MST treatment model, which posits that therapists should establish a strong working alliance during the initial phase of the therapy, and should maintain this throughout the treatment (Henggeler et al., 2009). Early adherence levels, on the other hand, were more

moderate and increased more pronouncedly throughout treatment. Previous studies, both within MST as well as within adult psychotherapy, have shown that alliance may function as a catalyst for therapist adherence. Establishing a good working relationship and setting common goals may facilitate implementing the treatment model adherently (Lange et al., 2016; Weck et al., 2015). This may explain why adherence is more moderate at the start of the treatment and improves as treatment progresses. It is surprising that we did not find a group of families reporting decreasing levels of alliance or adherence. We suspect that these families were present in our sample but that this group may have been too small to be extracted in a distinct trajectory class.

Both alliance and adherence predicted treatment outcomes. For alliance, increasing levels predicted lower levels of rule-breaking behavior and decreased the odds of police contact at the end of the treatment, but did not predict any of the treatment outcomes at 18 months post-treatment. It is interesting to note that change over time, instead of the initial level of alliance, was most predictive of outcome, which is in accordance with previous findings (e.g., Owen et al., 2016). Yet, these results did not hold after controlling for adherence, nor when using the person-centered approach. Previous meta-analyses have suggested that alliance has a small to medium effect on outcome in family and systemic therapy (Friedlander et al., 2011; Karver et al., 2006), and qualitative research within MST also supports the central role of alliance in fostering and sustaining change (Kaur, Pote, Fox, & Paradisopoulos, 2015; Paradisopoulos, Pote, Fox, & Kaur, 2015). There may have been too little variance in the current sample to find more robust associations. Nevertheless, a recent systematic review on the parent-therapist alliance reported that almost 20% of the included studies found non-significant associations only (De Greef et al., 2016). The majority of the studies reported a combination of positive significant and non-significant associations. This is contrary to the meta-analysis on adult psychotherapy which reported a consistent association between alliance and outcome across studies (Martin et al., 2000). This may indicate that the alliance-outcome relationship is more complex within family and systemic therapy than within individual adult psychotherapy, as several family members are participating in treatment. In such a context, the individual-therapist alliance may be of less importance. Rather, alliances of different family members with the therapist may need to be 'in balance' and goals may need to be shared within the whole family instead of only within the client-therapist dyad (Escudero, Friedlander, Varela, & Abascal, 2008; Friedlander et al., 2011; Robbins, Turner, Alexander, & Perez, 2003). Alliance measures which specifically target these systemic aspects of alliance, such as the System for Observing Family Therapy Alliances (SOFTA; Friedlander et al., 2006), may be better suited to study the role of alliance in family and systemic therapy.

For adherence, higher initial levels of therapist adherence and increasing levels of adherence during treatment were both related to better treatment outcomes at the end

of the treatment. This is in line with several previous studies on the association between adherence to the MST-model and outcome (Huey, Henggeler, Brondino, & Pickrel, 2000; Lange et al., 2017). Yet, the positive association between adherence and outcome in general is not undisputed, as was shown in the meta-analyses by Webb and colleagues (2010). One of the reasons could be that a too strict application of a treatment model may actually predict poorer treatment outcomes (Barber et al., 2006; Durlak & Dupre, 2008). The MST quality-assurance system may be well suited to prevent such rigidity as model adherence is discussed weekly during supervision. This can assist therapists in discerning the key ingredients and can help them applying these ingredients appropriately in each unique family situation.

It was surprising that adherence did not predict any of the long-term treatment outcomes. MST outcomes have been reported to sustain over time and therapist adherence within MST has been found to predict long-term lower rates of criminal charges as well (Schoenwald, Chapman, Sheidow, & Carter, 2009).

This study provided little evidence on how alliance and adherence jointly relate to outcome. We did find that a strong alliance in combination with high adherence related to less rule-breaking behavior at the end of treatment than moderate levels of alliance and adherence, yet such an association was not found for the other two outcomes. The small number of participants with low adherence, and the absence of a group of clients with low alliance, further prevented us from analyzing the role of alliance and adherence when one or the other is very low.

Several caveats should be kept in mind. Firstly, our study was restricted to MST, an intervention that has an intensive quality-assurance system to continuously monitor and sustain alliance and adherence to the MST model (Henggeler & Schoenwald, 1999; Schoenwald, 2008). As such, there may have been less variation in the current sample than may be the case in other settings or other interventions. On the one hand, this may have contributed to the limited number of significant associations between alliance and outcome, as well as the lack of findings regarding the joint role of alliance and adherence. On the other hand, our results may not apply to situations where alliance or adherence are extremely low. It is salient to understand whether alliance and adherence can compensate for one another in such a situation or whether they are both needed for good outcomes. Future research should realise that evidence-based interventions with intensive quality assurance may produce too little variance in processes and outcome, even when observed in clinical practice.

Secondly, we had to deal with fair amounts of missing data at different points in time. These high percentages of missing data are common in routine outcome monitoring (e.g., De Beurs et al., 2011; Hoenders et al., 2014). Although high attrition is undesirable, it is advised to include all individuals with any outcome data (Wood, White, & Thompson, 2004).

Indeed, analytical approaches for handling missing data, such as maximum likelihood, have been found to perform well even in samples with 50% missing information (Graham, 2009). In our sample, only 12% of families lost at follow-up refused to participate. Most families could not be reached due to incorrect contact details or language barriers. Although the group of families lost to follow-up had more frequently dropped out of treatment, they did not report poorer treatment outcomes at the end of the treatment. The non-response on rule-breaking behavior at the end of treatment may represent a more selective group of clients. On average, non-responding families seem to have experienced less successful treatments, as alliance and adherence scores were lower, a larger portion of the families dropped out of treatment, and a larger portion reported having had their adolescent being placed out-of-home. Nevertheless, one would not expect the association between alliance / adherence and outcome to be different for these non-responders than for the responding families. In addition, the overall conclusions of our study would not change if we would only focus on the two outcome variables with high response rates. We, therefore, felt safe to use these data, even though we must keep these details in mind when interpreting the results.

Thirdly, therapist adherence was rated by caregivers. Caregiver responses may, however, be biased. For example, ratings may be influenced by caregivers' loyalty towards the therapist, or their satisfaction with treatment. The TAM-R is assessed anonymously to counter biases due to loyalty (therapists only receive feedback on their adherence levels in an aggregated form) and caregivers are asked to rate the presence of concrete behaviors to minimize the influence of satisfaction on the scores. Previous research further has shown the TAM-R scores to predict reductions in criminal charges up to four years posttreatment (Schoenwald, Chapman et al., 2009). Nevertheless, previous research in the Netherlands has indeed shown the Dutch TAM to have a moderate association with satisfaction (r = .40; Manders et al., 2011). We must, therefore, keep in mind that the assessments of alliance and adherence in this study may have been influenced by factors such as satisfaction.

Conclusions

The current study is novel in that it measured the development of adherence over time. Researchers, managers, and professionals need to be aware of the fact that adherence is a dynamic process and that they may need to attend to this, for example when designing a study or when implementing and delivering an intervention. Attending to the level of adherence within a treatment episode may aid in achieving behavioral change in therapy.

Alliance did predict treatment outcomes at the end of treatment but the results seemed less robust than the results for adherence. Yet, most professionals would underscore the relevance of developing a good working relationship with their clients. Standard alliance measures, which have been found to perform well in adult psychotherapy, may not

necessarily be appropriate for family and systemic therapy, and other instruments should be considered.

Acknowledgements

Funding for this work was provided by ZonMw (729101006). The authors thank the participating treatment centers (de Viersprong, St. Ottho Gerhard Heldring and Ambulatorium, Vincent van Gogh Instituut) for making their data accessible, Praktikon for collecting the alliance and adherence assessments through phone interviews, Femke Kuijpers for reading the manuscript, and the panel of clinicians, for their clinical view of the results.

References

Achenbach, T., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth & Families.

Barber, J. P., Gallop, R., Crits-Christoph, P., Frank, A., Thase, M. E., Weiss, R. D., & Gibbons, M. (2006). The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: Results from the national institute drug abuse collaborative cocaine treatment study. *Psychotherapy Research, 16,* 229-/240.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research, and Practice, 16*, 252-260.

Boswell, J. F., Gallagher, M. W., Sauer-Zavala, S. E., Bullis, J., Gorman, J. M., Shear, M. K., Woods, S., & Barlow, D. H. (2013). Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder. *Journal of Consulting and Clinical Psychology*, *81*, 443-454.

Burnham, K. P., & Anderson, D. R. (2004). Multimodal inference. Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*, 261-304.

Chiapa, A., Smith, J. D., Kim, H., Dishion, T. J., Shaw, D. S., & Wilson, M. N. (2015). The trajectory of fidelity in a multiyear trial of the family check-up predicts change in child problem behavior. *Journal of Consulting and Clinical Psychology, 83*, 1006-1011.

Chu, B. C., Skriner, L. C., & Zandberg, L. J. (2014). Trajectory and predictors of alliance in cognitive behavioral therapy for youth anxiety. *Journal of Clinical Child & Adolescent Psychology, 43*, 721-734.

De Beurs, E., den Hollander-Gijsman, M. E., van Rood, Y. R., van der Wee, N. J. A., Giltay, E. J., van Noorden, M. S., van der Lem, R., van Fenema, E., & Zitman, F. G. (2011). Routine outcome monitoring in the Netherlands: Practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clinical Psychology and Psychotherapy, 18*, 1-12.

De Greef, M., Pijnenburg, H. M., van Hattum, M., McLeod, B. D., & Scholte, R. H. J. (2016). Parent-professional alliance and outcomes of child, parent, and family treatment: A systematic review. *Journal of Child and Family Studies*. doi: 10.1007/s10826-016-0620-5

Durlak, J. A., & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327-350.

Ellis, M. L., Weiss, B., Han, S., & Gallop, R. (2010). The influence of parental factors on therapist adherence in Multi-systemic therapy. *Journal of Abnormal Child Psychology, 38*, 857-868.

Escudero, V., Boogmans, E., Loots, G., & Fiedlander, M. L. (2012). Alliance rupture and repair in conjoint family therapy: An exploratory study. *Psychotherapy, 49*, 26-37.

Escudero, V., Friedlander, M., Varela, N., & Abascal, A. (2008). Observing the therapeutic alliance in family therapy: Associations with participants' perceptions and therapeutic outcomes. *Journal of Family Therapy, 30*, 194-214.

Friedlander, M. L., Escudero, V., Heatherington, L., & Diamond, G. M. (2011). Alliance in couple and family therapy. *Psychotherapy, 48*, 25-33.

Friedlander, M. L., Escudero, V., Horvath, A. O., Heatherington, L., Cabero, A., & Martens, M. P. (2006). System for observing family therapy alliances: A tool for research and practice. *Journal of Counseling Psychology, 53*, 214-224.

Goense, P. B., Assink, M., Stams, G. J., Boendermaker, L., & Hoeve, M. (2016). Making 'what works' work: A meta-analytic study of the effect of treatment integrity on outcomes of evidence-based interventions for juveniles with antisocial behavior. *Aggression and Violent Behavior.* doi: 10.1016/j.avb.2016.08.003

Goense, P., Boendermaker, L., van Yperen, T., Stams, G., & van Laar, J. (2014). Implementation of treatment integrity procedures. An analysis of outcome studies of youth interventinos targeting externalizing behavioral problems. *Zeitschrift für Psychologie, 222,* 12–21.

Goldfried, M. R., & Davila, J. (2005). The role of relationship and technique in therapeutic change. *Psychotherapy: Theory, Research, Practice, Training, 42*, 421–430.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Reviews of Psychology, 60*, 549-576.

