The factor structure of the Working Alliance Inventory Short-form in youth psychotherapy: an empirical investigation

Authors:

Antonella Cirasola[1,2]*, Nick Midgley[1,2], Peter Fonagy[1,3], IMPACT Consortium[4], and Peter Martin[2,5]

[1]*Research Department of Clinical, Educational and Health Psychology, University College London, London, UK*

[2]*Child Attachment and Psychological Therapies Research Unit (ChAPTRe), Anna Freud National Centre for Children and Families, London, UK*

[3]*Anna Freud National Centre for Children and Families, London, UK*

[4]*Department of Psychiatry, University of Cambridge, Cambridge, UK*

[5]*Department of Applied Health Research, University College London, London, UK*

*Corresponding author: Antonella Cirasola; Email: Antonella.cirasola@annafreud.org; Postal address: Anna Freud National Centre for Children and Families 4-8 Rodney Street London N1 9JH

Clinical or Methodological Significance of this Article: This paper addresses the methodological issues on the operationalization of the alliance in youth psychotherapy, by assessing the factor structure of one of the most popular measures in the field. Progress achieved in research regarding the methods used to measure the alliance is crucial for a deeper understanding of nature and role of the alliance in the treatment of young people.

Abstract

Objective: The Working Alliance Inventory short form (WAI-S) is one of the most commonly used alliance measures with adolescents. Yet, its factor structure has received minimal attention in the youth alliance literature. This study investigated the factor structure of the WAI-S in psychotherapy for adolescent depression and explored its measurement invariance across time, therapeutic approaches and patients' and therapists' perspectives. The existence of method effects associated with the negatively worded items of the scale was also assessed. **Method:** The setting of this study is the IMPACT trial, a randomised controlled trial assessing the effects of three therapeutic interventions in the treatment of adolescent depression. The WAI-S was completed at 6, 12 and 36 weeks after randomisation by 338 adolescents and 159 therapists. Data were analysed using confirmatory factor analysis. **Results:** The hypothesised Bond-Task-Goal alliance structure was not supported and a general, one-factor model was found to be more psychometrically valid. The existence of a method effect and measurement invariance across time and treatment arms were also found. **Conclusions:** While the distinction between the specific alliance dimensions is conceptually and clinically interesting, at an empirical level the alliance features of the WAI-S in youth psychotherapy remain strongly intercorrelated.

Keywords. Alliance; youth psychotherapy; Working-Alliance-Inventory; confirmatory factor analysis; method effect; measurement invariance.

The therapeutic alliance is at the heart of clinical work with both adults and youth, and there is a long history of empirical research exploring the concept (Flückiger, Del, Wampold, & Horvath, 2018; Karver, De Nadai, Monahan, & Shirk, 2018). Yet there is no consensually accepted definition of the alliance and there have been few attempts to differentiate its content

and structure in an empirically meaningful way (Horvath, 2018; McLeod, 2011). Despite the lack of consensus about its features, the alliance has been mainly described as a multidimensional phenomenon (Bordin, 1979; DiGiuseppe, Linscott, & Jilton, 1996; Freud, 1946; Shirk & Saiz, 1992). Consistent with the adult literature, Bordin's (1979) definition of the alliance, as the agreement between patient and therapist on therapy goals and tasks in the context of an emotional bond, is one of the most widely used with younger patients too. This definition involves three interrelated dimensions: 1) agreement on goals relates to a shared understanding of the changes that the therapeutic process aims towards; 2) agreement on tasks relates to a shared understanding of the activities necessary to meet such goals; and 3) the bond refers 'to the nature of the human relationship between therapist and patient' (Bordin, 1979, p. 254).

Despite its popularity, it remains uncertain whether Bordin's three-dimensional model adequately captures the construct of the alliance with young people. The alliance structure and the relative importance of its dimensions might in fact be different with this age group. From a theoretical perspective, more has been written about the affective/relational dimension (i.e. bond) than any other alliance dimensions with youth (Freud, 1946; Karver et al., 2008; Shirk & Saiz, 1992; Zack, Castonguay, & Boswell, 2007). The emphasis on the emotional connection between patient and therapist is based on the assumption that a positive bond is an essential prerequisite to foster young people's participation in the therapeutic work. Furthermore, important developmental considerations concern the task and especially the goal dimension of the alliance with youths. Firstly, a variety of cognitive skills is necessary to formulate long term-therapeutic goals and to elaborate the link between such broad, sometimes abstract, goals and the specific tasks of therapy (Shirk, 2013; Zack et al., 2007). Accordingly, it has been argued that young people might find it difficult to discriminate between therapeutic goals and tasks (Zack et al., 2007). Besides, developing an agreement on goals may be particularly

challenging in an age group that is often referred to treatment by others, may not see the problem as being 'in' them, and whose developmental needs for independence may interfere with the establishment of a collaborative relationship with an adult (Kazdin, 2003). Some have, however, criticized the exclusive emphasis on the bond for failing to recognise the importance of the more contractual features of the alliance, especially in the treatment of older children and adolescents (DiGiuseppe et al., 1996; Sandler, Kennedy, & Tyson, 1980).

Developmental issues should be taken into account when assessing the alliance with young people (Norcross, 2011); yet in most cases the operationalisation of the construct in youth psychotherapy has been directly imported or only mildly revised from adult core alliance scales. While the alliance is considered a multidimensional concept from a theoretical point of view, empirical evidence has thus far failed to fully support this multidimensionality in adolescent samples (Diamond, Hogue, Liddle, & Dakof, 1999; Faw, Hogue, Johnson, Diamond, & Liddle, 2005; Fjermestad et al., 2012; Hogue, Dauber, Stambaugh, Cecero, & Liddle, 2006; Meyer et al., 2002; Shelef & Diamond, 2008; Shelef, Diamond, Diamond, & Liddle, 2005). Research on the characteristics of the alliance with young people is underdeveloped and little is known about the factor structure and measurement invariance of alliance scales for this age group (Elvins & Green, 2008; Karver et al., 2018; McLeod, 2011).

The Working Alliance Inventory

Among the most frequently used measures of the alliance in adult and youth psychotherapy are the Working Alliance Inventory (WAI; Horvath & Greenberg, 1989) and its short form (WAI-S; Tracey & Kokotovic, 1989). The WAI was originally designed for adult therapy and then adapted for use with young people (DiGiuseppe et al., 1996, Figueiredo, Dias, Lima, & Lamela, 2016). Nevertheless, the original, and especially the short version of the WAI (WAI-S; Tracey & Kokotovic, 1989) are the most commonly chosen instrument for measuring youth alliance (Karver et al., 2018; McLeod, 2011; Shirk, Karver, & Brown, 2011). The WAI was developed

from a theory-based approach with the aim to measure Bordin's (1979) alliance model. Accordingly, it includes three subscales assessing each of the hypothesised dimensions of goal, task and bond. While the correlation between the subscales was shown to be high in the original measure development study (especially between Goal and Task), no use was made of factor analysis methods in the development of the subscales (Horvath & Greenberg, 1989). Subsequent research attempted to explore whether the theoretical dimensions of the alliance were empirically supported but yielded mixed results.

