Psychotherapy Research

ISSN: 1050-3307 (Print) 1468-4381 (Online) Journal homepage: www.tandfonline.com/journals/tpsr20

Evidence of Measurement Invariance in the Working Alliance Inventory Across In-Person and Videoconferencing Psychotherapy

Alberto Stefana

To cite this article: Alberto Stefana (2026) Evidence of Measurement Invariance in the Working Alliance Inventory Across In-Person and Videoconferencing Psychotherapy, Psychotherapy Research, 36:3, 583-593, DOI: 10.1080/10503307.2025.2495827

To link to this article: https://doi.org/10.1080/10503307.2025.2495827

Published online: 08 May 2025.

Psychotherapy Research, 2026
Vol. 36, No. 3, 583–593, https://doi.org/10.1080/10503307.2025.2495827

RESEARCH ARTICLE

Evidence of Measurement Invariance in the Working Alliance Inventory Across In-Person and Videoconferencing Psychotherapy

ALBERTO STEFANA [1,2]

[1]*Department of Psychiatry and Behavioral Health, Ohio State University, Columbus, OH, USA &* [2]*Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy*

(Received 28 November 2024; revised 13 February 2025; accepted 15 April 2025)

Abstract
Objective: The Working Alliance Inventory-Short Revised (WAI-SR) is a self-report measure of therapeutic alliance from the patient's perspective. This study aimed to examine the factor structure of the WAI-SR and evaluate its measurement invariance across in-person and videoconferencing psychotherapy sessions.
Method: The study sample consisted of 1043 adult patients. Exploratory and confirmatory factor analyses were conducted to evaluate the WAI-SR dimensions and structure. Measurement invariance was tested using multi-group confirmatory factor analysis, multiple indicator multiple causes model analysis, and item bias analysis.
Results: The findings supported a three-factor structure of the WAI-SR, encompassing goal, task, and bond dimensions. Results demonstrated configural, metric, scalar, and residual invariance across in-person and videoconferencing formats. The factor structure remained consistent after accounting for patient age and gender. Additionally, no differential item functioning or bias was observed between groups. The WAI-SR exhibited excellent internal consistency and composite reliability.
Conclusions: The WAI-SR is a reliable and valid tool for assessing therapeutic alliance across both in-person and videoconferencing sessions. The equivalence in measurement properties and mean scores across modalities highlights the adaptability of therapeutic alliance to virtual environments, supporting the broader use of telehealth in psychotherapy.

Keywords: alliance; psychotherapy; teletherapy; videoconferencing; measurement invariance

Clinical or methodological significance of this article: This study established that the Working Alliance Inventory-Short Revised (WAI-SR) is a psychometrically robust instrument for assessing therapeutic alliance in both in-person and telehealth settings, free from measurement bias. The demonstrated equivalence ensures a reliable method for monitoring alliance quality across session formats, facilitating the early identification of alliance-related issues that could hinder treatment outcomes.

The social restrictions related to the coronavirus 2019 pandemic (Stefana et al., 2020) compelled most mental health clinicians to adopt hybrid or fully remote care models based on telecommunication technologies (American Psychological Association, 2023). After the lifting of pandemic-related restrictions, the number of clinicians offering psychological assessment, counseling, and psychotherapy exclusively via telephone or videoconferencing decreased by about two-thirds (American Psychological Association, 2023). However, since teletherapy was perceived as valuable by most clinicians

Correspondence concerning this article should be addressed to Alberto Stefana, Department of Psychiatry and Behavioral Health, Ohio State University, Columbus, OH 43210, USA. Email: alberto.stefana@gmail.com, stefana.1@osu.edu

(American Psychological Association, 2022) and satisfactory by many patients—particularly females and younger individuals with mild symptoms (Neumann et al., 2023)—, the percentage of clinicians offering a hybrid option in the post-pandemic period has approached 70% and continues to rise (American Psychological Association, 2023).

Teletherapy, along with the broader fields of telepsychology and telepsychiatry, has become a significant component of standard practice for many mental health services and clinicians (American Psychological Association, 2024; Mishkind et al., 2024). Following the transition to telehealth, researchers have increased their efforts to investigate this modality (Feng et al., 2024). A growing body of research supports the effectiveness and efficacy of teletherapy across various settings and clinical populations, with many studies indicating that teletherapy can usually be as effective as in-person therapy in terms of treatment outcomes (Crowe et al., 2023; Giovanetti et al., 2022; Krzyzaniak et al., 2024; Lin et al., 2022; Wirth et al., 2023). Beyond psychotherapy outcome research, another major area of study explores therapists' and patients' experiences of the therapeutic relationship (e.g., Aafjes-van Doorn et al., 2024; Davis et al., 2024; Seuling et al., 2024), which is closely linked to treatment outcomes (Aafjes-van Doorn et al., 2024). This line of research examines how specific elements of the therapeutic relationship manifest and evolve in teletherapy compared to in-person psychotherapy. However, it raises a significant potential methodological problem.

For valid comparisons, the same phenomena must be observed or measured across contexts (Leitgöb et al., 2023). Comparative research requires that psychological inventories measure constructs with the same meaning across groups/settings, ensuring that quantitative group comparisons reflect true differences, uncontaminated by group/setting-specific attributes unrelated to the target construct (Gregorich, 2006). In other words, for differences and similarities to be meaningful when comparing in-person and videoconferencing psychotherapy, constructs must be measured consistently across session formats. Establishing measurement invariance is essential, ensuring that scores are comparable across groups and that items contribute equally based on factor loadings and intercepts or thresholds (Maassen et al., 2023; Sass, 2011). This invariance ensures that constructs are operationalized similarly across different psychotherapy formats. However, it must be noted that while measurement invariance is a prerequisite for tests of changes, mean comparisons, or differential relations across groups or settings, noninvariance can be informative (Putnick & Bornstein, 2016). The latter can lead to important theoretical and clinical insights about how the same construct is interpreted in different groups and should not preclude further analyses of group or setting differences (Putnick & Bornstein, 2016).