Granic, I., Otten, R., Blokland, K., Solomon, T., Engels. R. C. M. E., & Ferguson, B. (2012). Maternal depression mediates the link between therapeutic alliance and improvements in adolescent externalizing behavior. *Journal of Family Psychology, 26*, 880-885.

Henggeler, S. W., Borduin, C. M., Schoenwald, S. K., Huey, S. J., & Chapman, J. E. (2006). *Multisystemic Therapy Adherence Scale-Revised (TAM-R).* Unpublished instrument. Charleston, SC: Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina.

Henggeler, S. W., & Schoenwald, S. K. (1999). The role of quality assurance in achieving outcomes in MST programs. *Journal of Juvenile Justice and Detention Services, 14(2)*, 1-17.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (2009). *Multisystemic therapy for antisocial behavior in children and adolescents.* New York: Guilford.

Henggeler, S. W., Schoenwald, S. K., Liao, J. G., Letourneau, E. J., & Edwards, D. L. (2002). Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs. *Journal of Clinical Child Psychology, 31*, 155-167.

Hill, C. E. (2005). Therapist techniques, client involvement, and the therapeutic relationship: Inextricably intertwined in the therapy process. *Psychotherapy: Theory, Research, Practice, Training, 42*, 431-442.

Hoenders, R. H. J., Bos, E. H., Bartels-Velthuis, A. A., Vollbehr, N. K., van der Ploeg, K., de Jonger, P., de Jong, J. T. V. M. (2014). Pitfalls in the assessment, analysis, and interpretation of routine outcome monitoring (ROM) data: Results from an outpatient clinic for integrative mental health. *Administration and Policy in Mental Health, 41*, 647-659.

Hogue, A., Dauber, S., Stambaugh, L. F., Cecero, J. J., & Liddle, H. A. (2006). Early therapeutic alliance and treatment outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology, 74*, 121-129.

Hougaard, E. (1994). The therapeutic alliance - A conceptual analysis. *Scandinavian Journal of Psychology, 35*, 67-85.

Hudson. J. L., Kendall, P. C., Chu, B. C., Gosch, E., Martin, E., Taylor, A., & Knight, A. (2014). Child involvement, alliance, and therapist flexibility: Process variables in cognitive-behavioral therapy for anxiety disorders in childhood. *Behavior Research and Therapy, 52*, 1-8.

Huey, S., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology, 68*, 451-467.

Karver, M. S., Handelsman, J. B., Fields, S., & Bickman, L. (2006). Meta-analysis of therapeutic relationship variables in youth and family therapy: The evidence for different relationship variables in the child and adolescent treatment outcome literature. *Clinical Psychology Review, 26*, 50-65.

Kaur, P., Pote, H., Fox, S., & Paradisopoulos, D. A. (2015). Sustaining change following multisystemic therapy : Caregiver's perspectives. *Journal of Family Therapy.* doi:10.1111/1467-6427.12093

Kazdin, A. E. (2005). Treatment outcomes, common factors, and continued neglect of mechanisms of change. *Clinical Psychology: Science and Practice, 12,* 184-188.

Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, 44*, 1116-1129.

Keeley, M. L., Geffken, G. R., Ricketts, E., McNamara, J. P. H., & Storch, E. A. (2011). The therapeutic alliance in the cognitive behavioral treatment of pediatric obsessive-compulsive disorder. *Journal of Anxiety Disorders, 25*, 855-863.

Kendall, P. C., Comer, J. S., Marker, C. D., Creed, T. A., Puliafico, A., Hughes, A. A., Martin, E. D., Suveg, C., & Hudson, J. (2009). In-session exposure tasks and therapeutic alliance across the treatment of childhood anxiety disorders. *Journal of Consulting and Clinical Psychology, 77*, 517-525.

Lange, A. M. C., van der Rijken, R. E. A., Busschbach, J. J. V., Delsing, M. J. M. H., & Scholte, R. H. J. (2017). It's not just the therapist: Therapist and country-wide experience predict therapist adherence and adolescent outcome. *Child and Youth Care Forum*. doi:10.1007/s10566-016-9388-4.

Lange, A. M. C., & van der Rijken, R. E. A. (2014). *Structuur van de 'Therapist Adherence Measure-Revised' (TAM-R) binnen Multisysteem Therapie (MST)* [Structure of the 'Therapist Adherence Measure-Revised' (TAM-R) within Multisystemic Therapy (MST)]. Halsteren, the Netherlands: Viersprong Institute for Studies on Personality Disorders.

Lange, A. M. C., van der Rijken, R. E. A., Delsing, M. J. M. H., Busschbach, J. J. V., van Horn, J. E., & Scholte, R. H. J. (2016). Alliance and adherence in a systemic therapy. *Child and Adolescent Mental Health*. doi:10.1111/camh.12172.

Lebensohn-Chialvo, F., Hasler, B. P., Rohrbaugh, M. J., & Shoham, V. (2016). *Does therapist responsiveness influence fidelity in family therapy for adolescent substance use?* Poster presented at the Society for Psychotherapy Research International Annual Meeting, Jerusalem, Israël.

Lerner, M. D., Mikami, A. Y., & McLeod, B. D. (2011). The alliance in a friendship coaching intervention for parents of children with ADHD. *Behavior Therapy, 42*, 449-461.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88,* 767–778.

Manders, W. A., Deković, M., Asscher, J. J., van der Laan, P. H., & Prins, P. J. M. (2011). De implementatie van Multisysteem Therapie in Nederland: De invloed van behandelintegriteit en nonspecifieke factoren op behandeluitkomsten. [The implementation of Multisystemic Therapie in the Netherlands: The effect of adherence and nonspecific factors on treatment outcomes] *Gedragstherapie, 44*, 327-340.

Marker, C. D., Comer, J. S., Abramova, V., & Kendall, P. C. (2013) The reciprocal relationship between alliance and symptom improvement across the treatment of childhood anxiety. *Journal of Clinical Child & Adolescent Psychology, 42*, 22-33.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68*, 438-450.

McLeod, B. D. (2011). Relation of the alliance with outcomes in youth psychotherapy: A meta-analysis. *Clinical Psychology Review, 31*, 603-616.

McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodríguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health. *Clinical Psychology: Science and Practice, 20*, 14-32.

Mihalic, S. (2004). The importance of implementation fidelity. *Report on Emotional & Behavioral Disorders in Youth, 4*, 83-105.

MST Institute (n.d.a). *MST Therapist Adherence-Revised (TAM-R): Guidelines for administration.* Retrieved from ://www.mstinstitute.org/qa_program/tam.shtml

MST Institute (n.d.b). *MST Therapist Adherence Measure – Revised (TAM-R): Guidelines for interpretation*. Retrieved from ://www.mstinstitute.org/qa_program/tam.shtml

MST Institute (2016). *Frequently Asked Questions. MST Institute Enhanced Website*. Retrieved May 2016 from https://www.msti.org/documents/EW_FAQs.pdf.

MST Services (2014). *Multisystemic Therapy® (MST®) Organizational Manual*. Charleston, SC: Author.

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus user's guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*, 882-891.

Nagin, D. S., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development, 70*, 1181-1196.

Nagin, D. S., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological Methods, 6*, 18-34.

Norcross, J. C., & Wampold, B. E. (2011). Evidence-based therapy relationships: Research conclusions and clinical practices. *Psychotherapy, 48*, 98-102.

Owen, J., Miller, S. D., Seidel, J., & Chow, D. (2016). The working alliance in treatment of military adolescents. *Journal of Consulting and Clinical Psychology, 84*, 200-210.

Paradisopoulos, D., Pote, H., Fox, S., & Kaur, P. (2015). Developing a model of sustained change following MST: Young people's perspectives. *Journal of Family Therapy, 37*, 450-470.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change. Issues and research recommendations. *Clinical Psychology: Science and Practice, 12*, 365-383.

Prince, M. A., Connors, G. J., Maisto, S. A., & Dearing, R. L. (2016). Within treatment therapeutic alliance ratings profiles predict posttreatment frequency of alcohol use. *Psychology of Addictive Behaviors, 30*, 184-193.

Raftery, A. E. (1999). Bayes factors and BIC: Comment on "A critique on the Bayesian Information Criterion for Model Selection". *Sociological Methods & Research, 27*, 411-427.

Robbins, M. S., Feaster, D. J., Horigian, V. E., Puccinelli, M. J., Henderson, C., & Szapocznik, J. (2011). Therapist adherence in brief strategic family therapy for adolescent drug abusers. *Journal of Consulting and Clinical Psychology, 79*, 43-53.

Robbins, M. S., Turner, C. W., Alexander, J. F., & Perez, G. A. (2003). Alliance and dropout in family therapy for adolescents with behavior problems: Individual and systemic effects. *Journal of Family Psychology, 17*, 534-544.

Schoenwald, S. K. (2008). Toward evidence-based transport of evidence-based treatments: MST as an example. *Journal of Child and Adolescent Substance Abuse, 17*, 69-91.

Schoenwald, S. K. (2011). It's a bird, it's a plane, it's ... fidelity measurement in the real world. *Clinical Psychology: Science and Practice, 18*, 142-147.

Schoenwald, S. K., Chapman, J. E., Sheidow, A. J., & Carter, R. E. (2009). Long-term youth criminal outcomes in MST transport: The impact of therapist adherence and organizational climate and structure. *Journal of Clinical Child and Adolescent Psychology, 38*, 91–105.

Schoenwald, S. K., Sheidow, A. J., & Chapman, J. E. (2009). Clinical supervision in treatment transport: Effects on adherence and outcomes. *Journal of Consulting and Clinical Psychology, 77*, 410-421.

Sexton, T., & Turner, C. W. (2010). The effectiveness of functional family therapy for youth with behavioral problems in a community practice setting. *Journal of Family Psychology, 24*, 339-348.

Stiles, W. B., Glick, M. J., Osatuke, K., Hardy, G. E., Shapiro, D. A., Agnew-Davies, R., Rees, A., & Barkham, M. (2004). Patterns of alliance development and the rupture-repair hypothesis: Are productive relationships U-shaped or V-shaped? *Journal of Counseling Psychology, 51*, 81-92.

Tschuschke, V., Crameri, A., Koehler, M., Berglar, J., Muth, K., Staczan, P., von Wyl, A., Schulthess, P., & Koemeda-Lutz, M. (2015). The role of therapists' treatment adherence, professional experience, therapeutic alliance, and clients' severity of psychological problems: Prediction of treatment outcome in eight different psychotherapy approaches. Preliminary results of a naturalistic study. *Psychotherapy Research, 25*, 420-434.

Verhulst, F. C. & van der Ende, J. (2001). *Gedragsvragenlijst voor kinderen van 6 tot 18 jaar [CBCL 6-18].* Rotterdam, the Netherlands: Erasmus MC-Sophia Kinderziekenhuis.

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*, 200-211.

Weck, F., Grikscheit, F., Jakob, M., Höfling, V., & Stangier, U. (2015). Treatment failure in cognitive-behavioral therapy: Therapeutic alliance as a precondition for an adherent and competent implementation of techniques. *British Journal of Clinical Psychology, 54* , 91–108.

Weiss, M., Kivity, Y., & Huppert, J. D. (2014). How does the therapeutic alliance develop throughout cognitive behavioral therapy for panic disorder? Sawtooth patterns, sudden gains, and stabilization. *Psychotherapy Research, 24,* 407-418.

Wood, A. M., White, I. R., & Thompson, S. G. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials, 1*, 368-376.

Chapter 7

General discussion

Summary of the Main Findings

The aim of this dissertation was twofold. The first aim was to evaluate some of the factors that may affect the assessment of therapist adherence when disseminating evidence-based interventions. Using the Therapist Adherence Measure-Revised (TAM-R) as an example, we set out to investigate how the Dutch TAM-R related to the US original and how adherence scores may be affected by the level of experience that therapists, teams, or countries have with the treatment model.