Tracey and Kokotovic (1989) were the first to examine the construct validity and factor structure of the WAI. They conducted confirmatory factor analysis (CFA) on 84 adult patients' and 123 therapists' ratings of the WAI after the first session of counselling. Three alliance models were tested: 1) a model with only one general alliance factor, 2) Bordin's three-factor model, and 3) a bi-factor model with one common factor for all items and three separate factors for the three subscales. The authors found that the bi-factor structure fit the data best for both patients' and therapists' ratings. Based on this first CFA, Tracey and Kokotovic (1989) selected the four highest loading items on each of the three dimensions (Task, Goals and Bond) to develop a short form of the WAI (the WAI-S). A second set of CFAs was then conducted on these 12 items, which also supported the bi-factor structure of the WAI-S. However, methodological issues might have influenced these finding. For instance, the sample size was small and the fit indexes for the bi-factor solution were not within the currently accepted ranges. Further, the method of extracting the items from the original CFA to form the WAI-S should ideally have been validated in a different sample.

Subsequent research failed to confirm the hypothesised three-factor structure of the WAI. In the adult alliance literature, a few studies using exploratory factor analytic techniques found two rather than three factors, with one factor including Goal and Task items together, and the other factor comprising Bond items (Andrusyna, Tang, DeRubeis, & Luborsky, 2001; Hatcher

et al., 1996). CFA studies also failed to support the three-factor structure of the WAI and WAI-S in adult psychotherapy (Corbière, Bisson, Lauzon, & Ricard, 2006; Falkenström, Hatcher, & Holmqvist, 2015; Hatcher & Gillaspy, 2006).

Less research is available in youth psychotherapy. To our knowledge, only three studies have assessed the factor structure of different version of the WAI with young people and none of them supported the hypothesised Goal-Task-Bond alliance model (Anderson et al., 2012; G. S. Diamond et al., 2006; DiGiuseppe et al., 1996). DiGiuseppe and colleagues (1996) were the first to explore the factor structure of WAI in youth psychotherapy. Their findings were mixed: while patients' ratings of the adolescent version of the scale (WAI- A) resulted in a single, large alliance factor; therapists' ratings of the WAI yielded one general factor and the three separate factors of goal, task and bond. Similarly, a principal component analysis of adolescents' and therapists' ratings of the original version of the WAI yielded a one-factor solution for both adolescent and therapist perspectives, in the context of specific treatments for adolescents with cannabis dependence (Dennis et al., 2002; Diamond et al., 2006). A general one-factor model was also found in a CFA study assessing the structure of an online adaptation of the WAI-S in therapist-assisted online cognitive behavioural therapy for youth anxiety, where a single-factor alliance was found for the adolescent ratings; while the parents' ratings were explained by a two-factor model (i.e. bond and combined task and goal) (Anderson et al., 2012).

Overall, empirical evidence seems to suggest that the WAI measures a general, one-factor alliance construct when rated by young people. This is in line with the literature on the factor structure of a range of youth alliance measures, which also found a single, general alliance factor (Diamond et al., 1999; Faw et al., 2005; Fjermestad et al., 2012; Hogue et al., 2006; Meyer et al., 2002; Shelef & Diamond, 2008; Shelef et al., 2005). However, caution should be taken when interpreting these findings due to some common methodological limitations.

The majority of the studies on the factor structure of the WAI with adolescents had a relatively small sample size and their generalisability was limited due to the inclusion of specific target groups (e.g., hard-to-treat samples of primarily male, substance-abusing adolescents) or type of treatment (e.g. online therapy). This might result in factors that are specific to one data set, difficult to replicate, and not necessarily representative of a larger population (Lingard & Rowlinson, 2005). Additionally, measurement invariance has been neglected in youth alliance research. Measurement invariance, a statistical property of a measurement that indicates that the same underlying construct is being evaluated across groups or time, is a crucial prerequisite for valid comparisons of test scores over time or between different groups. This is particularly relevant since the WAI-S has increasingly been used in longitudinal research as well as across different types of treatment, rater perspectives and client group. Yet, to our knowledge no research has investigated whether the WAI-S structure differs across time, various psychotherapy modalities or raters in youth psychotherapy.

Another critical aspect is related to the alliance measures themselves and concerns the item wording. Although it is common for alliance measures, like the WAI, to include items worded in opposite direction, this might result in response bias, difficulties in processing reverse-worded items, and therefore possible method effects; which can affect the results of factor analytic studies, creating polarities (DeVellis, 2016). There is, however, lack of research on the impact of item wording on the latent structure of the WAI-S in youth psychotherapy.

In the context of these limitations and existing gaps, recent major reviews of the empirical literature on the alliance have called for further investigation of the dimensionality of the alliance measures (Horvath, 2011; Karver et al., 2018). If alliance scales should help to better understand the nature and role of the alliance in youth psychotherapy, it is essential to identify if they measure different features and, if so, to determine if particular aspects of the alliance are more associated with outcomes than others.

The Current Study

Despite the WAI-S (Tracey & Kokotovic, 1989) being the most used alliance measure with adolescents (Karver et al., 2018; McLeod, 2011; Shirk et al., 2011), little research has been conducted on its structure with this age group. The aim of this study is to address this gap by investigating the factor structure of the WAI-S and its measurement invariance in youth psychotherapy. This study has the following specific aims:

1) To investigate whether the hypothesised Bond-Task-Goal alliance model of the WAI-S is empirically supported, or whether a different alliance structure represents a better fit in three types of time-limited psychotherapy for adolescent depression.

2) To evaluate whether the alliance structure is invariant a) over the course of therapy; b) across rater groups (adolescents and therapists); and c) across different therapeutic approaches.

3) To examine the existence of method effects associated with the negatively worded items of the WAI-S.

Method:

Participants

The setting for this study is the Improving Mood with Psychoanalytic and Cognitive Therapies (IMPACT) trial, the first multicentre, pragmatic, RCT assessing the medium-term effects of three therapeutic interventions in the treatment of adolescent depression (Goodyer et al., 2017, 2011). 465 adolescents (aged between 11 and 17 years) with diagnosis of depression were randomised to receive either cognitive behavioural therapy (CBT), short-term psychoanalytic psychotherapy (STPP) or brief psychosocial intervention (BPI). Full details of the procedure of the IMPACT study are reported in Goodyer and colleagues (2011, 2017). The present study included only participants who received treatment and had at least one rating of the alliance

completed by the adolescent or their therapist. Sample sizes vary between different analyses due to missing values.

WAI-S sample. This sample consisted of 338 adolescents, i.e. all participants who completed one or more WAI-S over time. The adolescent in this sample were treated by 157 therapists. The median and mode of number of patients treated by each therapist was 1, with only a few therapists treating more than one patient. Specifically, 64% of therapists had only 1 patient, 19% had 2 patients and 18% had 3 patients or more.

WAI-S-T sample. This sample consisted of 159 adolescents with at least one rating of the alliance completed by 72 therapists. The median and mode of number of patients treated by each therapist was 1: 61% of therapists had only 1 patient, 15% had 2 patients and 24 % had 3 patients or more.

Demographic information for the adolescents in both samples is displayed in Table 1. No demographic information was collected for the therapists.

[Table 1 near here]

Measures

Demographics. Age, sex, and ethnicity were assessed with a demographic questionnaire at baseline.

Therapeutic alliance. The therapist and patient short-version of the WAI-S (Horvath & Greenberg, 1989; Tracey & Kokotovic, 1989) were completed at 6, 12 and 36 weeks after randomisation. The WAI-S aims to measure Bordin's (1979) conceptualization of the working alliance. Accordingly, it includes three subscales assessing: (a) agreement on goals, (b) agreement on tasks and (c) the emotional bond between patient and therapist. Both the WAI-S and its therapist version (WAI-S-T) consist of 12 items (10 positively worded and 2 negatively worded), 4 in each subscale, rated on a 7-point response scale (from 1=Never to 7=Always). The scale yields different scores for each subscale as well as an aggregate overall score, with

higher scores reflecting a stronger working alliance. The WAI-S demonstrated good validity and reliability (Horvath & Greenberg, 1989, Tracey & Kokotovic, 1989) as well as good internal consistency within youth samples (Capaldi, Asnaani, Zandberg, Carpenter, & Foa, 2016; Dennis et al., 2002; Hawley & Garland, 2008; Tetzlaff et al., 2005).