Therefore, testing the measurement invariance of scales originally developed and validated in the context of in-person encounters but currently used in teletherapy research is crucial. Among the various elements of the therapeutic relationship (Norcross & Lambert, 2019), the working alliance (Wampold & Flückiger, 2023) and the real relationship (Gelso, 2011) are particularly important for two main reasons. First, they have theoretical and clinical significance, as evidenced by their central role in models of the psychotherapy relationship such as Bruce Wampold's contextual model (Wampold, 2015, 2017) and Charles Gelso's tripartite model (Gelso, 2014; Hill et al., 2024), and by substantial empirical evidence demonstrating their predictive ability for treatment outcomes across various patient disorders, treatment settings, and psychotherapy approaches (Flückiger et al., 2020; Gelso et al., 2018). Second, both constructs are widely investigated in teletherapy and comparative research (e.g., Aafjes-van Doorn et al., 2024; Davis et al., 2024; Seuling et al., 2024). The most commonly used self-report tools are the Real Relationship Inventory (RRI; Gelso et al., 2005; Kelley et al., 2010) and the Working Alliance Inventory (WAI; (Hatcher & Gillaspy, 2006; Horvath & Greenberg, 1989)). However, while the patient version of the RRI (Stefana et al., 2024; Stefana et al., 2025a) has been tested for measurement invariance between in-person and remote psychotherapy formats, the WA I, to the best of our knowledge, has not yet been evaluated in this regard.

In light of the above, the aim of this study was to determine whether the patient version of the Working Alliance Inventory–Short Revised (Hatcher & Gillaspy, 2006) measures the therapeutic alliance construct equivalently in in-person and videoconferencing individual psychotherapy sessions, thus allowing for meaningful comparisons of therapeutic alliance levels.

Methods

Data Set

This secondary analysis utilized baseline data from two prior studies: a longitudinal investigation (Stefana et al., 2023) and a randomized clinical trial (RCT; (Stefana et al., 2024)). For the RCT, baseline data were collected before participant

randomization. Both studies received Institutional Review Board (IRB) approval from the University of North Carolina at Chapel Hill (IRB No. 23-0216 for the longitudinal study, approved March 6, 2023; IRB No. 23-1067 for the clinical trial, approved 31 July 2023).

Sampling Procedures

Participants in both studies were adults aged 18 or older, fluent in English, and currently engaged in individual psychotherapy. The RCT additionally required a minimum therapy frequency of two sessions per month.

Recruitment for the longitudinal study occurred between March and April 2023, while the RCT recruited participants from September to November 2023. Both studies utilized two online patient registries: ResearchMatch and Research for Me. The former is a national resource supported by the National Institutes of Health's Clinical and Translational Science Awards Program, hosts over 158,000 registered volunteers. Previous research indicates that ResearchMatch volunteers typically provide accurate self-report data, especially in the absence of monetary incentives (Chandler & Shapiro, 2016), and have higher study completion rates compared to peer-recruited participants (Faro et al., 2021).

Electronic informed consent was obtained via Qualtrics, which also facilitated data collection for both studies.

In both the longitudinal study ($N = 700$) and the RCT ($N = 475$), participants were assessed after a session within their ongoing individual psychotherapy treatment. Participants who attended sessions by telephone or in person while lying on a couch were excluded because their numbers were too small to provide sufficient sample size for reliable analysis.

Participants

The combined sample comprised 1043 adults undergoing various individual psychotherapeutic treatments for diverse mental health conditions, who completed the assessment after either a video call session (60%) or an in-person session (40%). The majority identified as female (74%). Participants were primarily aged 18–29 years (31%), followed by those aged 30–39 (29%), and those 50 years or older (25%). Most participants were White (81%). A substantial proportion (87%) had at least one psychiatric diagnosis, with anxiety disorders (70%) and unipolar depressive disorders (59%) being most prevalent. Nearly half (47%) had been in therapy for over 24 months, and the majority attended therapy sessions 2–3 times per month (45%). Table I presents additional demographic, clinical, and treatment details, which are reported separately for the in-person and video call groups.

Measure

The Working Alliance Inventory–Short Revised (WAI-SR; (Hatcher & Gillaspy, 2006)) is a 12-item self-report instrument designed to assess the quality of the therapeutic alliance from the patient's perspective. The measure includes three subscales, derived from Bordin's theory of the therapeutic alliance, each comprising four items: (i) agreement on therapeutic tasks, (i) agreement on therapeutic goals, and (iii) the affective bond between the patient and the therapist. Participants rate each item on a 6-point Likert scale, ranging from 0 ("Not at all") to 5 ("Completely") (Falkenström et al., 2015).

Data analysis

Data suitability for factor analysis was evaluated using the Kaiser-Meyer-Olkin test and Bartlett's test of sphericity. Parallel analysis, based on multiple retention criteria, was conducted via the R package *EFAtools* (v. 0.4.4. (Steiner & Grieder, 2020)) to determine the number of factors. The sample was split, with one half for exploratory factor analysis (EFA) and the other for confirmatory factor analysis (CFA). EFAs were performed using *EFAtools*, extracting factor solutions identified through parallel analysis. CFAs were conducted using *lavaan* (v. 0.6-12; Rosseel, 2012) to evaluate model fit. Fit indices included comparative fit index (CFI) and Tucker-Lewis index (TLI) ($\geq$.95 for excellent fit; $\geq$ .90 for acceptable fit), root mean square error of approximation (RMSEA) ($\leq$.06 for excellent fit; $\leq$ .08 for acceptable fit), and standardized root mean square residual (SRMR) ($\leq$.05 for excellent fit; $\leq$ .08 for acceptable fit) (Hoyle, 2023; Stefana et al., 2025b).

Using the identified factor structure, the full sample was analyzed for measurement invariance of the WAI-SR across psychotherapy formats (in-person vs. video). Descriptive analyses examined means and standard deviations for total and subscale scores, with independent samples t-tests comparing formats. Latent mean differences were tested using the scalar invariance model, with the in-person group as the reference (latent means fixed at zero). Effect sizes were interpreted using Cohen's *d* thresholds (Cohen, 1977), where values of.2, .5,

Table I. Demographics, clinical, and treatment characteristics of participating patients.