The second aim was to assess the unique and joint role of adherence and alliance within system therapy. Little is known on how adherence to the specific evidence-based elements of interventions relates to common factors such as alliance, nor on how they jointly or individually affect treatment outcome. This dissertation set out to investigate this topic within the context of Multisystemic Therapy.

This chapter provides a summary of the main findings, followed by reflections on these findings, strengths and limitations, directions for future research, and recommendations for health policy and clinical practice.

Aim I. Factors affecting reliable assessment of therapist adherence after cross-national dissemination

1. Is the Dutch TAM-R equivalent to the original US version?

The TAM-R (Henggeler, Borduin, Schoenwald, Huey, & Chapman, 2006) measures adherence to Multisystemic Therapy (MST), an evidence-based, intensive home- and community-based intervention for 12-18 years old adolescents with antisocial and/or delinquent behavioral problems (Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 2009). The TAM-R is one of the few adherence measures which is known to predict post-treatment outcomes, and to be reliable and valid (e.g., Henggeler, Halliday-Boykins et al., 2006; Schoenwald, Carter, Chapman, & Sheidow, 2008). Nevertheless, marked differences were observed between its functioning in the United States (US), where MST was originally developed, and in Europe. This led clinicians and scholars to question its reliability and validity after dissemination of MST and the TAM-R to European countries. Therefore, the first research question investigated how the Dutch TAM-R relates to the original US version using 1875 Dutch TAM-R reports and the response category frequency distributions of 1875 US TAM-R reports.

Chapter 2 showed some significant differences between the Dutch TAM-R and the original US TAM-R. In the first place, differences were observed regarding the level of difficulty of some of the items. This means that some Dutch items were very 'easy' to score (most Dutch therapists had a high score on these items), whereas high scores on the same items proved very difficult to achieve in the US (few US therapists had a high score on them). For scores on other items, it was just the other way around, being hard to achieve in

the Netherlands and easy in the US. This could point to bias regarding these specific items. In the second place, differences were observed on the response frequencies in general: The upper response categories were used more frequently by US respondents than by Dutch respondents for almost all items. This pointed to problems of nonequivalence at a more general level, affecting how the TAM-R as a whole was functioning across countries. Thus, there was not a one-to-one relation between the Dutch and US TAM-R. Chapter 2 set out to investigate potential sources of this bias.

Based on feedback from MST consultants, the role of the translation was tested as a potential source of bias. For this purpose, 237 families participating in MST were randomly assigned to one of three versions of the TAM-R: 1) the original Dutch TAM-R, 2) a new translation of the items, or 3) a new translation of the items and a new formulation of the response categories. The results showed that translation can have a significant impact on scores, as a different wording of the response categories led to very different respondent scoring, increasing the differences between the Dutch and US TAM-R instead of reducing them. Yet, the new translation of the items did not result in a reduction of the number of items that differed on the level of difficulty. Thus, bias could not be reduced with the translation.

Although no further sources of bias could be tested, the discussion section of Chapter 2 did review other potential sources of bias. Response style was identified as the most promising. As described above, US respondents used the upper response categories more frequently than Dutch respondents. This may reflect different response tendencies rather than true differences. Indeed, large true differences between US and Dutch levels of therapist adherence would not be expected given the elaborate quality-assurance system of MST, as well as the fact that the Dutch MST outcomes were comparable to the US outcomes (based on Dutch dashboards and yearly MST Data reports). This hypothesis regarding the role of differing response tendencies was further supported by findings from a large cross-national survey showing that US participants are more prone to using positive, but not negative, extreme response categories than Dutch respondents (Harzing, 2006).

2. How does experience affect therapist adherence?

Chapter 3 analyzed the effect of therapist experience, team experience, and experience at the country-level on therapist adherence. For this purpose, a Swedish study (Löfholm, Eichas, & Sundell, 2014) was replicated using data from 4290 families treated by 222 different therapists working across 27 teams during the first 10 years after implementation of MST in the Netherlands.

The study found that both therapist experience and country-wide experience predicted higher therapist adherence, leading to better treatment outcomes at the end of treatment. Individual therapists may need some time to familiarize themselves with the MST model

before being optimally adherent, as has been suggested by previous research as well (MST Institute, 2014). This chapter further showed that experience may also be required at a country-wide level before high levels of therapist adherence can be expected. For example, organizations and MST-the Netherlands may need some time to learn how to adherently implement the intervention in the new context and familiarize themselves with the quality-assurance system to support therapist adherence. This is in line with other research suggesting that effective implementation may take some time (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005; McLeod, Southam-Gerow, Tully, Rodriguez, & Smith, 2013; Schoenwald & Garland, 2013).

Contrary to the Swedish study, no effect was found for team experience on adherence and outcome. More so, there was a tendency in the other direction, suggesting that more team experience may actually lead to lower therapist adherence and hence worse treatment outcomes. Anecdotal evidence suggested that team supervisors may pay less attention to therapist adherence as time progresses. Alternatively, supervisor turnover may have played a role.

Aim II: Assess the unique and joint role of adherence and alliance within system therapy

3. How do adherence and alliance interrelate over time during treatment?

Chapter 5 studied how therapist adherence and alliance interrelate over time during treatment using data of 1970 families who had participated in MST. It was found that alliance in a particular month predicted adherence in the subsequent month throughout treatment. Adherence on the other hand predicted alliance in the subsequent month, but only in the middle part of treatment. These results suggest that adherence and alliance can reinforce one another. A strong therapeutic relationship at the start of treatment may facilitate adherent implementation of the model, whereas adherence may further strengthen the working alliance.

Alliance is important for creating engagement and motivation for treatment, enhancing confidence in the skills of the therapist, and foster hope for change. When alliance is good, clients may be more likely to accept the suggestions and interventions proposed by the therapist than in the absence of alliance. Adherence may further enhance the alliance as clients' confidence in the skills of the therapist further increases (Goldfried & Davila, 2005; Hill, 2005).

Adherence only predicted alliance in the middle part of treatment, but not at the start and end of treatment. This may be a consequence of the adherence-measure used in the current study, which heavily focused on those therapist behaviors that are most relevant during the middle part of the treatment. Alternatively, it may also be characteristic of the alliance-adherence association, as previous studies have suggested that a strong working alliance may be a prerequisite for adherence (Hukkelberg & Ogden, 2013; Weck, Grikscheit, Jakob, Höfling, & Stangier, 2015).

4. How do adherence and alliance develop over time during treatment?

Chapter 6 analyzed the development of adherence and alliance during treatment using data from 848 families who had participated in MST. Using latent growth models and latent class growth analysis, it was found that both therapist adherence and the therapeutic alliance change during the treatment. Within MST, alliance is generally well established after the first few weeks and this is maintained throughout treatment. This is consistent with the MST treatment model (Henggeler et al., 2009). Early adherence levels, on the other hand, were more moderate and increased more pronouncedly throughout treatment. As was shown in Chapter 5, alliance may function as a catalyst for therapist adherence: Establishing a good working relationship and setting common goals may facilitate implementing the treatment model adherently. This may explain why adherence is only moderate at the start of the treatment and improves as treatment progresses.

These results do not mean that adherence and alliance will always be characterized by moderate to high increasing levels. Instead, the results most probably reflect the setting of MST. MST pays a lot of attention to adherence and alliance during supervision and consultation, thereby helping therapists improve these two aspects in their treatments. This is part of the MST quality-assurance system, intended to ensure the same quality of treatment-delivery throughout settings. In fact, studies have reported that adherence and alliance may deteriorate during other types of treatment, which has been associated with poorer treatment outcomes (Chiapa et al., 2015; Hudson et al., 2014; Weiss, Kivity, & Huppert, 2014).

5. How do adherence and alliance uniquely and jointly contribute to short- and long-term treatment outcomes?

Chapter 6 (N = 848) also studied how these trajectories of adherence and alliance predicted treatment outcomes (no out-of-home placement, no police contact, and adolescent attending school or work) at the end of treatment and at 18-months post-treatment. Both adherence and alliance contributed to short-term but not to long-term treatment outcomes. Higher levels of adherence, and increasing levels of adherence and alliance were associated with better treatment outcomes at the end of the treatment. Similar to previous findings, change in alliance during treatment was most predictive of outcome (e.g., Owen, Miller, Seidel, & Chow, 2016). The effects of alliance were less robust than the effects of adherence. This may be due to the fact that very little change was observed in alliance.

Unfortunately, the current study provided little information regarding the unique or joint role of adherence and alliance. As adherence and alliance were reported to be high in almost all of the participating families, it was not possible to evaluate the role of adherence when alliance was low or the other way around. Chapter 5 showed that they interact and reinforce one another. This suggests that adherence and alliance may jointly lead to better

treatment outcomes. However, it remains unclear to what extent they are both necessary to achieve good treatment outcomes, or whether high adherence or strong alliance, in the absence of the other, can lead to good outcomes as well.

Reflections on the Main Findings

Bias or True Differences

As was set out in the introduction of this dissertation, differences had been observed between the functioning of the TAM-R in and outside of the US, namely: 1) mixed findings were reported regarding the relation between the TAM-R and outcome in Europe, and 2) adherence was systematically lower outside than within the US. Dutch therapists further doubted its validity and reliability. These observations could point to a lack of equivalence of the TAM-R, but could also reflect true differences in adherence. This dissertation investigated both perspectives.

In the first place, this dissertation has shown that the Dutch TAM-R is reliable and does predict post-treatment outcomes (Chapter 3 and 6). It further showed that both bias and true differences may have led to the observation that adherence was lower outside than within the US, as described above. Chapter 2 suggested that bias played a role: the Dutch and US TAM-R were not equivalent. Differences in response styles between the countries were put forward as one of the most likely explanations for this bias. Chapter 3 suggested that true differences may also have played a role, since country-wide experience predicted therapist adherence. As the amount of experience with MST differs between the US and Europe, the levels of adherence may have truly been lower in the Netherlands than in the US. These differences between the Netherlands and the US, however, have diminished over time, as therapists, teams, organizations, and MST-the Netherlands have become increasingly proficient in adherently providing MST to the Dutch youth population (Chapter 2 and 3). It is unclear whether remaining differences in adherence-scores can be attributed to bias, and if so, to what extent.

Reducing all bias may be impossible to achieve (see, for example, Gudmundsson, 2009), and it may not be necessary either. Implementation of interventions in a new setting often requires some adaptations to be made in order for the intervention to be effective (Durlak & Dupre, 2008). For example, a recent study showed that evidence-based interventions in a new setting were less effective than adaptations of evidence-based interventions or novel programs developed for that setting (Sundell, Beelmann, Hasson, & von Thiele Schwarz, 2016). Thus, adaptations may sometimes be necessary to fit to the new setting and warrant good outcomes.

For the Dutch TAM-R, as for any other adherence-instrument, this means that cross-national comparisons should be made with caution and interpretation of low adherence

scores should be discussed with clinical practice to obtain an understanding of the causes and consequences of these scores. In these discussions, one should be wary of the numerous factors that can affect the *assessment* of adherence on the one hand, such as different response styles or differences in the administration of the instrument, as well as factors affecting the adherence itself. Low adherence scores do not necessarily represent a problematic situation, but may represent effective adaptations to the new or ever-changing environment (e.g., organizational structure, funding, community relations; Fixsen et al., 2005; Schoenwald, Heiblum, Saldana, & Henggeler, 2008). Although adherence is an essential element in monitoring successful implementation, there is a balance to strike between adherence and adaptation.