Statistical Analyses

WAI-S factor structure. CFA estimated using maximum likelihood estimation with robust standard errors (MLR) and a Satorra-Bentler (S-B) scaled test statistic was used to investigate the factor structure of the WAI-S. Four alliance models were tested and compared to each other: 1) a three-factor model with the hypothesised three correlated subscales of Task, Goal and Bond (Horvath & Greenberg, 1989); 2) a one-factor model with all items loading into a general alliance factor; 3) a two-factor model with Collaboration (Goal and Task items combined) as one factor, and the Bond items as the other correlated factor (Shirk & Saiz, 1992), and 4) a bi-factor model with one general alliance factor at one level, and the three subfactors of Goal, Task, and Bond (Tracey & Kokotovic, 1989). In the bifactor model, all correlations between latent variables were constrained to be zero. The hypothesised models are described graphically in Figure 1 to 4 in the supplementary material.

Following guidelines from Kenny (2015) and Hu and Bentler (1999), to assess model fit we evaluated fit statistics from different categories, including the root mean square error of approximation (RMSEA), the standard root mean square residual (SRMR) as well as two incremental fit measures: the comparative fit index (CFI) and the Tucker-Lewis index (TLI). According to the most used conservative criteria, CFI and TLI values between 0.90 and 0.95 were considered an indication of acceptable fit, and values above 0.95 indicated good fit (Hu & Bentler, 1999; Kenny, 2015). SRMR values less than 0.08 were considered an indication of a good fit, as RMSEA values below .06; though RMSEA values between 0.06 and 0.08 were deemed acceptable, values in the 0.08 to 0.10 range were considered marginal fit, and values

> 0.10 poor fit (Hu & Bentler, 1999; Kenny, 2015). A model was determined to be well-fitting if at least three of these four indices demonstrated good fit. If there were problems with the model estimation, including model non-convergence, correlations between the latent variables over 1.0 (i.e., 'out of bound'), negative measurement error variances or invalid values for path estimates, the factor solution was considered 'improper' (Kyriazos, 2018). No post-hoc modifications were performed to improve model fit. Owing to the non-nested nature of the different hypothesised models, a statistical test of model comparison was not available and two model-fit criteria, the Akaike information criterion (AIC) and the Expected cross-validation index (ECVI), were used to compare the quality of models. Smaller values on AIC and ECVI indicate better fit.

The amount of missing WAI-S items was very low across samples and ranged from 0 to 1.7%. Missing data were assumed to be missing at random (MAR) (Rubin, 1987) and were handled using full information maximum likelihood (FIML). In FIML procedure, missing values are not imputed, but coefficients (such as loadings and variances) are estimated using all available data rather than complete cases only.

Nesting within therapists. Although in both samples patients were nested within therapists, the majority of therapists treated only one patient. Therefore, we expected this statistical dependency not to be high. Nevertheless, we attempted multilevel CFA to estimate both within- and between-person variation in the assessment of the alliance structure. This failed to converge in both samples likely because the median and mode number of cases per therapist was 1, limiting distinction between case and therapist.

Method effect. To assess the performance of the negatively worded items of the WAI-S, a residual method effect was investigated by allowing the errors of the two negatively worded items to be correlated in the CFA analyses. Each alliance model was tested twice: the first time with uncorrelated errors, the second time the model specification included the

correlation between the error of the two negatively worded items (e.g. accounted for the method effect). These two nested models were then compared against each other using the chi-square difference test.

Measurement Invariance. Once the best fitting model(s) of the WAI-S was established in both samples at the first time-point assessment (6 weeks), we conducted a series of CFAs for each time point and for each treatment arm separately to assess configural invariance. Configural invariance refers to a qualitatively invariant measurement pattern of latent constructs across groups and/or over time. If the same measurement model had the best fit in all groups, as well as having at least adequate fit (in terms of normative fit indices), we considered this as an indication of configural invariance. Subsequently, to test for metric and scalar measurement invariance across raters and treatment arms we conducted multiple groups CFA with MLR on the best fitting model(s) using a series of increasingly stringent model comparisons: configural (i.e. with no constraints) to metric (i.e. with factor loadings constrained to be equivalent across the groups) to scalar model (i.e. with both factor loadings and item intercepts constrained to be equivalent across the groups). This was done to assess whether constraining specified model parameters across groups resulted in a significant improvement or worsening of model fit. To assess longitudinal measurement invariance, we specified a longitudinal structural equation model for which we then progressively add the invariance constraints by successively setting the equality of the parameters of the measurement model across time points. Correlations among residuals for the same items at different time-points were estimated freely.

Change in model fit was evaluated by differences in CFI and S-B scale-corrected chi-square difference tests. Following the most used guidelines, a difference in CFI (ΔCFI) less than .01 was considered indicative of no meaningful difference in model fit and, therefore, indicative

of measurement invariance (Cheung & Rensvold, 2002). Since chi-square tests are sensitive to sample size, ΔCFI > 0.01 was our primary indication of violation of measurement invariance. All sets of analyses were performed separately on the adolescents' and therapists' ratings of the WAI-S. Descriptive analyses were conducted using SPSS (IBM, 2016). All other analyses were conducted in R (R Core Team, 2018) using the Lavaan package (Rosseel, 2012).

Results

Descriptive Statistics

The number of WAI-S completed by adolescents was 223 at 6 weeks, 247 at 12 weeks and 222 at 36 weeks. Therapists completed 139, 119 and 63 WAI-S-T at 6, 12 and 36 weeks, respectively. Descriptive statistics and correlations matrix of WAI-S and WAI-S-T items for both samples at all time points can be found in the supplementary Table 2-7. Since the negatively worded items were reversed, all correlations were positive. In both samples, but especially in the WAI-S, the correlations involving the negatively worded items (item 4 and 10) were lower compared to those resulting from the associations of the positively worded items. This raised questions about the performance of the negatively worded items, which was further assessed using CFA.

Information about the number of therapists and the patients treated by each therapist in both samples are provided in the Supplementary Table 8.

Factor Structure of WAI-S

Table 2 provides an overview of the results from the CFAs conducted on the adolescents' ratings of the WAI-S at 6 weeks. Of all models tested, the two-factor had the best fit, with all fit indices within the threshold for good model fit, and the lowest values in both comparative indices (AIC and ECVI). The fit indices for the one-factor solution were almost as good, ranging between acceptable to good fit. Despite its better fit, the two-factor model estimated a high correlation between its two factors (r=0.91), which questions whether the WAI-S does

meaningfully differentiate between Bond and Collaboration. In contrast, both the three-factor and the bi-factor models could not be reliably estimated in this sample and therefore are not reported. In particular, the three-factor solution, despite converging normally, it was considered unreliable due to out of bound correlation (r>1) between the Goal and Task scales. A correlation of >1 indicates the presence of problems with the model - often referred to as Heywood case-, which renders the factor solution invalid or 'improper' (Kenny, 2015; Kyriazos, 2018). The bi-factor model failed in this sample and at all time points due to identification problems given the covariance matrix could not be inverted. When a model is not identified, that generally means it is too complex given the amount of information in the covariance matrix and it is advisable to reduce the complexity of the model or to increase the number of items (Hartman; 2017). Information about factor loadings for the models tested are provided in Supplementary Table 9.