Demographics	In person ($N = 413$)	Video call ($N = 630$)	Total ($N = 1043$)
Age (years)			
18–29	29% (120)	32% (200)	31% (320)
30–39	27% (113)	30% (187)	29% (300)
40–49	15% (61)	16% (100)	15% (161)
50–59	15% (62)	12% (76)	13% (138)
≥ 60	14% (57)	11% (67)	12% (124)
Gender			
Woman	75% (309)	74% (467)	74% (776)
Man	19% (80)	16% (100)	17% (180)
Other / Prefer not to say	6% (24)	10% (63)	8% (87)
Ethnicity			
White	83% (342)	80% (502)	81% (844)
Black or African American	11% (44)	8% (49)	9% (93)
Other	6% (27)	13% (79)	10% (106)
Education			
Less than high school	0% (2)	0% (2)	0% (4)
High school graduate	5% (19)	3% (19)	4% (38)
Some college	19% (78)	18% (112)	18% (190)
2-year degree	9% (38)	7% (46)	8% (84)
4-year degree	34% (142)	33% (205)	33% (347)
Professional degree	26% (108)	31% (196)	29% (304)
Doctorate	6% (26)	8% (50)	7% (76)
Clinical characteristics [a]			
Any psychiatric disorder	84% (346)	89% (558)	87% (904)
Any anxiety disorder	269	461	730
Any (unipolar) depressive disorder	230	387	617
Any trauma- and stressor-related disorders	164	211	375
Any neurodevelopmental disorder	93	180	273
Any cluster A personality disorder	57	87	144
Any bipolar or related disorder	63	73	136
Any eating disorder	40	72	112
Any disruptive behavior and dissocial disorder	12	7	19
Schizophrenia or any other psychotic disorders	5	8	13
Treatment characteristics			
In psychotherapy from	18% (76)	13% (80)	15% (156)
0–3 months	26% (106)	24% (153)	25% (259)
4–12 months	10% (40)	15% (94)	13% (134)
13–24 months	46% (191)	48% (303)	47% (494)
>24 months	18% (76)	13% (80)	15% (156)
Session frequency			
1 or less per month	11% (47)	10% (64)	11% (111)
2–3 per month	46% (188)	45% (281)	45% (469)
1 per week	37% (154)	42% (262)	40% (416)
2 or more per week	6% (24)	4% (23)	5% (47)
Therapy location			
Private practice	72% (297)	73% (459)	73% (756)
Private health institution	9% (37)	11% (71)	10% (108)
Public health institution	11% (45)	8% (50)	9% (95)
Other	8% (34)	8% (50)	8% (84)
Therapist gender (Female)	75% (311)	81% (511)	79% (822)

Note. [a]*N* sums to more than the sample size because 647 patients had multiple psychiatric diagnoses.

and .8 represent "merely statistical," "subtle," and "obvious" effects, respectively (Fritz et al., 2012). Statistical significance was established at $p < .05$. Reliability for each subscale was calculated using Cronbach's α (Cronbach, 1951) and McDonald's ω_{total} (Feuerstahler et al., 2020) via the R package *semTools* (v. 0.5–6; (Jorgensen et al., 2022)).

Multigroup CFA with robust maximum likelihood estimation (MLR) was conducted to test measurement invariance while accounting for nonnormality in the data, following established guidelines (Luong & Flake, 2023; Putnick & Bornstein, 2016). Sequential steps included configural, metric (weak factorial), scalar (strong factorial), and residual (strict)

invariance. Model fit was evaluated using CFI, TLI, RMSEA, and SRMR, with $\Delta CFI \leq -.01$, $\Delta TLI \leq -.01$, $\Delta RMSEA \leq .015$, and $\Delta SRMR \leq .03$ (metric) or $\leq .01$ (scalar) indicating invariance (Chen, 2007). The Satorra-Bentler scaled chi-square difference test assessed model comparisons while adjusting for nonnormality (Satorra & Bentler, 2001).

As a subsequent step, a multiple indicator multiple causes (MIMIC) model (Jöreskog & Goldberger, 1975) further evaluated invariance while controlling for patient age and gender (women vs. men). Robust chi-square difference tests with the Satorra-Bentler correction examined demographic effects on measurement structure. Lastly, since the traditional method of multi-group CFA does not provide a measure of effect size for item bias, the *lavaan* package was employed to implement the standardized effect size metric (d_{MACS}). d_{MACS} was used to quantify differential item functioning (DIF). This metric captures differences in intercepts and slopes across groups, capturing the magnitude of disparities in the mean and covariance structures between the two groups (Nye & Drasgow, 2011). Small measurement non-equivalence is indicated by d_{MACS} values ranging from .20 to .40, medium non-equivalence corresponds to values between .40 and .70, and large non-equivalence is indicated by values of .70 or higher (Nye et al., 2019).

There were no missing data, as the Qualtrics survey required responses to all questions, ensuring complete data collection. Analyses were conducted using R statistical software (v. 4.4.2; R Core Team, 2024).

Results

Preliminary Analyses

The Kaiser–Meyer–Olkin test (.947) and the Bartlett test of sphericity ($X^2_{(66)} = 12171$, $p < .001$) confirmed the suitability of the data for factor analysis.

Factor Structure

Parallel analysis. Parallel analysis with exploratory factor analysis (Horn, 1965) identified three factors, whereas the Hull method (Lorenzo-Seva et al., 2011) and comparison data (Ruscio & Roche, 2012) suggested four factors. Given the limited number of items ($k = 12$), four factors appear unlikely. Consequently, EFA was conducted by extracting three factors and, additionally, two factors as proposed by some studies (Paap et al., 2022).