Adherence and Adaptability

This balance between adherence and adaptation is not only relevant at an implementation level, where interventions need to fit in the new setting, but also at the level of the clinician. As was set out in the general introduction, some clinicians dislike focusing on adherence, as they believe that manuals and protocols may be too rigid, and that they may hamper client-adjusted treatment and successful formation of alliance (Gyani, Shafran, Rose, & Lee, 2015; Nelson, Steele, & Mize, 2006). This dissertation showed that, contrary to this believe, alliance and adherence can actually reinforce instead of hinder one another. Thus, where alliance may form the basis to deliver the evidence-based treatment components, adherence may be important to strengthen the client-therapist relationship. This is novel information which may help tempering worries of clinicians regarding the effect of adopting evidence-based interventions into their clinical work.

However, it is likely that not *all* manuals and protocols will by default reinforce client-therapist alliances. Instead, the current findings have been gathered in a rather unique setting and this setting must be taken into account in the interpretation of the results. All data were collected within the context of MST, an evidence-based program that promotes 'flexibility within fidelity'. Indeed, there is some evidence that a too strict application of a manual can lead to poorer treatment outcomes (Barber et al., 2006; Hogue et al., 2008). The MST model is largely principle-based so that therapists can adjust the treatment to the individual client while remaining adherent. For example, the MST model provides a framework for how to identify the problems and their causes, but therapists can then choose which elements or techniques they use to tackle the client-specific problems. The weekly supervision and consultation sessions, as well as the quarterly booster sessions, may further help therapists to maintain their focus on adherence while at the same time taking into account the unique needs of the families (Henggeler et al., 2009; Schoenwald, 2008).

More research is needed on this balance between adherence and adaptation of interventions. Flexibility at the level of clinicians requires well-thought interventions or

programs which clearly distinguish between core elements and optional elements, and which carefully think through how 'flexibility within fidelity' can be facilitated. Adaptability of interventions at an implementation level may require a mind-switch regarding the role of adherence. Adherence should be monitored and reported as it is a valuable tool in the evaluation of implementation success and an informative validity check within research (was the intervention delivered as intended). At the same time, low levels of adherence should not by default be characterized as implementation failures. Adaptations to a program can be very beneficial, but can only be characterized as such if they are monitored and described in publications so that the community of researchers, program-developers, and clinicians can start to develop an understanding of which elements are truly crucial and which may be altered to improve effectiveness (see Sundell et al., 2016 for a similar plea). Thus, clinicians, program-developers, and researchers should be conscious of adaptations and report these, whereas publishers and reviewers should realise that low adherence, or descriptions of adaptations, are not necessarily an indication of an implementation failure and therefore a disqualification, but rather may be interesting and informative.

Strengths and Limitations

This dissertation has several strengths. For example, we used longitudinal data (monthly assessments during treatment and half-yearly assessments up to 18 months post-treatment) from a few hundred up to almost two thousand clients, and data were analyzed from different perspectives using Rasch analysis, multilevel structural equation modeling, growth curve modeling, and latent class growth analysis. Nevertheless, several limitations should be considered when judging the conclusions.

Firstly, this dissertation was restricted to MST. MST was chosen as it is one of the few evidence-based interventions with a valid and reliable adherence-instrument that has been found to relate to treatment outcome. MST is at the top in several databases for evidence-based interventions, such as the Blueprints Programs in the United States (http://www.blueprintsprograms.com/) and the Dutch Database for effective interventions in the Netherlands (http://nji.nl/nl/Databank/Databank-Effectieve-Jeugdinterventies). MST thus has a model function for other programs and interventions and is, therefore, an interesting example to investigate.

Nevertheless, the sole focus of this dissertation on MST has some limitations. Several chapters noted little variance in one or more of the variables of interest. For example, treatment outcomes were good for the majority of the participating families and almost all families reported sustained high levels of alliance. This means that these data cannot inform us on treatment failure or the role of very low alliance scores (Chapter 3 and 6). The focus on MST may also limit the generalizability of the findings to other interventions and populations. Although we have no evidence to assume that the processes described in this

dissertation would operate differently in other situations, we have not actually tested this assumption.

Secondly, the TAM-R has some limitations as well. Ideally, adherence should be scored by trained raters or treatment experts based on video observations (Goense, Assink, Stams, Boendermaker, & Hoeve, 2016; Schoenwald et al., 2011). However, this method is rather expensive and is, therefore, not often used in clinical practice. As caregiver-reports are more efficient and were found to be a good predictor of outcome (Henggeler, Melton, Brondino, Scherer, & Hanley, 1997), the developers of MST chose to implement caregiver-rated adherence scores in the MST clinical practice. Yet, caregiver-reports have been reported to show ceiling effects, something which has also been observed in our studies. Caregivers may be less able to critically judge whether a treatment component was delivered, for example because they have not been trained to appraise treatment delivery or because they feel loyalty towards their therapist to assess him or her favorably (Chapman, McCart, Letourneau, & Sheidow, 2013; Schoenwald et al., 2011). As Löfholm and colleagues (2014) observed, the TAM-R may have been developed to identify unexpected low adherence, but not to capture minute variation. Thus, although the TAM-R has been found to be reliable and valid (e.g., Henggeler, Halliday-Boykins et al., 2006; Schoenwald et al., 2008), it may not have been sensitive enough to capture *all* variation and related associations. Future research may benefit from more sensitive measures.

Lastly, we used routinely collected data, which allowed us to use large sample sizes. Besides this benefit, the vast amount of data that is currently collected on a routine basis also *obliges* researchers to carefully consider the possibilities of these data to answer research questions and restrain from unnecessary collection of new data. This dissertation shows that available data can be used to answer research questions that are relevant to both research and clinical practice.

Using routinely collected data also means we had to deal with some challenges. The biggest challenge was the amount of missing data, ranging up to 54% at follow-up in Chapter 6. Although some may question the validity of the results, we do believe that this data can be informative provided that the data and results are critically appraised. To achieve this we analyzed the missing data patterns, collected data on the cause of the missing data, and used full maximum likelihood estimation to account for the missing data. This analytical method uses all available data instead of dropping cases with missing data to estimate the outcomes. It is comparable to an infinite number of imputations. These different steps (see Chapter 6 for details) support the reliability of the results.

Directions for Future Research

The topics described above provide a number of directions for future research. Several pleas have been made to clearly define adherence and develop uniform and well-validated

adherence-measures (McLeod et al., 2013; Schoenwald & Garland, 2013). Based on the results presented in this thesis, I can only but agree with this, as valid and reliable assessments are essential if anyone wants to draw any conclusion about what is actually being delivered. To achieve this, researchers and program-developers should at least agree upon the focus of adherence-items (e.g., context, material, therapist behavior, or client behavior) and the level of detail required (e.g., measuring principles or concrete behaviors). Given that there is great overlap between interventions for similar populations (Sexton & Kelley, 2010; Spanjaard, Veerman, & van Yperen, 2015) it may be worthwhile to develop instruments that measure a broad array of elements, so that the same instrument can be used for different interventions (McLeod et al., 2013).

Initiatives have been launched trying to list and measure the elements of interventions for specific populations, in the US (Chorpita & Daleiden, 2009; Michie et al., 2013), as well as in the Netherlands (Spanjaard et al., 2015). These initiatives move away from protocolized interventions towards valuing individual techniques within a treatment as components or practice elements that can be tested for their effectiveness. This allows for more flexible and modular treatments as clinicians can deliver a combination of elements best suited for each individual client (e.g., Chorpita et al., 2017). This is a promising new direction which nicely seems to tackle this issue around flexibility within fidelity.

Besides continuing efforts to distill practice elements and develop modular treatments, this direction of research may have several additional research implications. Adherence will continue to be a relevant measure as these individual elements equally need to be delivered as intended to be effective. It may furthermore require implementation of quality-control methods to support clinicians during their work in choosing how and when to provide the different elements. Such methods have been developed within evidence-based interventions to sustain adherence and assure good outcomes (Henggeler & Sheidow, 2012) and may consist of coaching, training, or monitoring instruments (Garland & Schoenwald, 2013). Similar methods could be very suitable and beneficial for quality-assurance within modular treatments. One example comes from Chorpita and colleagues, who have developed monitoring instruments for a modular treatment for children and adolescents (Chorpita, Bernstein, Daleiden, & The Research Network on Youth Mental Health, 2008). These instruments consist of dashboards and roadmaps helping the clinician in the decision making process.

The costs associated with such a quality-assurance system, be it for modular treatments or to assure the quality of care as usual, will not always be feasible in daily practice (Garland & Schoenwald, 2013; Schoenwald et al., 2011). It may not be needed in all situations either. More research is needed to identify the specific role of the diverse elements of quality-assurance systems in order to evaluate which methods are needed in which situations.

When measuring adherence to a specific program or intervention, for example as part of an assessment of the implementation success, it seems important for the adherence-

measure to being able to distinguish between beneficial adaptations and actual deviations from adherence. To successfully come at such an instrument, researchers and program developers need to identify what elements are crucial and which may be altered without loss of effectiveness. Studies could try to discern what *type* of elements (e.g., material, content, format, population, setting, order/schedule of the content) are more or less suitable for adaptation. Program-developers should clearly describe which elements in their program are crucial to achieve good outcomes.

This information can subsequently be used to support individual therapists in striking a balance between adherence and flexibility. Some evidence-based interventions have attempted to describe how they promote flexibility within the treatment (Forehand, Dorsey, Jones, Long, & McMahon, 2010; Mazzucchelli & Sanders, 2010). More research on how therapists can be supported herein is necessary, both within the field of protocolized interventions, as well as within usual care, where clinicians may be working with protocols or manuals, but with much less attention to sustaining adherence.

Recommendations for Health Policy and Clinical Practice

The above discussions and research findings also provide some guidance and recommendations for practice, which will be outlined below.

Policymakers and program managers

- ***Adaptations to a program or intervention may be beneficial but should be included with care***

Programs may benefit from adaptations to fit to the new setting, but adaptations should only be applied if the following points are being taken into account: 1) Discuss with the program developers what elements of the program belong to the effective core and what elements may be suitable for changes and 2) Monitor adaptations to the program itself as well as adherence of the clinicians. This allows outcomes to be linked to what has actually been delivered and provides clinicians, program developers, and researchers with information on the effect of such adaptations or deviations from the model.

- ***Continuously assess adherence and discuss these scores with stakeholders to understand underlying processes***

Implementation of an intervention is not a single act with ever-lasting results. Rather, it is an ongoing process. Hence, adherence levels may vary over time. Implementation of evidence-based interventions thus requires continued assessment of adherence to evaluate whether the intervention is being delivered as intended.

Adherent delivery of an intervention is affected by many factors in the environment, such as the organizational structure, the clinicians, the client population, funding and regulation, and even time or the experience of clinicians and organizations with the intervention. Therefore, there should be an ongoing dialogue between different stakeholders and parties involved to evaluate the meaning of scores, taking these changing factors into account.

- ***Use validated and appropriate measures to assess adherence***

Measuring adherence is essential for a valid implementation of a therapy or program. Lack of adherence, thus failure to implement the core elements of an effective program or intervention, can lead to poorer or even detrimental treatment outcomes. Measuring adherence can help to identify and address problems and improve the delivery of the program and its outcomes.

An instrument transported to a new country or cultural group may function differently due to cultural differences. Translation of the instrument, as well as appropriate administration, differences in the client or clinician populations, and possible differences in response style should be carefully considered before and during translation (see for example Gudmundsson, 2009; Peña, 2007; Sousa & Rojjanasrirat, 2010 for extensive descriptions of optimal translation procedures).