[Table 2 near here]

Factor Structure of WAI-S-T

The bi-factor model could not be identified in the WAI-S-T sample too due to convergence problems. All other models showed an overall acceptable fit, with only minor differences in fit between models (see Table 2). Although the differences between the models were small, the two-factor solution had the lowest values for both the AIC and ECVI comparative indices, demonstrating a better fit compared to all rivalling models. The latent variable correlations were high in both multifactor models, especially in the three-factor solution where the correlation between the Goal and Task scales was 0.98, and those between the Bond scale and the Goal and Task scales were both 0.96. In the two-factor model, the correlation between Bond and Collaboration was also high (r=0.96). Information about factor loadings for all models are provided in Supplementary Table 10.

CFA results of the best fitting models (e.g. the one-factor and two-factor models) in both samples at all time points are reported in the configural analyses section; all sets of CFA testes can be obtained from the authors.

Method Effect

As shown in the Table 2 (last row for each group), chi-square difference tests suggested that in both samples each model that accounted for the method effect showed a significantly better fit compared with the equivalent model with uncorrelated errors. This indicated the existence of a method effect associated with the negative item phrasing on the WAI-S. To account for this issue all models tested with CFA included the correlation between the error of the two negatively worded items in their model specification.

Longitudinal Measurement Invariance

Table 3 and 4 report the results of CFAs for the one-factor and two factor models conducted on the WAI-S and WAI-S-T samples separately for all three assessment time points. In both samples, the two-factor model consistently had the best fit for the data across time, as suggested by lower scores on the AIC and ECVI compared to the one-factor model. However, the estimated inter-factor correlation was consistently very high. Since in both samples each model demonstrated a similar model fit across time, longitudinal configural invariance was supported and we proceeded with multigroup CFA to test for metric and scalar invariance.

[Table 3 and 4 near here]

In order to formally test measurement invariance, a longitudinal CFA model was set up for the three time points in both samples separately. In the WAI-S sample, for both the one-factor and two-factor WAI-S structure the ΔCFI criterion (ΔCFI<0.01) and the chi-square difference tests indicated no significant difference in model fit from the configural to the metric model (1-factor model: $\Delta\chi2$ (22)= 27.05 p= 0.209, ΔCFI =.001; two-factor model: $\Delta\chi2$ (22)= 22.09 p= 0.335, ΔCFI = 0) and from the metric to the scalar model (one-factor model: $\Delta\chi2$ (24)= 24.38

p= 0.439 ΔCFI =0; two-factor model: Δχ2 (24)= 23.74 p= 0.467, ΔCFI = 0). Similarly, in the WAI-S-T sample, both the chi-square difference test and ΔCFI criterion supported metric invariance of both the one and two-factor model (one-factor model: Δχ2 (22)= 30.63 p= 0.104; two-factor model: ΔCFI =0.001; Δχ2 (22)= 21.63 p= 0.361, ΔCFI = 0). From metric to scalar invariance, the chi-square difference tests yielded small p-values, the ΔCFI criterion again supported scalar invariance (one-factor model: Δχ2 (24) = 64.29, p <0.001, ΔCFI =.001; two-factor model: Δχ2 (24) = 71.31 p <0.001, ΔCFI = 0.008). Since in both samples the differences in CFI across the increasingly constrained models did not indicate any meaningful difference in model fit (ΔCFI < 0.01) for both the one-factor and two-factor WAI-S structure across time (Cheung & Rensvold, 2002), longitudinal measurement invariance was supported for both the adolescents' and therapists' WAI-S ratings.

 [Table 5 near here]

Measurement Invariance Across Raters

As shown in table 3 and 4, there was a high level of convergence between the therapist-reported and adolescent-reported alliance structure, with the two-factor model showing the best fit for the data in both samples. As such, configural invariance between rater groups was also supported. To assess for metric and scalar invariance we conducted multiple group CFA across the WAI-S and WAI-S-T samples at 6 weeks. When adolescents' and therapists' ratings were compared on the one-factor model, metric and scalar invariance did not hold, as indicated by both a significant chi-square difference test and a difference in CFI larger than .01 (Δχ2 (22)= 47.5 p<.001, ΔCFI =.012; Δχ2 (22)= 191.2 p<.001, ΔCFI = .058, respectively for metric and scalar invariance). For the two-factor model, instead, metric invariance passed the ΔCFI criterion (Δχ2 (10) =35.3 p<.001, ΔCFI = 0.008) and only scalar invariance did not hold (Δχ2 (10)=169.0 p<.001, ΔCFI = 0.051). Therefore, across raters there was no support for full

measurement invariance of the WAI-S, but configural and metric invariance were found (weak invariance) for the two-factor model only.

Measurement Invariance Across Treatments

The WAI-S-T sample was deemed too small to be divided into subgroups, therefore, measurement invariance was tested in the WAI-S sample only. Table 5 shows the results of the CFA conducted on the adolescents' ratings for each type of treatment at 6 weeks (BPI: N=72; CBT: N=78; STPP: N=73). In line with previous results, the two-factor model consistently had a better fit compared to the one-factor model across therapeutic approaches. This finding supported configural invariance across treatment groups and we proceeded to test full measurement invariance. For the one-factor structure, although the scalar model failed the chi-square difference test ($\Delta\chi2$ (22) =34.7 p=0.042), according to the ΔCFI criterion both metric and scalar invariance held ($\Delta\chi2$ (22) =22.8 p=0.413, ΔCFI =.0; ΔCFI=.007 respectively). Metric and scalar invariance were also found for the two-factor structure according to both the chi-square difference test and the ΔCFI criterion ($\Delta\chi2$ (20) =28.3 p=0.105, ΔCFI = 0.005; $\Delta\chi2$ (20) = 26.7 p=0.142, ΔCFI = 0.004 respectively).

[Table 5 near here]

Discussion

The current study is the first to evaluate the factor structure of the WAI-S and its measurement invariance in psychotherapy for adolescent depression. Although the WAI-S is based on Bordin's (1979) definition of the alliance and therefore structured in three subscales (Task, Bond and Goal), the results of this study do not provide empirical support for this measurement model in youth psychotherapy. Of the four alliance models tested, the two-factor (Bond and Task-Goal combined) and the general one-factor model seemed to represent more adequately the WAI-S structure from both the adolescent and therapist perspective. Despite some evidence

for two-dimensionality, given the high correlation between the factors, the instrument might in practice be best treated as unidimensional.

The overall poor fit of the three-factor and the bi-factor models of the WAI-S, as well as the high levels of correlation between the subscales, are common findings in youth alliance research (Anderson et al., 2012; Diamond et al., 2006; DiGiuseppe et al., 1996). Developmental issues might be responsible for the failure to support Bordin's (1979) definition of the alliance with youth. For instance, it has been argued that young people might not discriminate between different aspects of the alliance (DiGiuseppe et al., 1996; Zack et al., 2007). This could be not only because the ability to differentiate tasks and goals of therapy might require complex cognitive skills (i.e. the ability to think hypothetically); but also because adolescents might not be familiar with the activities expected in therapy. Furthermore, young people are often referred to treatment rather than seeking therapy themselves, which might further complicate the establishment of an agreement on therapy goals. However, factor analytic research on the WAI in adult samples has also failed to support the distinction between goal and task (Andrusyna et al., 2001; Corbière et al., 2006; Falkenström et al., 2015; Hatcher et al., 1996; Hatcher & Gillaspy, 2006); which might suggest that the poor distinction between these subscales might be due to the measure itself.