Exploratory factor analysis. In the three-factor model, items loaded onto factors consistent with the WAI-SR validation study: (1) Bond: items 3, 5, 7, 9 (smallest loading = .639); (2) Goal: items 4, 6, 8, 11 (smallest loading = .429); and (3) Task: items 1, 2, 10, 12 (smallest loading = .489). Items 11 and 12 exhibited cross-loadings (> .350) on both Goal and Task but had clear dominant loadings on their assigned factors.

In the two-factor model, the Goal and Task factors were combined (smallest loading = .679), while the Bond factor remained separate (smallest loading = .631). No cross-loadings were observed.

Confirmatory factor analysis. The three-factor model demonstrated good fit indices: $\chi^2_{(51)} = 243.45$, CFI = .955, TLI = .942, RMSEA = .094 (90% CI [.083, .106]), and SRMR = .036. These indices were superior to those of the two-factor model: $\chi^2_{(53)} = 297.28$, CFI = .942, TLI = .928, RMSEA = .105 (90% CI [.00, .19]), and SRMR = 0.041. Therefore, we conducted measurement invariance tests on the three-factor model.

Score and Subscores

The means and standard deviations for each WAI-SR item, grouped by session format (in-person vs video call), are reported in Table II. For the in-person format, item means ranged from 3.61 (*SD* = 1.29) to 5.12 (*SD* = 1.07). For the video call format, item means ranged from 3.70 (*SD* = 1.32) to 5.13 (*SD* = 1.03).

The overall WAI-SR scores showed no significant differences between formats. The mean total WAI score was 52.7 for in-person sessions and 53.0 for video calls, $t_{(1041)} = -.40$, $p = .70$. The Goal subscore was identical for both groups (mean = 17.7), $t_{(1041)} = -.02$, $p = 1.00$. The Task subscore means were 16.5 (in-person) and 16.6 (video call), $t_{(1041)} = -.50$, $p = .60$. he Bond subscore means were 18.5 (in-person) and 18.7 (video call), $t_{(1041)} = -.60$, $p = .60$. These results indicate no significant differences between session formats across the WAI-SR total score and its subscores, supporting the use of the WAI-SR in both in-person and video call therapeutic settings.

Latent Mean Differences

With scalar invariance confirmed, latent mean differences were analyzed to assess WAI-SR subscale differences between session formats, adjusted for age and gender. The in-person group served as the reference, with latent means fixed at zero. Results showed no significant differences between video call and in-person formats across all subscales (p

Table II. Descriptive statistics and unstandardized factor loadings from baseline models.

	WAI-SR items	In-person			Video call				
		Factor loadings	*M*	*(SD)*	Factor loadings	*M*	*(SD)*	DIF	d_{MACS}
Goal	4. My therapist and I collaborate on setting goals for my therapy.	1	4.31	(1.42)	1	4.27	(1.45)	–.089	.043
	6. My therapist and I are working towards mutually agreed upon goals.	1.115*	4.45	(1.43)	1.135*	4.44	(1.45)	–.068	.050
	8. My therapist and I agree on what is important for me to work on.	0.998*	4.58	(1.28)	0.977*	4.59	(1.24)	–.059	.001
	11. My therapist and I have established a good understanding of the kind of changes that would be good for me.	1.103*	4.38	(1.36)	1.084*	4.41	(1.38)	–.036	.003
Task	1. As a result of these sessions I am clearer as to how I might be able to change.	1	3.61	(1.29)	1	3.70	(1.32)	.022	.077
	2. What I am doing in therapy gives me new ways of looking at my problem.	0.924*	4.15	(1.29)	1.008*	4.12	(1.34)	–.067	.106
	10. I feel that the things I do in therapy will help me to accomplish the changes that I want.	1.107*	4.43	(1.37)	1.116*	4.43	(1.36)	–.073	.028
	12. I believe the way we are working with my problem is correct.	1.166*	4.29	(1.36)	1.123*	4.36	(1.37)	–.010	.028
Bond	3. I believe my therapist likes me.	1	4.43	(1.35)	1	4.49	(1.27)	.002	.073
	5. My therapist and I respect each other.	0.775*	5.12	(1.07)	0.744*	5.13	(1.03)	–.037	.047
	7. I feel that my therapist appreciates me.	1.107*	4.44	(1.37)	1.158*	4.42	(1.37)	–.086	.035
	9. I feel my therapist cares about me even when I do things that he/she does not approve of.	1.028*	4.53	(1.37)	0.956*	4.63	(1.29)	.036	.096

Note. DIF = Differential Item Functioning; d_{MACS} = effect size metric for differences in mean and covariance structures. The factor loading for the first item of each subscale was fixed to 1 to set the metric for the latent variable.
*$p < .001$.

> .05). Specifically, the Goal subscale difference was –0.104 ($SE = 0.266$, $p = .697$), the Task subscale difference was –0.260 ($SE = 0.270$, $p = .335$), and the Bond subscale difference was –0.032 ($SE = 0.271$, $p = .907$). Effect sizes were small: $d = –0.11$ for Goal, $d = –0.27$ for Task, and $d = –0.03$ for Bond. These findings indicate comparable levels of therapeutic alliance across formats.

Reliability

The internal consistency and composite reliability of the WAI-SR subscales and total score were evaluated for in-person and video call sessions using Cronbach's α and McDonald's ω_{total}. As shown in Table III, all coefficients were excellent (≥ .895), demonstrating the WAI-SR's reliability in assessing therapeutic alliance across both session formats.

Table III. Reliability coefficients for WAI-SR scale and subscale across in-person and video call sessions.

Scale	In-person		Video call	
	α	ω	α	ω
Total score	.949	.950	.951	.952
Goal	.911	.910	.917	.917
Task	.895	.895	.910	.906
Bond	.906	.910	.903	.912

Note. α = Cronbach's alpha, ω = McDonald's omega total.

Multi-Group Confirmatory Factor Analysis

A multigroup CFA was conducted on the whole sample to assess the measurement invariance of the three-factor WAI-SR between video call and in-person session formats.

