Current Psychology
https://doi.org/10.1007/s12144-021-01721-8

Psychometric properties and longitudinal invariance of the session rating scale in Chinese clinical samples

Zhuang She[1,2] · Yanwei Shi[3] · Barry L. Duncan[4] · Dong Xie[5] · Juzhe Xi[1,2] · Qiwu Sun[6] · Weidong Ji[2]

Accepted: 6 April 2021

Abstract

The Session Rating Scale (SRS) is a widely used clinical tool to measure the client-therapist working alliance. This study investigated the psychometric properties, the cut-off value, and longitudinal invariance of the SRS in a Chinese clinical population. The analyses were conducted separately in a sample of college students in counseling ($n = 403$) and in a sample of clients in outpatient therapy ($n = 246$). Participants completed the Outcome Rating Scale (ORS) before each session and the SRS after each session. To test concurrent validity, a subset of participants also completed the Working Alliance Questionnaire at the end of each session. In both samples, the results indicated high internal consistency ($\alpha = .92 \sim 0.97$) and adequate test-retest reliability over four sessions (university sample: $r = .69 \sim .78$; outpatient sample: $r = .52 \sim .66$). Construct validity was evident in a one-factor structure, and concurrent validity was established based on a strong correlation with the Working Alliance Questionnaire ($r = .64 \sim .70$). In addition, regression analysis indicated that early alliance ratings (at the third session) on the SRS predicted post-ORS scores. The cut-off value for the SRS in the Chinese context was established as 34. The longitudinal measurement invariance was tested by a longitudinal confirmatory factor analysis. Full scalar invariance of the SRS was supported. This study supported the use of the SRS in China and that a single-factor solution stayed stable over time, providing preliminary evidence for subsequent mean comparisons.

Keywords Session rating scale (SRS) · Client feedback · Working alliance · Psychometric properties · Cross-cultural · Longitudinal invariance

Introduction

The therapeutic alliance is an essential element of psychotherapy. Meta-analysis has consistently shown a moderate but robust positive correlation between the alliance and outcomes across clients, therapy modality, and clinical problems (Fluckiger et al., 2018; Fluckiger et al., 2020). More specifically, clients with higher alliance scores are more likely to achieve better therapeutic improvement (Fluckiger et al., 2020), while those with a weaker alliance are more likely to end up with an inferior outcome or drop out of therapy early (Sharf et al., 2010). The relationship of the alliance with successful outcomes suggests that therapists may enhance treatment outcomes by monitoring and adjusting the alliance in the course of treatment (Duncan, 2014).

One commonly used instrument to assess and monitor the therapeutic alliance is the Session Rating Scale (SRS; Duncan et al., 2003). Although the SRS can be used independently, it is part of a systematic client feedback intervention, the Partners for Change Outcome Management System (PCOMS; Duncan, 2012). PCOMS employs two, four-item scales: one focuses on outcome (the Outcome Rating Scale) and the other assesses the therapeutic alliance (the SRS). It includes a real-time collaborative discussion of client views of outcome and alliance. With this information, providers and

✉ Juzhe Xi
jzxi@psy.ecnu.edu.cn

1 Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention, Affiliated Mental Health Center (ECNU), School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China

2 Shanghai Changning Mental Health Center, Shanghai 200335, China

3 Department of Human Resource Management, Shanghai Normal University, Shanghai 200234, China

4 Better Outcomes Now, West Palm Beach, FL 33401, USA

5 Department of Psychology and Counseling, University of Central Arkansas, Conway, AR 72035, USA

6 School of Psychology, Central China Normal University, Wuhan 430079, China

Published online: 14 April 2021

clients have an opportunity to shift focus, re-visit goals, or alter interventions before deterioration or dropout. PCOMS, an evidence-based practice, has been shown to improve outcomes in nine randomized controlled trials (RCT) (Duncan & Reese, 2015; Duncan & Sparks, 2019).