In this study the two- and the one-factor model both had adequate to good fit for the data, with the former showing a slightly better fit than the simpler, one-factor model. This might suggest that a general alliance factor on its own does not sufficiently represent the data and provides some support for the bidimensional (Bond-Collaboration) structure of the WAI-S. The two factor model has also found some support for the parent, but not the self-report, version of the WAI-S in online youth psychotherapy (Anderson et al., 2012); and for the observer and the self-report ratings of the WAI-S in adult psychotherapy (Andrusyna et al., 2001; Hatcher & Gillaspy, 2006, respectively). Furthermore, the Therapeutic Alliance Scale for Children

(TASC; Shirk & Saiz, 1992), the other popular alliance measure in youth psychotherapy, has also displayed a similar structure (Ormhaug, Shirk, & Wentzel-Larsen, 2015; Shirk & Saiz, 1992). However, the two-factor model yielded a high correlation between Bond and Collaboration, which raises questions about the practical and statistical distinction between the two latent variables.

On the one hand, the strong association between the two specific alliance dimensions might not be sufficient to demonstrate that youth alliance is a one-factor phenomenon. Bond and Collaboration have strong face validity as being indicative of two distinct, but mutually dependent aspects of the alliance (Hougaard, 1994; Shirk & Saiz, 1992). Collaboration comprises patient and therapist negotiation and agreement on the work of therapy; bond refers to the affective aspects of the relationship. Despite being unique in their content, they are supposed to be linked: positive emotional bonding likely heightens the patient's motivation and involvement in therapy; similarly high levels of collaboration foster the development of a strong bond. This might mean that each alliance dimension cannot be achieved without the other. For instance, it might not be possible to establish a positive bond with young people without developing an agreement on therapy goals and tasks. Accordingly, it has been argued that while the bond might be the basis for the therapeutic work with children, developing a shared understanding of therapy goals/task might be essential with older children and adolescents (Sandler, Kennedy, & Tyson, 1980; DiGiuseppe et al., 1996). This might be because the bond and the wish for help might not be sufficient to carry the therapeutic work due to "the mistrust, suspicion, scepticism, and doubt that they often experience in association with their effort to break the ties with the parental figure" (Sandler, Kennedy, & Tyson, 1980, p.50). As such, to collaboratively work in treatment the adolescent is expected to develop "a proportionally greater awareness of his problems and greater wish to work towards their

solutions; for him less of the work should depend on a positive relationship with the therapist" (Sandler, Kennedy, & Tyson, 1980, p. 45).

On the other hand, the high correlation between Bond and Collaboration might imply that these factors cannot be meaningfully differentiated with the WAI-S. This could be either because they are poorly represented as distinct in this measure (i.e. the WAI-S items are designed in a way that does not allow for this subtler distinction), or because youth alliance is an integrated phenomenon. As such, it could be argued that the parsimonious, one-factor structure is psychometrically more valid. This is in line with the majority of the empirical literature on the structure of the WAI-S, as well as of a range of other alliance measures, which also supports the acceptance and further use of a single, general alliance factor in youth psychotherapy (Diamond et al., 2006; DiGiuseppe et al., 1996; Faw et al., 2005; Fjermestad et al., 2012; Hogue et al., 2006).

The measurement invariance analyses showed that from both rater groups there was evidence of longitudinal measurement invariance; which suggests that the way adolescents and therapists understand and rate the scale items does not change over the course of treatment. This finding is in line with the result of a previous study assessing longitudinal measurement invariance of a revised version of the WAI in adult counselling/psychotherapy (Falkenström, Hatcher, Skjulsvik, Larsson, & Holmqvist, 2014). In our sample, from the adolescent perspective, both the one- and especially the two-factor structure of the WAI-S also showed measurement invariance across treatments. This might suggest that the WAI-S items have the same meaning across different treatment modalities when rated by adolescents. Finally, in our sample measurement invariance across raters (i.e. adolescents and therapists) was not supported for the one-factor structure of the WAI-S; while the two-factor model showed configural and metric invariance, but not scalar invariance. Since scalar invariance was not

attained, there might be differences in the way adolescents and therapist interpret the scale; thus, mean ratings of adolescents and therapists cannot strictly be compared. Notably, the WAI-S was originally created as a patient-report measure and subsequently adapted for therapists, so perhaps caution should be taken in the future when developing measures for different raters. Overall, given the dearth of research on the measurement invariance of the WAI-S in youth psychotherapy, this aspect of the scale requires further investigation.

Another contribution of this study is that it showed the presence of a method effect associated with the two negatively worded items of the WAI-S. This finding extends previous results on the distinctive performance of the negatively worded items in both the WAI and WAI-S (Hatcher & Gillaspy, 2006), an issue that led to a decision to include only positively worded items in the latest revised version of the scale (WAI-SR, Hatcher & Gillaspy, 2006). In this regard, it is worth noticing that the WAI-S contains only two of the fourteen negatively worded items of the original scale, which are both included in the Goal subscale. This aspect of the scale has not received much empirical attention; nor has the exclusion of negatively worded items from the Task and Bond subscales been investigated for potential loss of relevant information. There is a general lack of attention to the presence of method effects associated with item wording in alliance research. This is, however, an important issue when developing alliance measures and assessing their factor structure, since method effects can affect the results of exploratory factor analyses, creating polarities (DeVellis, 2016). For instance, a few factor analytic studies (Accurso, Hawley, & Garland, 2013; Ormhaug et al., 2015) showed that adolescents' ratings of the alliance were organised by item valence (i.e., whether they were positively or negatively worded) and not by content. Results of this kind might be due to a method effect, rather than reflect the real structure of the scales. Whether positively and negatively worded items are included in self-report questionnaires, they should be equally

distributed within subscales, their wording should be clear and simple, and their association with method effects should be investigated to avoid some of the bias discussed.

Strengths and Limitations

This study had several strengths, including being the first known study to carry out an in-depth exploration of the factor structure of WAI-S and to evaluate its measurement invariance in youth psychotherapy. This is an important research topic given the alliance is among the most investigated variables in psychotherapy and the WAI-S is in widespread use. Further, assessing measurement invariance is an essential prerequisite in studies comparing people from different groups or evaluating change over time since it ensures that different respondents interpret questionnaire items in the same way. Other strengths concern the inclusion of three distinct treatment modalities as well as of ratings of the alliance at different time points and from the perspective of both adolescents and therapists. This study is also the first to report on the method effect of the WAI-S, a neglected area in the alliance literature. This issue demands attention in future developments of the scale and further research is needed to explore the substantive nature of such method effects.

Nevertheless, important considerations need to be kept in mind when interpreting the results of this study. One limitation concerns the relatively small sample size of the therapists' ratings of the alliance. As a consequence, the results of the measurement invariance across treatment arms were limited to adolescents' ratings of the WAI-S. Further research using larger samples across different therapeutic approaches is needed to replicate this finding and to also test for measurement invariance across treatments from the therapist perspective. Furthermore, given the importance of parental alliance in youth psychotherapy (Kazdin, Whitley, & Marciano, 2006; McLeod, 2011), future studies should investigate the factorial structure of the parent version of the WAI-S too. Additionally, we were unable to control for the clustering within the therapist, yet this is likely to be the result of there being too many therapists with a single case.