Configural invariance. Configural invariance was tested to determine if the factor structure was consistent across session formats. The analysis indicated strong model fit across groups (CFI = .957, TLI = .944, RMSEA = .094, SRMR = .036), supporting the model's adaptability and validity. Minimal changes in fit indices confirmed metric invariance, indicating conceptual consistency of the WAI-SR three-factor model across formats.

Metric invariance. Metric invariance was assessed by constraining factor loadings across groups to evaluate whether items contributed similarly to latent constructs. The model fit indices

indicated minimal changes: CFI decreased from .957 to .956 (ΔCFI = –.0002), TLI increased from .944 to .948 (ΔTLI = + .004), RMSEA decreased from .094 to .090 (ΔRMSEA = –.004), and SRMR increased slightly from .036 to .038 (ΔSRMR = + .002). These minimal changes confirm that item contributions are consistent across modalities, supporting metric invariance.

Scalar invariance. Scalar invariance was assessed by constraining item intercepts across groups to enable valid latent mean comparisons. Model fit indices showed minimal changes: CFI remained at .956 (ΔCFI = –.0002), TLI increased from .948 to .952 (ΔTLI = + .004), RMSEA decreased from .090 to .087 (ΔRMSEA = –.003), and SRMR remained at .038 (ΔSRMR = + .0004). These minimal changes support scalar invariance, indicating comparable item intercepts across groups.

Residual invariance. Strict invariance was assessed by constraining residual variances across groups to evaluate consistency in measurement error between session formats. The model fit indices were: CFI = .957, TLI = .957, RMSEA = .082, and SRMR = .038, indicating acceptable model fit with these added constraints. These results support strict invariance, demonstrating that the WAI-SR reliably assesses therapeutic alliance in both formats without bias.

Model Comparisons

The Satorra-Bentler scaled chi-square difference test comparing the metric (weak) invariance model to the configural model indicated no significant difference (χ^2 diff = 11.2, *df* diff = 9, $p = .261$), suggesting that factor loadings are consistent across groups. Changes in fit indices remained within acceptable thresholds, supporting metric invariance.

Similarly, the chi-square difference test comparing the scalar (strong) invariance model to the metric invariance model showed no significant difference (χ^2 diff = 10.7, *df* diff = 9, $p = .299$), indicating that both factor loadings and item intercepts are equivalent across groups, thus supporting scalar invariance.

For residual (strict) invariance, the chi-square difference test comparing the residual invariance model to the scalar invariance model also indicated no significant difference (χ^2 diff = 9.67, *df* diff = 12, $p = .645$). Changes in fit indices were negligible, confirming that residual variances are consistent across groups and supporting strict invariance.

Overall, these findings demonstrate that factor loadings, item intercepts, and residual variances are invariant across session formats, confirming the stability of the measurement properties of the instrument across groups.

Multiple Indicator Multiple Causes Model Analysis

A MIMIC model was used to evaluate measurement invariance while adjusting for age and gender (woman and man) between session formats. The Satorra-Bentler robust chi-square difference test comparing the unconstrained (configural) and constrained (metric + scalar) models showed no significant difference, $\Delta X^2_{(18)} = 16.1$, $p = .585$. These findings indicate that adjusting for age and gender did not significantly affect model fit, suggesting that the factor structure remains consistent across groups after accounting for these demographics.

Item Bias Analysis

Item bias was assessed using DIF and d_{MACS}. As reported in Table II, the DIF values ranged from –.089 to .036, and none of the estimates were statistically significant (all $p > .33$). Furthermore, all d_{MACS} values were below .20, indicating no differential item functioning or bias between groups.

Discussion

This study aimed to evaluate the factor structure and measurement invariance of the patient version of the WAI-SR across in-person and video call psychotherapy sessions. The findings confirmed the robust psychometric properties of the WAI-SR across both session formats, aligning with the growing body of research that supports the adaptability of psychological inventories in telepsychology settings (e.g., Stefana et al., 2024).

Our findings support the three-factor model of the working alliance, as originally theorized by Bordin (Bordin, 1979) and validated through the development of the WAI full (Horvath & Greenberg, 1989) and short (Hatcher & Gillaspy, 2006) forms. While this factor structure has been confirmed by prior validation studies, some research has proposed a two-factor model, combining the Goal and Task factors (see (Paap et al., 2022), for a review of WAI measurement properties). Notably, only two studies identifying the two-factor model as the best fit met the COnsensus-based Standards for the selection of health Measurements INstruments (COSMIN; (Mokkink et al., 2016)) criteria for sufficient structural validity. Both of these studies examined the

working alliance in teacher-student relationships within classroom settings (Knowles et al., 2020; Toste et al., 2015). This divergence underscores the significance of context and relational dynamics in conceptualizing the therapeutic alliance, highlighting the necessity of consistent validation across diverse populations and settings.

Latent mean analyses showed no significant differences in therapeutic alliance scores across session formats, with negligible effect sizes. These findings indicate that participants reported comparable alliance levels in both modalities, demonstrating that the core dimensions of the therapeutic alliance remain intact regardless of the delivery format. Internal consistency and composite reliability were excellent across formats, with Cronbach's α and McDonald's ω_{total} reaching or exceeding the threshold of .90 (Youngstrom et al., 2017) for all subscales and the total score.

The WAI-SR exhibited configural, metric, scalar, and residual invariance across in-person and video call formats. The same three-factor structure was consistent across both groups. Minimal adjustments in fit indices across subsequent models confirmed the stability of factor loadings, item intercepts, and residual variances between modalities. These findings underscore the reliability and robustness of the WAI-SR as a psychometric tool, particularly in the context of the rapidly evolving landscape of telehealth practices. Furthermore, the Satorra-Bentler scaled chi-square difference tests revealed no significant differences, strengthening evidence for measurement equivalence. The demonstrated invariance across age and gender groups further supports the generalizability of the WAI-SR for diverse populations.