- ***Provide clinicians with feedback and support around adherence and alliance***

Both adherence and alliance change over time, even within a treatment. This means that therapists should be attentive to these aspects in their treatment in order for them to address drops in the level of adherence or alliance. To facilitate this, clinicians may benefit from feedback on their alliance and adherence (Bickman, Kelley, Breda, de Andrade, & Riemer, 2013). Moreover, they may need support to be adherent to the model while adapting to the individual client and unique setting. This requires support for clinicians to discern flexibility from drift, in other words, support to know what elements of an intervention can be adapted to the unique situation and what elements should be delivered as intended for the intervention to be effective. Such support and feedback should be facilitated to warrant and improve quality of care.

- ***Quality-assurance systems are a warranty for the quality of the youth care delivered***

Programs that have incorporated a quality-assurance system or quality-control methods may be a good choice. These quality-control methods (e.g., supervision and training) provide support and feedback to clinicians and thereby help them to achieve adherent and competent delivery of the program, leading to good outcomes.

- ***Implementation of evidence-based interventions may take some time before showing good outcomes***

Implementing evidence-based interventions in a new setting takes time. The new setting may need to adapt their processes or ways of thinking to incorporate the program, but also the program may need to be adapted to fit the new setting. It may take up to a couple of years before the program is achieving representative outcomes (see also Durlak & Dupre, 2008).

- ***The new Dutch translation of the TAM-R is preferable over the current translation***

The new translation of the TAM-R items was received favorably by caregivers and its functioning proved to be comparable to the original Dutch translation. Therefore, we advise MST-the Netherlands to implement the new translation.

Clinicians

First of all, the results show that most MST-therapists are adherent to the MST model, that they achieve good alliances with their clients, and create good treatment outcomes. This high quality of treatment delivery did, however, have a downside for our research project, as the lack of low scores meant there was only little variance to drive correlations between outcome and adherence. Nevertheless, we were able to come with some recommendations that may help to improve the clinical work of MST even further, and which might also have broader implications for psychotherapists.

- ***Adherence to an evidence-based program can reinforce the client-therapist alliance***

Adhering to an evidence-based program does not need to harm the client-therapist alliance. Adherence can actually reinforce alliance, most likely by reinforcing the confidence of the client in the skills of the therapist.

- ***Be attentive of changes in adherence and alliance during treatment***

Delivering the critical elements of an intervention, as well as developing a strong client-therapist relationship, is beneficial for good treatment outcomes. As alliance and adherence can change during treatment, it is important for clinicians to be attentive of this possibility and adapt their behavior in treatment as necessary.

- ***Be transparent about occasional deviations from the program and report them if these are systematic***

If clinicians feel that a deviation from the model or protocol is necessary in a certain situation, they should be transparent about this to allow for reflection on what kind of deviations

do and do not work. Systematic deviations should be written down, whereas occasional deviations should be discussed and evaluated with colleagues. As such, clinicians can help each other to identify what behaviors and elements may be suitable for adaptation and what should be retained.

- ***Becoming skillful in adherent delivery of a new program takes time***

Learning a new program takes time and continues while working with clients. Thus, most clinicians will not achieve high levels of adherence straight on, but this will most likely increase after having worked with a couple of families under supervision. This also applies at an organizational or country-wide level. It may take some time before a transported program reaches its full potential when implemented in a new setting.

- ***The Dutch TAM-R is reliable and associated to treatment outcomes***

This dissertation has shown the Dutch TAM-R to be reliable and to predict treatment outcomes. Although there are some disadvantages to the current TAM-R, as discussed above, the TAM-R seems to be useful as part of the quality-assurance system of MST to improve the quality of care. Given the limitations of the TAM-R, however, the adherence scores should be interpreted in conjunction with the other indices in the quality-assurance system to develop a comprehensive understanding of the situation. Moreover, the full range of response categories should be used to interpret the adherence score whenever possible, as the current study has shown that, contrary to the categories of the US TAM-R, the fourth category is frequently used in the Netherlands.

Clients

- ***A quality-assurance system provides a warrant for the quality of the program***

A quality-assurance system can consist of different methods incorporated in a program or treatment to monitor, evaluate, and improve the quality of the delivery of the program. Some more frequently-used methods are supervision of the clinicians and questionnaires to assess adherence or outcomes. Although many factors will play a role in the effectiveness of a specific program or treatment for a specific client, programs incorporating such quality-control methods do provide some warrant for the quality of the program. Therefore, this could be taken into consideration if clients are provided a choice between different treatment options.

- **Participate in initiatives to routinely collect data if requested**

Monitoring instruments, such as questionnaires to assess adherence or outcome, are used to routinely evaluate the quality of programs. They are important tools in providing high

quality care, but will only be a reliable measure if enough clients decide to participate. It is therefore warmly recommended to participate in such data collection as it provides a valuable tool to improve care.

Concluding Remarks

To conclude, it is essential to measure adherence well in order to assess what is actually being delivered to youth and their families, warrant high quality of care, and further improve and develop youth care. However, adequate assessment of adherence is not an easy task given the lack of consensus on how adherence should be measured (e.g., level of detail, type of elements), as well as the lack of a clear distinction between the actual core elements and other intervention-specific elements that *can* be changed without loss of effectiveness. Intriguing and relevant topics are thus awaiting researchers and clinicians. I hope this dissertation has contributed to the debate regarding the role of adherence within clinical practice and can assist in furthering the field to help provide the best possible care.

References

Barber, J. P., Gallop, R., Crits-Christoph, P., Frank, A., Thase, M. E., Weiss, R. D., & Gibbons, M. (2006). The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: Results from the national institute drug abuse collaborative cocaine treatment study. *Psychotherapy Research*, 16, 229-240.

Bickman, L., Kelley, S. D., Breda, C., de Andrade, A. R., & Riemer, M. (2013). Effects of routine feedback to clinicians on mental health outcomes of youths: Results of a randomized trial. *Psychiatric Services, 62*, 1423-1429.

Chapman, J. E., McCart, M. R., Letourneau, E. J., & Sheidow, A. J. (2013). Comparison of youth, caregiver, therapist, trained, and treatment expert raters of therapist adherence to a substance abuse treatment protocol. *Journal of Consulting and Clinical Psychology, 81*, 674-680.

Chiapa, A., Smith, J. D., Kim, H., Dishion, T. J., Shaw, D. S., & Wilson, M. N. (2015). The trajectory of fidelity in a multiyear trial of the family check-up predicts change in child problem behavior. *Journal of Consulting and Clinical Psychology*, 83, 1006-1011.

Chorpita, B. F., Bernstein, A., Daleiden, E. L., & The Research Network on Youth Mental Health (2008). Driving with roadmaps and dashboards: Using information resources to structure the decision models in service organizations. *Administration Policy & Mental Health, 35*, 114-123.

Chorpita, B. F., & Dalieden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology, 77*, 566-579.

Chorpita, B. F., Daleiden, E. L., Park, A. L., Ward, A. M., Levy, M. C., Crommley, T., Letamendi, A. M., Tsai, K. H., & Krull, J. L. (2017). Child STEPs in California: A cluster randomized effectiveness trial comparing modular treatment with community implemented treatment for youth with anxiety, depression, conduct problems, or traumatic stress. *Journal of Consulting and Clinical Psychology, 85*, 13-25.

Durlak, J. A., & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327-350.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Forehand, R., Dorsey, S., Jones, D. J., Long, N., & McMahon, R. J. (2010). Adherence and flexibility: They can (and do) coexist! *Clinical Psychology: Science & Practice, 17*, 258-264.

Garland, A., & Schoenwald, S. K. (2013). Use of effective and efficient quality control methods to implement pscyhosocial interventions. *Clinical Psychology: Science & Practice, 20*, 33-43.

Goense, P. B., Assink, M., Stams, G.-J., Boendermaker, L., & Hoeve, M. (2016). Making 'what works' work: A meta-analytic study of the effect of treatment integrity on outcomes of evidence-based interventions for juveniles with antisocial behavior. *Aggression and Violent Behavior, 31*, 106-115.

Goldfried, M. R., & Davila, J. (2005). The role of relationship and technique in therapeutic change. *Psychotherapy: Theory, Research, Practice, Training*, 42, 421-430.

Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology, 61*, 29-45.

Gyani, A., Shafran, R., Rose, S., & Lee, M. J. (2015). A qualitative investigation of therapists' attitudes towards research: Horses for courses? *Behavioural and Cognitive Psychotherapy*, 43, 436-448.

Harzing, A. (2006). Response styles in cross-national survey research. *International Journal of Cross-Cultural Management, 6*, 243-266.

Henggeler, S. W., Borduin, C. M., Schoenwald, S. K., Huey, S. J., & Chapman, J. E. (2006). *Multisystemic Therapy Adherence Scale-Revised (TAM-R).* Unpublished instrument. Charleston, SC: Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina.

Henggeler, S. W., Halliday-Boykins, C. A., Cunningham, P. B., Randall, J., Shapiro, S. B., & Chapman, J. E. (2006). Juvenile Drug Court: Enhancing outcomes by integrating evidence-based treatments. *Journal of Consulting and Clinical Psychology, 74*, 42-54.

Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic Therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65*, 821-833.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (2009). *Multisystemic therapy for antisocial behavior in children and adolescents*. New York: Guilford.

Henggeler, S. W., & Sheidow, A. J. (2012). Empirically supported family-based treatments for conduct disorder and delinquency in adolescents. *Journal of Marital and Family Therapy, 38*, 30-58.

Hill, C. E. (2005). Therapist techniques, client involvement, and the therapeutic relationship: Inextricably intertwined in the therapy process. *Psychotherapy: Theory, Research, Practice, Training, 42*, 431-442.

Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology, 76*, 544-555.

Hudson. J. L., Kendall, P. C., Chu, B. C., Gosch, E., Martin, E., Taylor, A., & Knight, A. (2014). Child involvement, alliance, and therapist flexibility: Process variables in cognitive-behavioral therapy for anxiety disorders in childhood. *Behavior Research and Therapy, 52*, 1-8.

Hukkelberg, S. S., & Ogden, T. (2013). Working alliance and treatment fidelity as predictors of externalizing problem behaviors in parent management training. *Journal of Consulting and Clinical Psychology, 81*, 1010-1020.

Löfholm, C. A., Eichas, K., & Sundell, K. (2014). The Swedish implementation of Multisystemic Therapy for adolescents: Does treatment experience predict treatment adherence? *Journal of Clinical Child & Adolescent Psychology, 43*, 643-655.

Mazzucchelli, T. G., & Sanders, M. R. (2010). Facilitating practitioner flexibility within an empirically supported intervention: Lessons from a system of parenting support. *Clinical Psychology: Science & Practice, 17*, 238-252.

McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodriguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. *Clinical Psychology: Science & Practice, 20*, 14-32.

Michie, S., Richardson, M., Jonhston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, 46, 81-95.

MST Institute (2014). *2014 MST Data overview report*. Retrieved July 2015 from http://www.mstinstitute.org.

Nelson, T. D., Steele, R. G., & Mize, J. A. (2006). Practitioner attitudes toward evidence-based practice: Themes and challenges. *Administration and Policy in Mental Health and Mental Health Services Research, 33*, 398-409.

Owen, J., Miller, S. D., Seidel, J., & Chow, D. (2016). The working alliance in treatment of military adolescents. *Journal of Consulting and Clinical Psychology, 84*, 200-210.

Pena, E.D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development, 78*, 1255-1264.

Schoenwald, S. K. (2008). Toward evidence-based transport of evidence-based treatments: MST as an example. *Journal of Child and Adolescent Substance Abuse, 17*, 69-91.

Schoenwald, S. K., Carter, R. E., Chapman, J. E., & Sheidow, A. J. (2008). Therapist adherence and organizational effects on change in youth behavior problems one year after Multisystemic Therapy. *Administration and Policy in Mental Health and Mental Health Services Research, 35*, 379-394.

Schoenwald, S. K., & Garland, A. F. (2013). A review of treatment adherence measurement methods. *Psychological Assessment, 25*, 146-156.

Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration Policy and Mental Health*, 38, 32-43.

Schoenwald, S. K., Heiblum, N., Saldana. L., & Henggeler, S. W. (2008). The international implementation of Multisystemic Therapy. *Evaluation & the Health Professions, 31*, 211-225.

Sexton, T. L. & Kelley, S. D. (2010). Finding the common core: Evidence-based practices, clinically relevant evidence, and core mechanisms of change. *Administration and Policy in Mental Health and Mental Health Services Research, 37*, 81-88.

Sousa, V. D., & Rojjanasrirat, W. (2010). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research : A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice, 17*, 268-274.

Spanjaard, H. J. M., Veerman, J. W., & van Yperen, T. A. (2015). De kern van effectieve jeugdhulp: Van erkende interventies naar werkzame elementen [The core of effective youth care: From approved interventions to effective elements]. *Orthopedagogiek: Onderzoek & Praktijk, 54*, 441-455.

Sundell, K., Beelmann, A., Hasson, H., & von Thiele Schwarz, U. (2016). Novel programs, international adoptions, or contextual adaptations? Meta-analytical results from German and Swedish intervention research. *Journal of Clinical Child & Adolescent Psychology, 45*, 784-796.

Weck, F., Grikscheit, F., Jakob, M., Höfling, V., & Stangier, U. (2015). Treatment failure in cognitive-behavioural therapy: Therapeutic alliance as a precondition for an adherent and competent implementation of techniques. *British Journal of Clinical Psychology, 54*, 91-108.

Weiss, M., Kivity, Y., & Huppert, J. D. (2014). How does the therapeutic alliance develop throughout cognitive behavioral therapy for panic disorder? Sawtooth patterns, sudden gains, and stabilization. *Psychotherapy Research, 24,* 407-418.

Appendices

Summary

Therapist adherence (the degree to which a clinician delivers the specified components of a treatment), is a quality indicator in routine clinical care when evaluating the success of the implementation of an intervention. Implementing evidence-based interventions into clinical practice with high adherence has proven challenging. Therefore, the **first aim** of this dissertation was to evaluate what factors affect therapist adherence scores when disseminating evidence-based interventions. This first aim is described in Chapter 1 and is subsequently addressed in Chapter 2 and 3.

A focus on adherence presupposes that the specific techniques of the intervention are important for achieving good outcomes. Some scholars, however, have stated that common factors play a much larger role. Common factors are factors common to all psychotherapy, such as the working relationship with the client (alliance). The **second aim** of this dissertation was to study how adherence and alliance interact and uniquely and jointly contribute to treatment outcomes. This second aim is described in Chapter 1 and subsequently addressed in Chapter 4 through 6.

Both aims were addressed using the Dutch Therapist Adherence Measure Revised (TAM-R), developed to asses therapist adherence to the treatment model of Multisystemic Therapy (MST). MST is an evidence-based family- and community-based intervention for antisocial adolescents aged 12 to 18 years. It focuses on addressing all environmental systems that impact the adolescent and works intensively with caregivers to achieve this. The TAM-R is routinely collected in clinical practice as part of the quality-assurance system of MST and is used to monitor and improve adherent delivery of MST worldwide. It is a 28-item questionnaire that is assessed through monthly phone interviews with the primary caregiver during treatment.

Chapter 2 compared the Dutch TAM-R to the original TAM-R from the United States (US) to evaluate equivalence of Dutch and US adherence scores. International implementation of the TAM-R assumes cross-national equivalence, even though this assumption had never been investigated. In Study 1 Rasch analysis was applied to 1,875 Dutch TAM-R reports and the frequency distributions of 1,875 US TAM-R reports. Several items showed differences in their level of difficulty. This means that some Dutch items were very 'easy' to score (most Dutch therapists had a high score on these items), whereas high scores on the same items proved very difficult to achieve in the US (few US therapists had a high score on them). For scores on other items, it was just the other way around. Furthermore, response frequencies were more heavily skewed in the US compared to the Netherlands, meaning that the upper response categories were used more frequently by US respondents than by Dutch respondents for almost all items. Study 2 investigated whether adaptations to the translation of the items and response categories could improve equivalence. For this purpose, 237

families were randomly allocated to 1 of 3 versions (original TAM-R, adapted items only, adapted items and response categories) and the analyses from Study 1 were replicated. Results indicated that equivalence was not improved by the adapted translations. Different response tendencies were identified as the most likely source of bias.

In **Chapter 3** we tested the hypothesis that experience affects adherence scores. To achieve this, it was investigated whether therapist adherence mediates the association between therapist, team, and country-wide experience (i.e., number of years since implementation in the country) on the one hand, and treatment outcome on the other hand. We replicated and extended a study by Löfholm, Eichas, and Sundell (2014). Data over a ten-year period were obtained from 4,290 adolescents who were treated with MST by 222 therapists, working in 27 different teams in the Netherlands. Multilevel structural equation modeling was used to assess the associations between experience, therapist adherence, and post-treatment outcomes. Both therapist experience and country-wide experience predicted therapist adherence, which subsequently predicted treatment outcomes. This suggests that therapists may need some time to familiarize themselves with the model. The effect of country-wide experience on outcome may reflect increasing experience of training and supporting the therapists. It suggests that nation-wide quality control may relate to better therapist adherence and treatment outcome for adolescents treated with systemic therapy. Team experience did not predict better adherence or treatment outcomes.

In **Chapter 4** the factor structure of the Dutch TAM-R was analyzed. We included 580 families who had been treated with MST. As the TAM-R was assessed on a monthly basis during treatment, a Principal Component Analysis (PCA) was conducted on two randomly selected subsamples of 1 TAM-R per family. The solutions of both PCA's were compared and merged to arrive at one optimal solution, which was then tested in a Confirmatory Factor Analysis (CFA) on an independent sample of 723 families treated with MST. The PCA's derived two components, namely 'therapist adherence' and 'client-therapist alliance'. Cross-loading items were removed to retain two clearly distinguishing components. These components showed good fit in the CFA.

In **Chapter 5** these components were used to investigate how alliance and adherence interrelate over the course of the therapy. The bidirectional associations between alliance and therapist adherence were analysed using cross-lagged panel analyses for a sample of 1,970 adolescents and their families having participated in MST. The study found that alliance in one month predicted therapist adherence in a subsequent month. Adherence only predicted subsequent alliance during the middle part of the treatment process. The results were not moderated by client characteristics. The results suggest that alliance and therapist adherence may reinforce one another during therapy. Whereas alliance may facilitate the development of therapist adherence, adherence may subsequently deepen and consolidate the client-therapist alliance.

In **Chapter 6** it was investigated how alliance and adherence develop during treatment and how this development is uniquely and jointly related to treatment outcomes up to 18-months post-treatment. For this purpose, a variable-centered (latent growth curve modeling) and a person-centered approach (latent class growth analysis) were used. A total of 848 adolescents and their caregivers having participated in MST were included. Outcomes were assessed at the end of the treatment and at 18-months post-treatment using the scale 'rule-breaking behavior' of the Child Behavior Checklist (CBCL) and two MST Ultimate Outcomes (i.e., police contact and out-of-home placement). Alliance and adherence showed an increasing and then flattening slope. We identified two trajectory classes for alliance and three classes for adherence, which were mainly characterized by different initial levels of alliance and adherence. Both alliance and adherence predicted treatment outcomes at the end of treatment, but not at 18 months post-treatment. The effects of alliance could not be replicated using the person-centered approach.

In **Chapter 7** we summarize and discuss the findings from these five studies. In the first place, we conclude that the Dutch TAM-R is reliable and does predict post-treatment outcome. Nevertheless, the Dutch adherence scores are affected by external sources (Aim 1), such as cultural response tendencies and the amount of experience a therapist or a country has with MST. This suggests that both bias, as well as true differences, may explain the observed lower adherence scores in the Netherlands compared to the US. We further conclude that alliance and adherence can reinforce one another, and that adherence is important for good outcomes (Aim 2). Yet, there needs to be a balance between adherence and adaptability. Treatment outcomes may improve if evidence-based interventions can be adapted to the unique needs of the new setting or individual client, but only if the core elements are retained. It is, therefore, essential to identify which elements of an intervention should be delivered in all situations, and which may be suitable for adaptation to new settings or may be flexibly used by clinicians.

Samenvatting

Behandelprogramma's bestaan doorgaans uit een set elementen, vastgelegd in een handleiding. Modeltrouw is de mate waarin een behandelaar deze elementen van het behandelprogramma uitvoert zoals bedoeld. Verondersteld wordt dat behandelprogramma's effectief zijn *vanwege* deze elementen en dat de handleiding dus trouw gevolgd moet worden om goede behandelresultaten te behalen. Zolang het programma nog onder toeziend oog van de ontwikkelaar uitgevoerd wordt, is dit doorgaans geen probleem. Lastiger wordt het, als het behandelprogramma op nieuwe locaties, soms zelfs in een nieuw land, geïmplementeerd moet worden. Onder nieuwe omstandigheden, uitgevoerd door andere behandelaren, blijkt het soms lastig om behandelresultaten te behalen die vergelijkbaar zijn met de oorspronkelijke resultaten. Ook de modeltrouw is vaak lager dan in de oorspronkelijke setting. Het eerste doel van dit proefschrift was een aantal factoren te onderzoeken die mogelijk van invloed zijn op modeltrouw na implementatie van een behandelprogramma in een nieuw land. Dit onderzoek wordt beschreven in de Hoofdstukken 2 en 3.

Onderzoek naar modeltrouw bij de implementatie van behandelprogramma's gaat ervan uit dat modeltrouw aan de elementen die specifiek zijn voor het behandelprogramma van belang is voor goede behandelresultaten. Sommige onderzoekers stellen echter dat niet de *specifieke* maar de *algemene* factoren van een behandeling belangrijk zijn voor het behalen van goede behandelresultaten. Algemene factoren zijn elementen die terugkomen in elke behandeling, zoals de werkrelatie tussen de behandelaar en de cliënt (de alliantie), of de verwachting en hoop van zowel de behandelaar als de cliënt dat de behandeling effectief zal zijn. Er is nog weinig bekend over het afzonderlijke en gecombineerde belang van alliantie en modeltrouw voor de totstandkoming van behandelresultaten. Het tweede doel van dit proefschrift was daarom te onderzoeken hoe modeltrouw en alliantie zich tot elkaar verhouden tijdens de behandeling en hoe ze samenhangen met behandeluitkomsten op de korte en lange termijn. Dit onderzoek wordt beschreven in de Hoofdstukken 3, 4 en 5.

Voor beide deelonderzoeken is gebruik gemaakt van de TAM-R (Therapist Adherence Measure-Revised), een vragenlijst die binnen Multisysteem Therapie (MST) gebruikt wordt om modeltrouw te meten. MST is een bewezen effectief behandelprogramma, ontwikkeld in de Verenigde Staten (VS) en in 2005 ingevoerd in Nederland. Het is een intensieve en ambulante systeembehandeling voor jongeren van 12 tot 18 jaar met ernstige gedragsproblemen. De TAM-R wordt routinematig verzameld als onderdeel van het kwaliteitsbewakingssysteem van MST met als doel te waarborgen dat MST correct wordt uitgevoerd. De vragenlijst bestaat uit 28 vragen op een 5-puntsschaal, die maandelijks telefonisch aan één van de ouders of verzorgers van de jongere voorgelegd worden.