Recent research on the factor structure of the WAI-SR showed that the model fit did not improve significantly when testing a model that separates variance due to therapists from variance due to patients (Falkenström et al., 2015); while a therapist effect was found for the therapist versions (Hatcher, Lindqvist, & Falkenström, 2019). Future research should endeavour to control for therapist effect, especially when using therapists' ratings of the alliance as they might be influenced by the therapist's rating style. Another limitation might come from making the decision of which factor solution is the most appropriate based on small differences in the model fit, which carries the risk of accepting an unnecessarily complex model instead of a more appropriate simple one. To address this concern, in addition to careful evaluation of statistical information, the theory underlying each model was used to inform all decisions. Hence, every attempt was made to select the alliance structure on the basis of both theoretical and statistical grounds.

Conclusion

This study did not confirm the hypothesised Task-Goal-Bond structure of the WAI-S in the context of psychotherapy with depressed adolescents but supported the use of the WAI-S total, general score. Nevertheless, a two-factor structure, in which Task and Goal are collapsed into one overall 'Collaboration' factor, had some empirical support and warrants further investigation. In addition, this study provided evidence of the WAI-S longitudinal measurement invariance and of the existence of a method effect associated with the negatively worded items of the WAI-S. Measurement invariance across therapeutic approaches was also found from the adolescent perspective, but there was no evidence of full measurement invariance across adolescent and therapist ratings. As the alliance is often used across different types of treatment, rater perspectives and client group, the assessment of measurement invariance is an important research question for future research.

Ultimately, the results of this study support the view of the measure's authors that 'one overriding alliance factor appears to be the most salient dimension measured by the WAI' (Tracey & Kokotovic, 1989, p. 209). Despite the conceptual value of considering different dimensions of the alliance, it seems challenging to isolate and tease apart different alliance dimensions in both empirical research and clinical practice. Future research is needed to confirm these findings and investigate if meaningful differences between the different alliance dimensions can be found. One way of testing this might be to assess whether Bond and Collaboration, when considered as two distinct dimensions, each have a unique relationship to outcomes across different forms of psychotherapy for young people. Qualitative research strategies might also be necessary to grasp the nature and the unique, holistic constellations of the alliance and its dimensions in youth psychotherapy. A renewal of the conversation between the theoretical and empirical definitions of the alliance in the context of therapy with young people might be needed for a deeper understanding of the alliance as a developmentally sensitive construct. Both theory-based (top-down) and empirically derived (bottom-up) approaches are needed to clarify the nature of the alliance and its function in youth psychotherapy.

References

Accurso, E. C., Hawley, K. M., & Garland, A. F. (2013). Psychometric properties of the Therapeutic Alliance Scale for Caregivers and Parents. *Psychological Assessment*, *25*(1), 244–252. https://doi.org/10.1037/a0030551

Anderson, R. E. E., Spence, S. H., Donovan, C. L., March, S., Prosser, S., & Kenardy, J. (2012). Working alliance in online cognitive behavior therapy for anxiety disorders in youth: Comparison with clinic delivery and its role in predicting outcome. *Journal of Medical Internet Research*, *14*(3), e88. https://doi.org/10.2196/jmir.1848

Andrusyna, T. P., Tang, T. Z., DeRubeis, R. J., & Luborsky, L. (2001). The Factor Structure

of The Working Alliance Inventory in Cognitive-Behavioral Therapy. *The Journal of Psychotherapy Practice and Research : A Journal of the American Psychiatric Association*, *10*(3), 173–178.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, *16*(3), 252–260.

Capaldi, S., Asnaani, A., Zandberg, L. J., Carpenter, J. K., & Foa, E. B. (2016). Therapeutic Alliance during Prolonged Exposure Versus Client-Centered Therapy for Adolescent Posttraumatic Stress Disorder. *Journal of Clinical Psychology*, *72*(10), 1026–1036.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902

Corbière, M., Bisson, J., Lauzon, S., & Ricard, N. (2006). Factorial validation of a French short-form of the Working Alliance Inventory. https://doi.org/10.1002/mpr.27

Dennis, M. L., Titus, J. C., Diamond, G. S., Donaldson, J., Godley, S. H., Tims, F. M., … Scott, C. K. (2002). *The Cannabis Youth Treatment (CYT) experiment: Rationale, study design and analysis plans*. *Addiction* (Vol. 97). UCHC. https://doi.org/10.1046/j.1360-0443.97.s01.2.x

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). Sage publications.

Diamond, G. M., Hogue, A., Liddle, H. A., & Dakof, G. A. (1999). Alliance-building interventions with adolescents in family therapy: A process study. *Psychotherapy*, *36*(4), 355–368.

Diamond, G. S., Liddle, H. A., Wintersteen, M. B., Dennis, M. L., Godley, S. H., & Tims, F. (2006). Early therapeutic alliance as a predictor of treatment outcome for adolescent cannabis users in outpatient treatment. *American Journal on Addictions*, *15*(SUPPL. 1),

26–33. https://doi.org/10.1080/10550490601003664

DiGiuseppe, R., Linscott, J., & Jilton, R. (1996). Developing the therapeutic alliance in child-adolescent psychotherapy. *Applied and Preventive Psychology*, *5*(2), 85–100. https://doi.org/10.1016/S0962-1849(96)80002-3

Elvins, R., & Green, J. (2008). The conceptualization and measurement of therapeutic alliance: An empirical review. *Clinical Psychology Review*, *28*(7), 1167–1187. https://doi.org/10.1016/j.cpr.2008.04.002

Falkenström, F., Hatcher, R. L., & Holmqvist, R. (2015). Confirmatory Factor Analysis of the Patient Version of the Working Alliance Inventory–Short Form Revised. *Assessment*, *22*(5), 581–593. https://doi.org/10.1177/1073191114552472

Falkenström, F., Hatcher, R. L., Skjulsvik, T., Larsson, M. H., & Holmqvist, R. (2014). Development and Validation of a 6-item Working Alliance Psychotherapy. *Psychological Assessment*, (27), 169–183. https://doi.org/10.1037/pas0000038

Faw, L., Hogue, A., Johnson, S., Diamond, G. M., & Liddle, H. A. (2005). The Adolescent Therapeutic Alliance Scale (ATAS): Initial psychometrics and prediction of outcome in family-based substance abuse prevention counseling. *Psychotherapy Research*, *15*(1–2), 141–154. https://doi.org/10.1080/10503300512331326994

Figueiredo, B., Dias, P., Lima, V. S., & Lamela, D. (2016). Working Alliance Inventory for Children and Adolescents (WAI-CA). *European Journal of Psychological Assessment*, 1–7. https://doi.org/10.1027/1015-5759/a000364

Fjermestad, K. W., McLeod, B. D., Heiervang, E. R., Havik, O. E., Öst, L. G., & Haugland, B. S. M. (2012). Factor Structure and Validity of the Therapy Process Observational Coding System for Child Psychotherapy-Alliance Scale. *Journal of Clinical Child and Adolescent Psychology*, *41*(2), 246–254. https://doi.org/10.1080/15374416.2012.651999

Flückiger, C., Del, A. C., Wampold, B. E., & Horvath, A. O. (2018). The Alliance in Adult

Psychotherapy: A Meta-Analytic Synthesis. *Psychotherapy*, *55*(4), 316–340.
https://doi.org/10.1037/pst0000172

Freud, A. (1946). *The psycho-analytical treatment of children.* Oxford: Imago Publishing Co.
Retrieved from http://psycnet.apa.org/record/1947-01846-000