The study's findings have significant implications for evidence-based assessment in videoconferencing psychotherapy for both clinical and research purposes, particularly as telehealth becomes increasingly prominent. The equivalence of WAI-SR total score and subscores between in-person and video call formats underscores the feasibility of reliably and validly assessing therapeutic alliance in online psychotherapy settings.

Clinical Implications

The findings hold significant implications for psychotherapy practice and research in the digital age. Clinicians and researchers can confidently utilize the WAI-SR to evaluate the therapeutic alliance in both in-person and telehealth settings without concern for measurement bias. This equivalence provides a dependable method for monitoring alliance quality across session formats, enabling the early detection of alliance-related issues that could impede treatment progress. Furthermore, the minimal differences observed between session formats reinforce that videoconferencing therapy is a viable and effective alternative to traditional face-to-face sessions, enhancing accessibility and flexibility in mental health care delivery.

Limitations and Future Directions

Despite its strengths, this study has limitations. The study focused exclusively on in-person and video call formats, which limits the generalizability of the findings to other teletherapy modalities, such as telephone-administered sessions. Future studies should broaden their scope to include telephone-based and asynchronous modalities, such as text-based therapy, to provide a more comprehensive understanding of the WAI-SR's applicability across diverse teletherapy methods. Additionally, the therapist version of the WAI-SR (Hatcher et al., 2019) should be tested for measurement invariance to ensure its reliability and validity across different psychotherapy formats.

Furthermore, investigating longitudinal patterns in therapeutic alliance across session formats could yield valuable insights into how these relationships evolve over time in telehealth contexts. Future research should also aim to validate the WAI-SR in culturally and demographically diverse populations, addressing potential disparities in digital literacy and access to technology that might influence perceptions of the alliance.

Conclusion

The WAI-SR is a reliable and valid tool for measuring therapeutic alliance across in-person and video call session formats. The equivalence in measurement properties and mean scores between modalities underscores the adaptability of therapeutic alliance to virtual environments, supporting the expanded use of telehealth in psychotherapy. These findings contribute to the growing body of evidence affirming the robustness of the WAI-SR and its utility in diverse clinical and research contexts.

In light of the COVID-19 pandemic and the ongoing digital transformation in healthcare, this study underscores the critical role of adaptable and robust tools like the WAI-SR in bridging the gap between traditional and remote therapeutic settings. As telehealth solidifies its position as a permanent feature of mental health care, ensuring the reliability of psychometric assessments across diverse formats will be vital for upholding high standards of care.

Acknowledgments

We would like to thank the Society for Psychotherapy Research (SPR) for allowing us to use the Working Alliance Inventory–Short Revised in our study. ChatGPT o1 was used to edit the language (grammar, syntax, clarity, and readability) of the original draft. No theoretical or statistical concepts were introduced or expanded, and no references were added through AI. Both the data and the analysis code that support the findings of this study are available from the corresponding author upon reasonable request.

Funding

The study has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 101030608.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Alberto Stefana http://orcid.org/0000-0002-4807-7184

References

Aafjes-van Doorn, K., Békés, V., Luo, X., & Hopwood, C. J. (2024a). Therapists' perception of the working alliance, real relationship and therapeutic presence in in-person therapy versus tele-therapy. *Psychotherapy Research*, *34*(5), 574–588. https://doi.org/10.1080/10503307.2023.2193299

Aafjes-van Doorn, K., Spina, D. S., Horne, S. J., & Békés, V. (2024b). The association between quality of therapeutic alliance and treatment outcomes in teletherapy: A systematic review and meta-analysis. *Clinical Psychology Review*, *110*, 102430. https://doi.org/10.1016/j.cpr.2024.102430

American Psychological Association. (2022). *Psychologists struggle to meet demand amid mental health crisis 2022—COVID-19 Practitioner Impact Survey*. Https://Www.Apa.Org. https://www.apa.org/pubs/reports/practitioner/2022-covid-psychologist-workload.

American Psychological Association. (2023). *Infographic: Percentage of psychologists offering remote or in-person treatment, 2020–2023*. https://www.apa.org/pubs/reports/practitioner/2023-infographics/telehealth-location.

American Psychological Association. (2024). *Proposed revision of guidelines for the practice of Telepsychology*. https://www.apa.org/practice/guidelines/telepsychology-revisions.pdf.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, *16*(3), 252–260. https://doi.org/10.1037/h0085885

Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology*, *12*(1), 53–81. https://doi.org/10.1146/annurev-clinpsy-021815-093623

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences, Rev. Ed (pp. xv, 474)*. Lawrence Erlbaum Associates, Inc.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Crowe, M., Inder, M., Manuel, J., & Carlyle, D. (2023). Characteristics of effective teletherapy for major depression: A systematic review. *Journal of Affective Disorders*, *327*, 175–182. https://doi.org/10.1016/j.jad.2023.02.019

Davis, K. A., Zhao, F., Janis, R. A., Castonguay, L. G., Hayes, J. A., & Scofield, B. E. (2024). Therapeutic alliance and clinical outcomes in teletherapy and in-person psychotherapy: A non-inferiority study during the COVID-19 pandemic. *Psychotherapy Research*, *34*(5), 589–600. https://doi.org/10.1080/10503307.2023.2229505

Falkenström, F., Hatcher, R. L., Skjulsvik, T., Larsson, M. H., & Holmqvist, R. (2015). Development and validation of a 6-item working alliance questionnaire for repeated administrations during psychotherapy. *Psychological Assessment*, *27*(1), 169–183. https://doi.org/10.1037/pas0000038

Faro, J. M., Nagawa, C. S., Orvek, E. A., Smith, B. M., Blok, A. C., Houston, T. K., Kamberi, A., Allison, J. J., Person, S. D., & Sadasivam, R. S. (2021). Comparing recruitment strategies for a digital smoking cessation intervention: Technology-assisted peer recruitment, social media, ResearchMatch, and smokefree.gov. *Contemporary Clinical Trials*, *103*, 106314. https://doi.org/10.1016/j.cct.2021.106314