The SRS is based on Bordin's (1979) classic delineation of the components of the alliance: the relational bond and the degree of agreement between the client and therapist about the goals and tasks of therapy (Duncan et al., 2003). Unlike other alliance measures, many of which are used for research purposes, the SRS was designed as a convenient assessment tool in clinical settings. It takes only two minutes to administer and score, minimizing the burden on both clients and therapists (Duncan, 2012). In addition, administering the SRS at each session encourages clients to report real-time alliance problems, dissatisfaction with the therapeutic process, and dissatisfaction with the therapist (Duncan et al., 2003). Therapists, then, have opportunities to improve the therapeutic relationship by identifying and repairing ruptures (Reese et al., 2009). Thus, the scale is a useful clinical tool in everyday clinical practice.

Given its brevity, a concern often arises regarding the reliability and validity of SRS generated scores. The clinical utility of ultra-brief measures can only be realized when they have sufficient reliability and validity (Crits-Christoph et al., 2011). Currently, psychometric studies of the SRS have been conducted in several countries, including the United States (US; Duncan et al., 2003), Australia (Campbell & Hemsley, 2009), the Netherlands (Janse et al., 2014), Spain (Moggia et al., 2020) and France (Cazauvieilh et al., 2020). Overall, these studies supported the basic psychometric properties of the SRS despite its brevity (Duncan, 2012; Murphy et al., 2020).

There are still questions, however, regarding the SRS. First, the construct validity of the SRS has not been adequately examined. Duncan et al. (2003) initially suggested that the SRS was a unidimensional scale like other alliance measures but they did not provide any empirical support for the unidimensional structure of SRS. Through exploratory factor analysis, Quirk et al. (2013) suggested a one-latent factor structure underlying the 4-item Group Session Rating Scale (the group version of the SRS), but no confirmatory analysis has been conducted. A recent study found that the SRS demonstrated a single factor solution; however, the small sample size limited the generalization of results (Cazauvieilh et al., 2020). Thus, one of the purposes of this study is to confirm the factor structure of SRS with a larger sample size.

In addition, the existing psychometric studies were limited to western countries. Murphy et al. (2020) suggested that a critical direction for future research should be to focus on cultural validation and include participants from non-western countries. Given that the SRS has been translated to over 31 languages and is widely used in clinical research and practice around the world, more studies are needed to validate its use in non-Western cultures. Therefore, the current study aimed to examine the reliability and validity of SRS in China.

Another question relevant to cross-cultural use of the SRS is about the cut-off score. The cut-off score for the US SRS was identified as 36 (Duncan & Reese, 2015). Clients who scored below this value are at a greater risk of negative outcome from treatment (Duncan & Reese, 2015). Although the US cutoff was determined more by a cautious clinical perspective than statistical analysis, several studies outside of the US found that the mean score for SRS was lower than the US standard (Hafkenscheid et al., 2010; Janse et al., 2014; She et al., 2018). For example, in the Dutch samples, the mean scores on the SRS were never more than 34 points during treatment, and the global mean for the SRS was identified as 32.4 points, which is far below the cut-off value of the US standard (Hafkenscheid et al., 2010; Janse et al., 2014). One study conducted in China found similar results; the average last SRS score was 33.71, still well below the American's cut-off score (She et al., 2018). Given the variability, researchers have suggested that the cut-off score of SRS should be determined based on specific cultures (Janse et al., 2014). To fill this gap, one goal of the present study was to determine a cut-off score for Chinese SRS.

In addition to testing the psychometric properties of the SRS in a Chinese sample, this investigation also evaluated a characteristic of the SRS that has not even been tested in the Western literature. In addition to testing whether the SRS has a one-factor structure, we also analyzed whether this factor structure is invariant over time, known as longitudinal measurement invariance (Vandenberg & Lance, 2000). Longitudinal invariance may have the same importance as reliability and validity (Vandenberg & Lance, 2000) because if the factor structure of a scale changes over time, inferences based on the results of tracking studies may be inaccurate (Horn & McArdle, 1992). In other words, the establishment of longitudinal measurement invariance is a prerequisite for making direct mean comparisons over time (Horn & McArdle, 1992). Given that the SRS is widely used in tracking studies without evidence of longitudinal invariance, the present study seeks to fill this gap and has important implications for use of SRS in both clinical practice and tracking studies on the alliance in psychotherapy research.

Addressing the noted gaps in the literature, this study evaluated the psychometric properties of the Chinese-language version of the SRS using an adult outpatient sample as well as a separate sample of university students receiving counseling services in university counseling centers. Specifically, this investigation tested the reliability, validity, and factor

invariance of the Chinese-language SRS, and established the clinical cut-off value.