Gaston, L. (1990). The concept of the alliance and its role in psychotherapy: Theoretical and
empirical considerations. *Psychotherapy*, *27*(2), 143–153. https://doi.org/10.1037/0033-
3204.27.2.143

Goodyer, I. M., Reynolds, S., Barrett, B., Byford, S., Dubicka, B., Hill, J., … Midgley, N.
(2016). Cognitive behavioural therapy and short-term psychoanalytical psychotherapy
versus a brief psychosocial intervention in adolescents with unipolar major depressive
disorder (IMPACT): a multicentre, pragmatic, observer-blind, randomised controlled
superiori. *The Lancet Psychiatry*, *0366*(16), 1–11. https://doi.org/10.1016/S2215-
0366(16)30378-9

Goodyer, I. M., Tsancheva, S., Byford, S., Dubicka, B., Hill, J., Kelvin, R., … Fonagy, P.
(2011). Improving mood with psychoanalytic and cognitive therapies (IMPACT): A
pragmatic effectiveness superiority trial to investigate whether specialised psychological
treatment reduces the risk for relapse in adolescents with moderate to severe unipolar
depres. *Trials*, *12*(1), 175. https://doi.org/10.1186/1745-6215-12-175

Hartman, R. (2017). CFA in lavaan. Retrieved January 26, 2019, from
http://www.understandingdata.net/2017/03/22/cfa-in-lavaan/

Hatcher, R. L., Barends, A. W., Burg, R., Reynolds, M., Shannon, P., Hansell, J., & Hatcher,
S. (1996). Patients ' View of the Alliance in Psychotherapy : Exploratory Factor
Analysis of Three Alliance Measures. *Journal of Consulting and Clinical Psychology*,
64(6), 1326–1336. https://doi.org/10.1037/0022-006X.64.6.1326

Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short

version of the Working Alliance Inventory. *Psychotherapy Research*, *16*(1), 12–25. https://doi.org/10.1080/10503300500352500

Hatcher, R. L., Lindqvist, K., & Falkenström, F. (2019). Psychometric evaluation of the Working Alliance Inventory—Therapist version: Current and new short forms. *Psychotherapy Research*, *0*(0), 1–12. https://doi.org/10.1080/10503307.2019.1677964

Hawley, K. M., & Garland, A. F. (2008). Working alliance in adolescent outpatient therapy: Youth, parent and therapist reports and associations with therapy outcomes. *Child and Youth Care Forum*, *37*(2), 59–74. https://doi.org/10.1007/s10566-008-9050-x

Hogue, A., Dauber, S., Stambaugh, L. F., Cecero, J. J., Howard, L. A., & Liddle, H. A. (2006). Early Therapeutic Alliance and Treatment Outcome in Individual and Family Therapy for Adolescent Behavior Problems Aaron. *J Consult Clin Psychol*, *74*(1), 121–129. https://doi.org/10.1037/0022-006X.74.1.121.

Horvath, A. O. (2011). Alliance in Common Factor Land: A view through the research lens. *Research in Psychotherapy: Psychopathology, Process and Outcome*, *14*(1), 121. https://doi.org/10.4081/ripppo.2011.45

Horvath, A. O. (2018). Research on the alliance: Knowledge in search of a theory. *Psychotherapy Research*, *28*(4), 499–516.

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, *36*(2), 233.

Horvath, A. O., & Greenberg, L. S. (1994). The Working alliance : theory, research, and practice. *Journal of Counseling Psychology*, *36*(2), 223.

Hougaard, E. (1994). The therapeutic alliance–A conceptual analysis. *Scandinavian Journal of Psychology*, *35*(1), 67–85.

Hu, L., & 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), 1–55. https://doi.org/10.1080/10705519909540118

IBM, C. (2016). IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.

Karver, M. S., De Nadai, A. S., Monahan, M., & Shirk, S. R. (2018). Meta-analysis of the prospective relation between alliance and outcome in child and adolescent psychotherapy. *Psychotherapy*, 55(4), 341–355. https://doi.org/10.1037/pst0000176

Karver, M. S., Shirk, S., Handelsman, J. B., Fields, S., Crisp, H., Gudmundsen, G., & McMakin, D. (2008). Relationship processes in youth psychotherapy: Measuring alliance, alliance-building behaviors, and client involvement. *Journal of Emotional and Behavioral Disorders*, 16(1), 15–28. https://doi.org/10.1177/1063426607312536

Kazdin, A. E. (2003). Psychotherapy for children and adolescents. *Annual Review of Psychology*, 54(1), 253–276.

Kazdin, A. E., Whitley, M., & Marciano, P. L. (2006). Child-therapist and parent-therapist alliance and therapeutic change in the treatment of children referred for oppositional, aggressive, and antisocial behavior. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 47(5), 436–445. https://doi.org/10.1111/j.1469-7610.2005.01475.x

Kenny, D. A. (2015). Measuring Model Fit. Retrieved January 5, 2019, from http://www.davidakenny.net/cm/fit.htm

Kyriazos, T. A. (2018). Improper Solutions in SEM. Retrieved January 19, 2019, from http://www.ssicentral.com/lisrel/techdocs/HowLargeCanaStandardizedCoefficientbe.pdf

Lingard, H. C., & Rowlinson, S. (2005). *Sample size in factor analysis: why size matters Helen C Lingard and Steve Rowlinson*.

McLeod, B. D. (2011). Relation of the alliance with outcomes in youth psychotherapy: A meta-analysis. *Clinical Psychology Review*, 31(4), 603–616. https://doi.org/10.1016/j.cpr.2011.02.001

Meyer, B., Pilkonis, P. A., Krupnick, J. L., Egan, M. K., Simmens, S. J., & Sotsky, S. M.

(2002). Treatment Expectancies, Patient Alliance, and Outcome: Further Analyses From

the National Institute of Mental Health Treatment of Depression Collaborative Research

Program. https://doi.org/10.1037/0022-006X.70.4.1051

Norcross, J. C. (2011). *Psychotherapy Relationships That Work*. Oxford University Press.

https://doi.org/10.1093/acprof:oso/9780199737208.001.0001

Ormhaug, S. M., Shirk, S. R., & Wentzel-Larsen, T. (2015). Therapist and client perspectives

on the alliance in the treatment of traumatized adolescents. *European Journal of

Psychotraumatology*, *6*. https://doi.org/10.3402/ejpt.v6.27705

R Core Team. (2018). R: A language and environment for statistical computing. *R

Foundation for Statistical Computing, Vienna, Austria*. Retrieved from https://www.r-

project.org/

Rosseel, Y. (2012). lavaan: an R package for structural equation modeling and more Version

0.5-12 (BETA). *Ournal of Statistical Software*, *48*(2), 1–36. Retrieved from

http://cran.r-project.org/.

Rubin, D. B. (1987). Multiple imputation for survey nonresponse. New York: Wiley.

Sandler, J., Kennedy, H., & Tyson, R. L. (1980). *The technique of child psychoanalysis:

Discussions with Anna Freud*. Harvard University Press.

Shelef, K., & Diamond, G. M. (2008). Short form of the revised vanderbilt therapeutic

alliance scale: Development, reliability, and validity. *Psychotherapy Research*, *18*(4),

433–443. https://doi.org/10.1080/10503300701810801

Shelef, K., Diamond, G. M., Diamond, G. S., & Liddle, H. A. (2005). Adolescent and Parent

Alliance and Treatment Outcome in Multidimensional Family Therapy. *Journal of

Consulting and Clinical Psychology*, *73*(4). https://doi.org/10.1037/0022-006X.73.4.689

Shirk, S. R. (2013). *Cognitive development and child psychotherapy*. Springer Science &

Business Media.