Feng, H., Kurata, K., Cao, J., Itsuki, K., Niwa, M., Aoyama, A., & Kodama, K. (2024). Telemedicine Research Trends in 2001-2022 and research cooperation between China and other countries before and after the COVID-19 pandemic: Bibliometric analysis. *Interactive Journal of Medical Research*, *13*(1), e40801. https://doi.org/10.2196/40801

Feuerstahler, L. M., Waller, N., & MacDonald, A. 3rd. (2020). Improving measurement precision in experimental psychopathology using item response theory. *Educational and Psychological Measurement*, *80*(4), 695–725. PubMed-not-MEDLINE. https://doi.org/10.1177/0013164419892049

Flückiger, C., Del Re, A. C., Wlodasch, D., Horvath, A. O., Solomonov, N., & Wampold, B. E. (2020). Assessing the alliance–outcome association adjusted for patient characteristics and treatment processes: A meta-analytic summary of direct comparisons. *Journal of Counseling Psychology*, *67*(6), 706–711. https://doi.org/10.1037/cou0000424

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338

Gelso, C. J. (2011). *The real relationship in psychotherapy*. American Psychological Association.

Gelso, C. J. (2014). A tripartite model of the therapeutic relationship: Theory, research, and practice. *Psychotherapy Research*, *24*(2), 117. https://doi.org/10.1080/10503307.2013.845920

Gelso, C. J., Kelley, F. A., Fuertes, J. N., Marmarosh, C., Holmes, S. E., Costa, C., & Hancock, G. R. (2005). Measuring the real relationship in psychotherapy: Initial validation of the therapist

form. *Journal of Counseling Psychology*, *52*(4), 640. https://doi.org/10.1037/0022-0167.52.4.640

Gelso, C. J., Kivlighan, D. M., & Markin, R. D. (2018). The real relationship and its role in psychotherapy outcome: A meta-analysis. *Psychotherapy*, *55*(4), 434–444. https://doi.org/10.1037/pst0000183

Giovanetti, A. K., Punt, S. E. W., Nelson, E.-L., & Ilardi, S. S. (2022). Teletherapy versus in-person psychotherapy for depression: A meta-analysis of randomized controlled trials. *Telemedicine and E-Health*, *28*(8), 1077–1089. https://doi.org/10.1089/tmj.2021.0294

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*(11 Suppl 3), S78–S94. https://doi.org/10.1097/01.mlr.0000245454.12228.8f

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*, *30*(6), 706–717. https://doi.org/10.1080/10503307.2019.1677964

Hill, E. M., An, M., Kivlighan, D. M., & Gelso, C. J. (2024). A tripartite model of the psychotherapy relationship: Interrelations among its components and their unfolding across sessions. *Psychotherapy*, *61*(2), 151–160. https://doi.org/10.1037/pst0000521

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, *36*(2), 223–233. https://doi.org/10.1037/0022-0167.36.2.223

Hoyle, R. H., Ed. (2023). *Handbook of structural equation modeling* (2nd ed.). The Guilford Press.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., Rosseel, Y., Miller, P., Quick, C., Garnier-Villarreal, M., Selig, J., Boulton, A., Preacher, K., Coffman, D., Rhemtulla, M., Robitzsch, A., Enders, C., Arslan, R., Clinton, B., Panko, P., Merkle, E., Chesnut, S., … Johnson, A. R. (2022). *semTools: Useful tools for structural equation modeling* (Version 0.5-6) [Computer software]. https://cran.r-project.org/web/packages/semTools/index.html.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, *70*(351a), 631–639. https://doi.org/10.1080/01621459.1975.10482485

Kelley, F. A., Gelso, C. J., Fuertes, J. N., Marmarosh, C., & Lanier, S. H. (2010). The Real Relationship Inventory: Development and psychometric investigation of the client form. *Psychotherapy: Theory, Research, Practice, Training*, *47*(4), 540. https://doi.org/10.1037/a0022082

Knowles, C., Murray, C., Gau, J., & Toste, J. R. (2020). Teacher–student working alliance among students with emotional and behavioral disorders. *Journal of Psychoeducational Assessment*, *38*(6), 753–761. https://doi.org/10.1177/0734282919874268

Krzyzaniak, N., Greenwood, H., Scott, A. M., Peiris, R., Cardona, M., Clark, J., & Glasziou, P. (2024). The effectiveness of telehealth versus face-to face interventions for anxiety disorders: A systematic review and meta-analysis. *Journal of Telemedicine and Telecare*, *30*(2), 250–261. https://doi.org/10.1177/1357633X211053738

Leitgöb, H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., De Roover, K., Jak, S., Meitinger, K., Menold, N., Muthén, B., Rudnev, M., Schmidt, P., & van de Schoot, R. (2023). Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social Science Research*, *110*, 102805. https://doi.org/10.1016/j.ssresearch.2022.102805

Lin, T., Heckman, T. G., & Anderson, T. (2022). The efficacy of synchronous teletherapy versus in-person therapy: A meta-analysis of randomized clinical trials. *Clinical Psychology: Science and Practice*, *29*(2), 167–178. https://doi.org/10.1037/cps0000056

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The Hull Method for selecting the number of common factors. *Multivariate Behavioral Research*, *46*(2), 340–364. https://doi.org/10.1080/00273171.2011.564527

Luong, R., & Flake, J. K. (2023). Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. *Psychological Methods*, *28*(4), 905–924. https://doi.org/10.1037/met0000441

Maassen, E., D'Urso, E. D., Van Assen, M. A. L. M., Nuijten, M. B., De Roover, K., & Wicherts, J. M. (2023). The dire disregard of measurement invariance testing in psychological science. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000624

Mishkind, M., Shore, J. H., Barrett, R., Caudill, R., Chiu, A., Hilty, D., Idigo, O. B., Kaftarian, E., Khan, S., Krupinski, E. A., Malik, T. S., Thackaberry, J., Torous, J., & Yellowlees, P. (2024). Resource document on best practices in synchronous videoconferencing-based Telemental health. *Telemedicine and e-Health*, *30*(5), 1330–1340. https://doi.org/10.1089/tmj.2023.0174

Mokkink, L. B., Prinsen, C. A. C., Bouter, L. M., de Vet, H. C. W., & Terwee, C. B. (2016). The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian Journal of Physical Therapy*, *20*(2), 105–113. https://doi.org/10.1590/bjpt-rbf.2014.0143

Neumann, A., König, H.-H., Bokermann, J., & Hajek, A. (2023). Determinants of patient use and satisfaction with synchronous Telemental Health Services during the COVID-19 Pandemic. *Systematic Review. JMIR Mental Health*, *10*(1), e46148. https://doi.org/10.2196/46148

Norcross, J. C., & Lambert, M. J. (2019). *Psychotherapy relationships that work (3rd ed). Volume 1: Evidence-based therapist contributions*. Oxford University Press.