In Hoofdstuk 2 is de Nederlandse TAM-R vergeleken met de Amerikaanse TAM-R om te bepalen of de vragenlijst in beide landen equivalent is. Met andere woorden: voert een Nederlandse behandelaar met eenzelfde score op de TAM-R als een Amerikaanse behandelaar de behandeling ook in dezelfde mate modeltrouw uit? In Studie 1 is een Rasch-analyse toegepast op 1.875 Nederlandse TAM-R vragenlijsten en de frequentieverdelingen van 1.875 Amerikaanse vragenlijsten. Hieruit bleek dat sommige vragen van de TAM-R 'moeilijker' waren in Nederland (Nederlandse behandelaren hadden gemiddeld genomen lagere scores op deze vragen dan Amerikaanse behandelaren) en dat andere vragen juist 'makkelijker' waren dan in de VS. Bovendien werden, over alle vragen heen, de hoogste antwoordcategorieën vaker gebruikt in de VS dan in Nederland. Dit betekent dat de vragenlijsten niet equivalent waren.

In een vervolgstudie werd nagegaan of een verbetering van de vertaling van de Nederlandse TAM-R deze verschillen tussen beide landen zou doen verdwijnen. Hiervoor werden 237 gezinnen willekeurig verdeeld over drie verschillende versies van de TAM-R, namelijk 1) de oorspronkelijke vertaling, 2) een nieuwe vertaling van de vragen en 3) een nieuwe vertaling van de vragen alsook de antwoordcategorieën. Uit dit vervolgonderzoek bleek dat de equivalentie niet verbeterde door de vertaling aan te passen. De vertaling werd wel als prettiger ervaren door ouders en was volgens MST-experts accurater. De blijvende verschillen tussen de Nederlandse en Amerikaanse scores zijn wellicht deels te wijten aan een cultuurverschil; Nederlandse gezinnen zijn minder snel geneigd een hoge score te geven dan Amerikaanse gezinnen. Daarnaast zijn de verschillen mogelijk een gevolg van het feit dat er in de VS meer ervaring is met MST, waardoor MST-therapeuten in de VS ook daadwerkelijk meer modeltrouw werken dan in Nederland.

Deze laatste hypothese werd onderzocht in Hoofdstuk 3 door te toetsen of méér ervaring met MST ook voorspellend was voor hogere TAM-R scores en vervolgens voor betere behandeluitkomsten. Hiervoor werd gekeken naar de ervaring van de behandelaar, de ervaring van het team, en de ervaring op landelijk niveau. De gegevens van alle MST-gezinnen die tijdens de eerste 10 jaar van MST in Nederland behandeld waren, werden voor dit onderzoek gebruikt. De steekproef bestond uit 4.290 gezinnen, die behandeld waren door 222 verschillende behandelaren, werkend in 27 teams. Dit onderzoek was een replicatie van Zweeds onderzoek naar de rol van ervaring binnen MST, uitgevoerd door Löfholm, Eichas en Sundell (2014). Het onderzoek in het voorliggende proefschrift vond dat zowel therapeutervaring als landelijke ervaring voorspellend waren voor modeltrouw en behandeluitkomsten aan het einde van de behandeling. Dit betekent dat therapeuten enige tijd nodig hebben om zich het behandelprogramma helemaal eigen te maken. De rol van landelijke ervaring doet vermoeden dat de landelijke kwaliteitscontrole, zoals trainingen en de continue ondersteuning van therapeuten, met de toegenomen ervaring beter werd. Dit zou betekenen dat zulke kwaliteitscontrole leidt tot betere modeltrouw en (hierdoor) betere behandeluitkomsten.

In Hoofdstuk 4 is onderzocht welke elementen de TAM-R precies meet, dat wil zeggen wat de onderliggende factor-structuur is, door middel van een 'Principal Component Analysis' (PCA) op de data van 580 gezinnen. Hieruit kwamen twee factoren naar voren, namelijk 'modeltrouw' en 'alliantie'. Deze twee factoren werden bevestigd met behulp van een 'Confirmatory Factor Analysis' (CFA) op een onafhankelijke steekproef van 723 gezinnen.

In Hoofdstuk 5 werden deze twee factoren gebruikt om te onderzoeken hoe modeltrouw en alliantie elkaar tijdens de behandeling onderling beïnvloeden. Hiervoor werd een 'cross-lagged panel analysis' uitgevoerd op een steekproef van 1.970 gezinnen. De resultaten toonden aan dat alliantie in de ene maand voorspellend was voor modeltrouw in de volgende maand. Dit betekent dat het ontwikkelen van een goede werkrelatie met de cliënt het vervolgens makkelijker maakt voor de behandelaar om het behandelprogramma modeltrouw uit te voeren. Modeltrouw in de ene maand was ook voorspellend voor alliantie in de volgende maand, maar alleen halverwege de behandeling. Het modeltrouw uitvoeren van de behandeling kan de werkrelatie dus verder versterken. Deze resultaten waren onafhankelijk van cliëntkenmerken.

In Hoofdstuk 6 is onderzocht hoe modeltrouw en alliantie zich tijdens de behandeling ontwikkelen en hoe ze gezamenlijk en los van elkaar de behandeluitkomsten voorspellen aan het einde van de behandeling en 18 maanden na afloop van de behandeling. Hiervoor werden twee verschillende analysemethodes gebruikt, namelijk latente groeicurve analyse en latente klasse groei analyse. De steekproef bestond uit 848 gezinnen. Zowel modeltrouw als alliantie werden gemiddeld genomen iets hoger in de beginfase van de behandeling en bleven de rest van de behandeling stabiel. Met de latente klasse groei analyse werden verschillende trajecten van alliantie en modeltrouw geïdentificeerd. Dit leidde tot twee verschillende trajecten voor alliantie en drie voor modeltrouw. Deze trajecten hadden een vergelijkbare ontwikkeling gedurende de behandeling, maar een verschillend startpunt. Zowel modeltrouw als alliantie voorspelden behandeluitkomsten aan het einde van de behandeling, maar niet 18 maanden na afloop van de behandeling. Modeltrouw was, in ons onderzoek, een robuustere voorspeller van de behandeluitkomsten dan alliantie.

In Hoofdstuk 7 worden de bevindingen van alle vijf studies uit dit proefschrift besproken. Allereerst concluderen we dat de Nederlandse TAM-R betrouwbaar is en behandeluitkomsten voorspelt. Desalniettemin worden de scores op de TAM-R beïnvloed door externe factoren zoals cultuurverschillen in responsstijl en de mate van ervaring met MST. Verschillen tussen Nederlandse en Amerikaanse scores zijn waarschijnlijk deels een gevolg van bias en deels een gevolg van daadwerkelijke verschillen in de mate waarin behandelaren in beide landen modeltrouw zijn.

Een andere belangrijke conclusie uit dit proefschrift is dat modeltrouw en alliantie elkaar kunnen versterken. Het is hierbij wel belangrijk op te merken dat zowel modeltrouw als alliantie in dit proefschrift door bijna alle gezinnen als hoog gescoord werden. Dit proefschrift

kan hierdoor geen uitspraken doen over de interactie tussen modeltrouw en alliantie als één van beide heel laag is. Verder blijkt dat modeltrouw bij MST van belang is voor het behalen van goede behandelresultaten. Modeltrouw zal naar alle waarschijnlijkheid echter niet bij elk behandelprogramma automatisch tot betere behandelresultaten en een sterkere alliantie leiden. Het lijkt van belang dat er een balans is tussen modeltrouw en flexibiliteit. Er zijn aanwijzingen dat behandelprogramma's tot betere uitkomsten leiden als het programma enigszins aangepast wordt aan de nieuwe setting. Ook binnen een behandeling met een cliënt lijkt het belangrijk aan te sluiten bij die unieke situatie. Desalniettemin kan niet zomaar elk element van een behandelprogramma aangepast worden. De *kern*elementen van een behandelprogramma moeten wel uitgevoerd worden zoals bedoeld. Meer kennis hierover is wenselijk, zodat ontwikkelaars en onderzoekers expliciet kunnen maken welke elementen van behandelprogramma's essentieel zijn voor de effectiviteit van het programma en welke elementen aangepast kunnen worden aan de situatie.

PhD Portfolio

Name PhD student	Aurelie M.C.M.J.G. Lange
Erasmus MC department	Medical Psychology & Psychotherapy
PhD period	2013-2017
Promotor(s)	Prof. dr. J.J. van Busschbach
	Prof. dr. R.H.J. Scholte
Supervisor	Dr. R.E.A. van der Rijken

PhD Training	Year	ECTS
General courses		
CPO Minicourse 'Methodologie van patiëntgebonden onderzoek en voorbereiding van subsidieaanvragen' (Graduate School, Erasmus MC)	2012	0.3
Research Integrity	2014	0.3
BROK ('Basiscursus Regelgeving Klinisch Onderzoek')	2014	1.5
Biomedical English Writing and Communication	2015	3
Specific courses		
Practical Rasch Measurement – Core Topics (online course statistics.com)	2011	2
Item Response Theory (Leiden University)	2011	0.3
Survival Analysis (NIHES, Erasmus Summer Programme)	2013	1.5
Advanced course on using Mplus (Utrecht Summer School)	2014	1.5
Advanced Topics in Causal Research (online course Graduate School of Life Sciences Utrecht University)	2017	1.5
Didactic courses		
Workshop 'omgaan met groepen' voor nieuwe tutoren	2015	0.2
Teach the Teacher 1	2016	0.5
Coachen van toekomstige Erasmusartsen, basis	2017	0.2
Seminars and workshops		
Studium Generale: Boos – Over agressie, opvoeding en ontwikkeling (Utrecht University)	2012	0.3
ISED Symposium: Special Issues in Research about Intellectual Disabilities (Groningen): Oral presentation	2012	1

Symposium Innovating Healthcare: An implementation science perspective (Trimbos Institute; Amsterdam, the Netherlands)	2017	0.3
Lock Symposium 'Uitkomsten wetenschappelijk onderzoek Kinderen uit de Knel	2017	0.2
Several presentations at de Viersprong and at the Erasmus Medical Centre	2011-2017	3

National Conferences

MST-MID Invitational Conference (Boxtel, the Netherlands): Oral presentation	2014	1
Met Het Oog op Behandeling 3 (Ermelo, the Netherlands): Oral presentation	2014	1
Jeugd in Onderzoek (Den Bosch, the Netherlands): Oral presentation	2016	1
Jeugd in Onderzoek (Den Bosch, the Netherlands): Leader network table	2017	1

International Conferences

1st European Multisystemic Therapy Conference (Oslo, Norway): Oral presentation	2012	1
2nd European Multisystemic Therapy Conference (London, United Kingdom): Oral presentation	2014	1
44th European Association for Behavioural and Cognitive Therapies Congress (EABCT; The Hague, the Netherlands): Oral presentation	2014	1
47th Annual Society for Psychotherapy Research Meeting (SPR; Jeruzalem, Israël): Oral presentation	2016	1
XIV European Scientific Association on Residential & Family Care for Children and Adolescents Conference (EUSARF; Oviedo, Spain): Oral presentation	2016	1

Reviewing Papers

Review for Psychotherapy Research	2016	0.3

Teaching Activities

	Year	ECTS
Tutoraat	2015	1.5
Tutoraat	2016	1.5
Tutoraat	2017	1.5
Coaching	2017-present	0.4

Parameters of Esteem	Year
Grants	
ZonMw 'Effectief werken in de Jeugdsector: Algemeen werkzame factoren' (729101006): Co-applicant	2014

Publications

Publications in this thesis

Lange, A.M.C., van der Rijken, R.E.A., Busschbach, J.J.V., Delsing, M.J.M.H., & Scholte, R.H.J.. Factorial structure of the Therapist Adherence Measure-Revised (TAM-R) within Multisystemic therapy. *Submitted*

Lange, A.M.C., van der Rijken, R.E.A., Busschbach, J.J.V., delsing, M.J.M.H., & Scholte, R.H.J. Development of alliance and therapist adherence in relation to treatment outcomes of adolescents with behavioral problems: A variable-centered and person-centered approach. *Submitted***.**

Lange, A.M.C., van der Rijken, R.E.A., Busschbach, J.J.V., Delsing, M.J.M.H., & Scholte, R.H.J. (2017). It's not just the therapist: country-wide experience also predicts therapist adherence and adolescent outcome. *Child and Youth Care Forum, 46*, 455-471**.**

Lange, A.M.C., van der Rijken, R.E.A., Delsing, M.J.M.H., Busschbach, J.J.V., van Horn, J.E., & Scholte, R.H.J. (2016). Alliance and adherence in a systemic therapy for antisocial adolescents. *Child and Adolescent Mental Health, published online***.** DOI: 10.1111/camh.12172.