Shirk, S. R., Karver, M. S., & Brown, R. (2011). The Alliance in Child and Adolescent Psychotherapy. *Psychotherapy*, *48*(1), 17–24. https://doi.org/10.1037/a0022181

Shirk, S. R., & Saiz, C. C. (1992). Clinical, empirical, and developmental perspectives on the therapeutic relationship in child psychotherapy. *Development and Psychopathology*, *4*(4), 713–728. https://doi.org/10.1017/S0954579400004946

Tarka, P. (2018). An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences. *Quality & Quantity*, *52*(1), 313–354. https://doi.org/10.1007/s11135-017-0469-8

Tetzlaff, B. T., Kahn, J. H., Godley, S. H., Godley, M. D., Diamond, G. S., & Funk, R. R. (2005). Working alliance, treatment satisfaction, and patterns of posttreatment use among adolescent substance users. *Psychology of Addictive Behaviors*, *19*(2), 199–207. https://doi.org/10.1037/0893-164X.19.2.199

Tracey, T. J., & Kokotovic, A. M. (1989). *Factor Structure of the Working Alliance Inventory*. *Psychological Assessment: A Journal of Consulting and Clinical Psychology* (Vol. 1).

Zack, S. E., Castonguay, L. G., & Boswell, J. F. (2007). Youth Working Alliance: A Core Clinical Construct in Need of Empirical Maturity. *Harvard Review of Psychiatry*, *15*(6), 278–288. https://doi.org/10.1080/10673220701803867

Table 1. Demographics for the WAI-S and WAI-S-T samples

	WAI-S sample (N=338)		WAI-S-T sample (N= 159)	
	Mean	**SD**	**Mean**	**SD**
Age	15.59	1.41	15.57	1.52
	N	**%**	**N**	**%**
Gender				
Female	247	73.1	110	69.2
Ethnicity [1]				
White British	261	77.4	114	71.7
Any other group	69	20.3	43	27

[1] 8 missing in the WAIS sample; 2 missing in the WAI-S-T sample

Table 2. CFAs of the WAI-S and WAI-S-T at 6 weeks: Model fit information using the Maximum Likelihood Estimator with Robust Standard Errors

Model	Robust Model Fit Indices								$\chi 2/df$
WAI-S sample [a]	$\chi 2$	**df**	**CFI**	**TLI**	**RMSEA**	**SRMR**	**AIC**	**ECVI**	
One-factor	147.39***	54	0.935	0.920	0.103	0.047	8757.58	1.23	13.13***
One-factor Method Effect	133.32***	53	0.944	0.939	0.096	0.040	8740.00	1.15	
Two- factor	108.77***	53	0.962	0.952	0.080	0.044	8705.0	0.99	12.71***
Two-Factor Method Effect	94.85***	52	0.971	0.963	0.071	0.037	8687.90	0.92	
WAI-S-T sample									
One-factor	135.77***	54	0.922	0.904	0.111	0.047	4328.7	1.63	13.75***
One-factor Method Effect	120.12***	53	0.936	0.920	0.102	0.040	4312.5	1.51	
Two- factor	132.12***	53	0.924	0.905	0.111	0.047	4327.3	1.62	14.11***

	χ2	df	CFI	TLI	RMSEA	SRMR	AIC	ECVI	
Two-Factor Method Effect	116.48***	52	0.938	0.921	0.101	0.040	4310.1	1.50	
Three-factor	133.01***	51	0.922	0.900	0.114	0.046	4329.8	1.64	15.96***
Three-factor Method Effect	117.46***	50	0.936	0.916	0.104	0.040	4314.4	1.53	

Note. Method effect= the model specification included the correlation between the error of the two negatively worded items. The bi-factor model was tested in both samples but could not be identified.

a= The three-factor solution was unreliable in this sample because the correlation between Task and Goal was >1.

***= p < .001

Table 3. CFAs of the WAI-S at 6, 12 and 36 weeks: Factor Intercorrelations and Model fit information using the Maximum Likelihood Estimator with Robust Standard Errors

Sample	Model	Robust Model Fit Indices								Factor correlations
6 weeks a		$\chi 2$	df	CFI	TLI	RMSEA	SRMR	AIC	ECVI	
	One-factor	133.32***	53	0.944	0.939	0.096	0.040	8740.00	1.15	\
	Two-factor	94.85***	52	0.971	0.963	0.071	0.037	8687.90	0.92	0.91***
12 weeks b										
	One-factor	167.30***	53	0.940	0.925	0.108	0.035	9374.8	1.198	\
	Two-factor	128.90***	52	0.960	0.949	0.088	0.032	9323.0	0.988	0.92***
36 weeks c										
	One-factor	175.89***	53	0.960	0.951	0.094	0.028	8157.12	1.13	\
	Two-factor	118.61***	52	0.967	0.958	0.086	0.026	8138.30	1.04	0.96***

Note. All models include the correlation between the error of the two negatively worded items.

a N= 223; b N= 247; c N= 222; ***= p < .001.

Table 4. CFAs of the WAI-S-T at 6, 12 and 36 weeks: Factor Intercorrelations and Model fit information using the Maximum Likelihood Estimator with Robust Standard Errors

Sample	Model	Robust Model Fit Indices								Factor
6 weeks a		$\chi 2$	df	CFI	TLI	RMSEA	SRMR	AIC	ECVI	correlations
	One-factor	120.12***	53	0.936	0.920	0.102	0.040	4312.47	1.51	\
	Two-factor	116.48***	52	0.938	0.921	0.101	0.040	4310.10	1.50	0.96***
12 weeks b										
	One-factor	90.36**	53	0.969	0.962	0.080	0.032	3470.60	1.45	\
	Two-factor	88.30***	52	0.971	0.963	0.079	0.035	3468.10	1.43	0.95***
36 weeks c										
	One-factor	80.01**	53	0.953	0.942	0.093	0.053	1750.50	2.53	\
	Two-factor	73.05*	52	0.963	0.954	0.083	0.049	1745.60	2.45	0.91***

Note. All models include the correlation between the error of the two negatively worded items.

a N= 139; b N=119; N=63; ***= p < .001.; **= p < .01.; *= p < .05.

Table 5. CFAs of the WAI-S at 6 weeks for each treatment arm: Factor Intercorrelations and Model fit information using the Maximum Likelihood Estimator with Robust Standard Errors

Treatment	Model	Robust Model Fit Indices								Factor
BPI a		$\chi 2$	df	CFI	TLI	RMSEA	SRMR	AIC	ECVI	correlations
	One-factor	99.59**	53	0.895	0.870	0.110	0.068	2791.84	2.63	\
	Two-factor	86.87**	52	0.923	0.902	0.096	0.064	2777.46	2.43	0.87***
CBT b										
	One-factor	94.39***	53	0.867	0.834	0.110	0.076	3067.38	2.42	\

Two-factor	71.30*	52	0.939	0.922	0.075	0.073	3040.15	2.07	0.77***
STPP c									
One-factor	93.16***	53	0.947	0.934	0.107	0.041	2878.97	2.42	\
Two-Factor	82.85**	52	0.961	0.949	0.094	0.041	2868.95	2.28	0.94***

Note: All models include the correlation between the error of the two negatively worded items.

a N= 72; b N= 78; c N= 73; ***= p < .001; **= p < .01; *= p < .05.