Nye, C. D., Bradburn, J., Olenick, J., Bialko, C., & Drasgow, F. (2019). How big are my effects? Examining the magnitude of effect sizes in studies of measurement equivalence. *Organizational Research Methods*, *22*(3), 678–709. https://doi.org/10.1177/1094428118761122

Nye, C. D., & Drasgow, F. (2011). Effect size indices for analyses of measurement equivalence: Understanding the practical importance of differences between groups. *Journal of Applied Psychology*, *96*(5), 966–980. https://doi.org/10.1037/a0022955

Paap, D., Karel, Y. H. J. M., Verhagen, A. P., Dijkstra, P. U., Geertzen, J. H. B., & Pool, G. (2022). The Working Alliance Inventory's measurement properties: A systematic review. *Frontiers in Psychology*, *13*, 945294. https://doi.org/10.3389/fpsyg.2022.945294

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team. (2024). *R: A language and environment for statistical computing. R Foundation for Statistical Computing. Https:*//www.R-project.org/ [Computer software].

Rosseel, Y. (2012). **lavaan**: An *R* Package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, *24*(2), 282–292. https://doi.org/10.1037/a0025697

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, *29*(4), 347–363. https://doi.org/10.1177/0734282911406661

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514. https://doi.org/10.1007/BF02296192

Seuling, P. D., Fendel, J. C., Spille, L., Göritz, A. S., & Schmidt, S. (2024). Therapeutic alliance in videoconferencing psychotherapy compared to psychotherapy in person: A systematic review and meta-analysis. *Journal of Telemedicine and Telecare*, *30*(10), 1521–1531. https://doi.org/10.1177/1357633X231161774

Stefana, A., Damiani, S., Granziol, U., Provenzani, U., Solmi, M., Youngstrom, E., & Fusar-Poli, P. (2025b). Psychological, psychiatric, and behavioral sciences measurement scales: Best practice guidelines for their development and validation. *Frontiers in Psychology*, *15*, 1494261. https://doi.org/10.3389/fpsyg.2024.1494261

Stefana, A., Fusar-Poli, P., Vieta, E., Gelso, C. J., & Youngstrom, E. A. (2024). Development and validation of an 8-item version of the Real Relationship Inventory–client form. *Psychotherapy Research*, *35*(3), 395–411. https://doi.org/10.1080/10503307.2024.2320331

Stefana, A., Fusar-Poli, P., Vieta, E., & Youngstrom, E. A. (2023). Therapeutic relationship elements and therapy session outcomes: Protocol for a longitudinal study of the patient's perspective. *Open Research Europe*, *3*, 133. https://doi.org/10.12688/openreseurope.16466.2

Stefana, A., Fusar-Poli, P., Vieta, E., & Youngstrom, E. A. (2025a). Evaluating the psychometric properties of the 24-item and 12-item Real Relationship Inventory-Client forms. *PLoS One*, *20*(3), e0311411. https://doi.org/10.1371/journal.pone.0311411

Stefana, A., Vieta, E., Fusar-Poli, P., & Youngstrom, E. A. (2024). Enhancing psychotherapy outcomes by encouraging patients to regularly self-monitor, reflect on, and share their affective responses toward their therapist: Protocol for a randomized controlled trial. *JMIR Research Protocols*, *13*, e55369. https://doi.org/10.2196/55369

Stefana, A., Youngstrom, E. A., Hopwood, C. J., & Dakanalis, A. (2020). The COVID-19 pandemic brings a second wave of social isolation and disrupted services. *European Archives of Psychiatry and Clinical Neuroscience*, *270*(6), 785–786. https://doi.org/10.1007/s00406-020-01137-8

Steiner, M., & Grieder, S. (2020). EFAtools: An R package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software*, *5*(53), 2521. https://doi.org/10.21105/joss.02521

Toste, J. R., Heath, N. L., McDonald Connor, C., & Peng, P. (2015). Reconceptualizing teacher-student relationships: Applicability of the working alliance within classroom contexts. *The Elementary School Journal*, *116*(1), 30–48. https://doi.org/10.1086/683110

Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, *14*(3), 270–277. https://doi.org/10.1002/wps.20238

Wampold, B. E. (2017). What should we practice? A contextual model for how psychotherapy works. In T. Rousmaniere, R. Goodyear, S. D. Miller, & B. E. Wampold (Eds.), *The cycle of excellence* (1st ed, pp. 49–65). Wiley. https://doi.org/10.1002/9781119165590.ch3

Wampold, B. E., & Flückiger, C. (2023). The alliance in mental health care: Conceptualization, evidence and clinical applications. *World Psychiatry*, *22*(1), 25–41. https://doi.org/10.1002/wps.21035

Wirth, E. v., Willems, S., Döpfner, M., & Kohl, L. T. (2023). Effectiveness of videoconference-delivered psychotherapy for children, adolescents, and their parents: A meta-analysis of randomized controlled trials. https://doi.org/10.1177/1357633X231199784

Youngstrom, E. A., Van Meter, A., Frazier, T. W., Hunsley, J., Prinstein, M. J., Ong, M., & Youngstrom, J. K. (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*, *24*(4), 331–363. https://doi.org/10.1111/cpsp.12207