Lange, A.M.C., Busschbach, J.J.V., van Geffen, W., Timman, R., Scholte, R.H.J., & van der Rijken, R.E.A. (2015). The lack of cross-national equivalence of a therapist adherence measure (TAM-R) in Multisystemic Therapy (MST). *European Journal of Psychological Assessment, 32*, 312-325.

Publications not in this thesis

Lange, A.M.C., van der Rijken, R.E.A., Delsing, M.J.M.H., Busschbach, J.J.V., Van Horn, J.E., & Scholte, R.H.J. Alliantie en modeltrouw kunnen elkaar versterken. *In preparation*

van Logten, A., **Lange, A.**, Cima, M., & van der Rijken, R. (2017). Functionele Gezinstherapie en Multi Systeem Therapie: Een onderzoek naar de stabiliteit van doelgroepen en behandelresultaten in Nederland. *Tijdschrift voor Orthopedagogiek, 55*, 51-64.

Hendriks, M., **Lange, A.**, Boonstoppel-Boender, M., & van der Rijken, R. (2014). Functional Family Therapy en Multi Systeem Therapie: Een vergelijking van doelgroepen. *Orthopedagogiek: Onderzoek & Praktijk, 53*, 355-366.

Lange, A. Vuyst de, K., Rijken van der, R. (2014) Multi systeem therapie voor mensen met een LVB: een pilotonderzoek. *Onderzoek & Praktijk*, 1, 15-22.

Lange, A., De Vuyst, K., De Bruijn, J., & Van der Rijken, R. (2013). Multisysteem therapie voor mensen met een LVB: Een pilot studie. In: R. Didden & X. Moonen (Red.), *Met het oog op behandeling 3: Diagnostiek en behandeling van gedragsproblematiek bij mensen met een licht verstandelijke beperking* (pp. 61-67). Amersfoort: Bergdrukkerij.

Rieffe, C., Camodeca, M., Pouw, L.B.C., **Lange, A.M.C.**, & Stockmann, L. (2012). Don't anger me! Bullying, victimization, and emotion dysregulation in young adolescents with ASD. *European Journal of Developmental Psychology, 9 (3)*, 351-370.

Curriculum Vitae

Aurelie Lange was born in Berlin on the 16th of January 1987 and completed high school at the Stedelijk Gymnasium Leiden in 2005. Although she had always wanted to be a primary school teacher, she decided to start with a bachelor in Pedagogical Sciences in Utrecht. She conducted her bachelor thesis in South Africa, where she worked in a centre for children with severe mental and physical disabilities and conducted a baseline assessment before introduction of a training programme to increase the independence of the children in daily activities. During her stay in South Africa she discovered how much she enjoyed doing research, which led her to decide to do a research master in Psychological Research Methods in Notthingham (UK). Back in Utrecht she started to work as a research assistant for the research group 'Focus on Emotions' at the University of Leiden, where she also assisted students writing their bachelor theses. She switched to de Viersprong, centre for Personality, behaviour and family, in 2011. Working as a junior researcher for the research team on Multisystemic Therapy (MST), she was involved in a project to investigate whether MST could be adapted to suit the needs of adolescents with externalizing behavioural problems in combination with an intellectual disability (MST-ID), as well as a study on therapist adherence within MST. The latter project is described in this thesis and became the main focus of her work from 2013 on, when she was employed at the Erasmus Medical Centre as a PhD student. In 2015 she also started to do some teaching activities at the Erasmus Medical Centre. In 2017 she had the opportunity to become the principal researcher of the research team at de Viersprong, which had now become the research team on Systemic Interventions. In this role, she is involved in several projects regarding the effectiveness of established and new systemic programs at de Viersprong.

Dankwoord

Een proefschrift schrijven betekent soms dat je dagen achter elkaar in je eentje achter je laptop zit. Als ik menig behandelaar binnen de Viersprong mag geloven, lijkt onderzoek doen een erg eenzame aangelegenheid te zijn. Desondanks was het schrijven van dit proefschrift dat allerminst. Het was een ontzettend boeiend en leerzaam proces, waar ik grotendeels veel van genoten hebben, maar wat ik allerminst alleen gedaan heb. Ik wil hier dan ook graag een aantal mensen bedanken.

Allereerst zou ik graag alle deelnemende instellingen willen bedanken voor het beschikbaar stellen van hun data, alsook de bellers van Kwestion voor het verzamelen van de TAM-R en de follow-up interviews. Ook bedankt aan alle ouders en verzorgers, dat zij, ondanks alle perikelen in hun persoonlijke leven, de tijd namen om de interviewvragen te beantwoorden.

Daarnaast wil ik graag mijn promotoren bedanken. Prof. Dr. van Busschbach, beste Jan, bedankt voor je kritische blik bij het meelezen van de artikelen. Jouw vragen over de tekst hielpen mij de artikelen aan te scherpen. Ik ben je dankbaar voor de mogelijkheid die je mij geboden hebt in dienst te komen van het Erasmus MC. Daardoor kreeg ik ook de kans studenten te coachen en te begeleiden in hun eerste jaar; een waardevolle en inspirerende aanvulling op mijn onderzoekswerkzaamheden.

Prof. Dr. Scholte, beste Ron, wat ontzettend fijn dat jij onze onderzoekslijn bent komen versterken. Je bent altijd enorm betrokken geweest en hebt een hele brede kennis van het vakgebied. Dit heeft echt voor verdieping gezorgd en daar ben ik je dankbaar voor.

Lieve Rachel, zonder jou was dit proefschrift er nooit gekomen. Dank je voor je vertrouwen in mij, toen je me aannam als junior onderzoeker, en voor je vertrouwen, tomeloze inzet, en je eindeloze bereikbaarheid en luisterend oor de jaren daarna. Dank je dat ik met al mijn vragen bij je terecht kon, en dat je helder kon blijven denken en me bij tijd en wijle een spiegel voorhield in de wirwar van keuzes.

Voor de analyses ben ik Marc Delsing en Reinier Timman dankbaar. Heel fijn dat ik altijd bij jullie terecht kon voor vragen of advies. Dr. Eichas and Dr. Löfholm, many thanks for your help in replicating your Swedish study on the role of experience on therapist adherence and outcome. I really appreciate that you were willing to answer all my questions.

Ik wil ook graag Roel Verheul, het management van de voormalige Unit Systeem-interventies en ZonMw bedanken voor het financieel mogelijk maken van dit promotietraject. Dank je Kim, dat je besloot ruimte te maken voor onderzoek binnen de Unit. Het is fantastisch om iemand op de werkvloer te hebben die zo voor het belang van onderzoek wil staan. Beste Roel, ik waardeer ik het enorm dat er ook in financieel zware jaren ruimte en steun was voor onderzoek binnen de Viersprong. Dat maakt de Viersprong een bijzondere plek om te werken.

MST-Nederland, oftewel Wim, Femke, Jo, Hans, Mai, Miranda, Diana en Lidewy, bedankt voor jullie hulp in het hele proces, zoals het faciliteren van de dataverzameling en het inzetten van jullie kennis tijdens klankbordgroep- of expertbijeenkomsten, of het lezen van de artikelen. Mijn dank gaat ook uit naar alle MST therapeuten en supervisoren die tijdens klankbordgroepbijeenkomsten, of gewoon tijdens de lunch, met mij meedachten of voor inspiratie zorgden.

I would also like to thank MST Services and the MST Institute, specifically Sonja Schoenwald, Molly Brunk, Kristen Dodd, and Lisa Reiter. Thank you for making US-data available, for your feedback on some of the papers as well as on the research setup for the new translation of the TAM-R. Your input has been most valuable.

Een speciale dank gaat uit naar Prof. Dr. Overbeek, Prof. Dr. Asscher, en Prof. Dr. Mulder voor het lezen en beoordelen van mijn proefschrift. Beste Dr. Boendermaker, Prof. Dr. Hox en Dr. van Hattum, bedankt dat u deel wilt nemen aan de oppositie. Ik kijk ernaar uit om met u allen in gesprek te gaan over mijn proefschrift op 31 januari 2018.

Lieve huidige en oud-collega's van 'mijn' onderzoeksteam Systeeminterventies, maar ook van de VISPD, de (toenmalige) unit systeminterventies en de locatie Den Bosch: Bedankt voor jullie gezelligheid, interesse, vragen en ideeën. Het is onder andere vanwege jullie dat ik al een heel aantal jaren met veel plezier bij de Viersprong werk. Collega's van MPP, hoewel ik beduidend minder in Rotterdam aanwezig was, werd ik desondanks altijd warm welkom geheten. Bedankt voor jullie gastvrijheid.

Lieve Ilona, Ines en Mindy. Bedankt dat jullie er zijn. Wat is het fijn om altijd bij jullie terecht te kunnen, even te kunnen 'luchten' als het iets minder gaat, maar vooral ook fijne tijd samen door te kunnen brengen. Speciaal vandaag, dank dat jullie mijn paranimfen (op afstand) willen zijn. Ik ben blij dat ik zulke lieve vriendinnen om mij heen heb.

Een goede manier om even mijn gedachten uit te zetten, waren de wekelijkse danslessen, de bijkomende sociale activiteiten en de shows. Lieve vrienden van Jumping Spectacles, bedankt dat ik al zoveel jaren elke week weer met jullie kan dansen, lachen, kletsen en gave (nieuwe) acro's kan uitproberen.

Lieve Yannick, bedankt voor het ontwerpen van de kaft van mijn proefschrift. Ik weet dat jij je hiermee niet helemaal in je comfort zone voelde, maar ik ben er desondanks erg blij mee! Cynthia, Glenn, Guillaume, Pauline, Janna en Benjamin, het is ontzettend fijn om jullie dichtbij te weten. Lieve Anne en Christine, bedankt voor jullie liefde en eeuwige bereidheid tot hulp. Oppassen, klusjes in huis, advies… Jullie zijn een fijne en onmisbare bron van steun geweest afgelopen jaren.

Bien cher papa et maman, votre amour et confiance en moi m'on donné tout dont j'avais besoin pour réussir cette épreuve. J'ose voler parce-que vous êtes la pour moi. Je ne peux pas vous remercier assez pour être qui vous êtes.

Lieve, lieve Daan, bedankt voor alles. Voor het aanhoren van al mijn werkverhalen en het meedenken over oplossingen (zelfs over statistiek of de opbouw van mijn inleiding), voor het koken, de papadagen, de was, de boodschappen… Maar vooral, dank je dat je van me houdt, vertrouwen in me hebt, me steunt en dat ik altijd bij je mag wegkruipen als het even teveel wordt.

Allerliefste Sam en Jona, jullie komst was het mooiste wat er is. Bedankt dat jullie me er dagelijks aan herinneren dat er belangrijkere dingen zijn dan mijn werk, dat jullie me laten zien hoe je met verwondering naar de wereld kunt kijken, dat geluk in de kleine dingen zit, en dat er niets belangrijker is dan goed te zorgen voor de wereld en alles wat daarop leeft, zodat ook jullie, en de generaties na jullie, van onze aarde kunnen blijven genieten.