Method

Participants

University Sample ($n = 403$) The students receiving services in university counseling centers came from two sources. The first source was a university counseling center where 119 students (98 females, 21 males) volunteered to participate. The average age was 21.4 years ($SD = 4.43$; range: 18 ~ 42). In this sample, therapists did not have access to the clients' completed questionnaires. The second source included university students who had participated in a previous randomized control trial (RCT) on the use of PCOMS in a college sample (She et al., 2018). There were 310 clients who enrolled in the RCT, but 26 cases were removed from the analysis because of incomplete data. The final sample in the RCT included 284 clients (228 female, 54 male, 2 unspecified) with an average age of 21.85 years ($SD = 2.70$; range: 18 ~ 32). All the clients in this RCT received regular treatment, with the only difference between groups being whether clients' ORS/SRS feedback was provided to their therapists. There were 150 participants in the feedback group (therapists received clients' ORS/SRS feedback) and 134 participants in the no-feedback group (therapists did not receive clients' ORS/SRS feedback). More details can be found in She et al. (2018).

The two samples were merged to increase the power of analysis (n = 403). All the participants were Chinese college students (undergraduates or graduates). This combined sample was used to conduct all the analyses in this study except that the feedback group from the RCT was included only in the analyses used to determine the cut-off value (see below).

Outpatient Sample ($n = 246$) This sample included 246 clients (165 women, 78 men, 3 unspecified; an average age of 28.1 years, $SD = 9.24$; range: 18 ~ 58) who were receiving counseling services at a large community counseling center. Almost 16% of clients had more than a university degree, 55% had a college degree, and 29% had a high school degree or below. About 50% were unmarried, 44% were married, and 6% were divorced. In this sample, clients were asked to complete pre-session ORS and post session SRS for each session. A feedback procedure was implemented in this sample, similar to the one used in the feedback arm in the RCT described above.

Procedure

New clients were invited to participate in this study at intake. All participants were informed about the purpose of the study and its risks and benefits, and they volunteered to participate. In both samples, clients were asked to complete a paper version of the Chinese-language ORS at the beginning of each session and the Chinese SRS at the end of each session. From the first source, the university sample, 119 clients also complete another well-established alliance scale, the Working Alliance Questionnaire (WAQ), along with the SRS, after each session. All measures were given by administrative staff separate from the therapy encounter. Clients received counseling once a week. This research was approved by all clinical sites.

Measures

The Session Rating Scale (SRS) is a 4-item visual analogue scale to measure the working alliance between the therapist and client (Duncan et al., 2003). Clients are asked to rate the relationship with their therapists, agreement on goals and topics, methods, and overall experience by using a 10-cm line with markers from left (negative) to right (positive). The total score ranges from 0 to 40, with higher scores indicating a better therapeutic alliance.

The original translation of the SRS from English to Chinese is available at betteroutcomesnow.com. With permission from Barry. L Duncan (the first author of SRS) (Duncan et al., 2003), we slightly altered the Chinese translation to make the language more concise, but without changing its intended meaning. This version of Chinese SRS was used in the previous RCT study (She et al., 2018).

Working Alliance Questionnaire (WAQ) is a Chinese-language 12-item questionnaire for measuring three dimensions of the working alliance in therapy (WAQ, Zhu & Jiang, 2011). Each dimension is assessed with 4 items: relationship bond, goal-task, and client engagement. Clients were asked to rate each item on a five-point Likert scale ranging from 1 (rarely) to 5 (always). The total score of WAQ is from 12 to 60, with a higher score indicating a better working alliance. The authors reported that the internal consistencies for each dimension were above .70. WAQ also has satisfactory concurrent validity ($r = .86$) with the Working Alliance Inventory-Short Revised (WAI-SR; Hatcher & Gillaspy, 2006), and good construct and predictive validity (Zhu & Jiang, 2011).

The Outcome Rating Scale (ORS) is another 4-item visual analogue scale developed as a brief outcome measure to track treatment progress (Miller et al., 2003). The ORS asks clients to rate their subjective experience on four domains: individual (symptomatic distress), interpersonal (family or any close relationship), social (work, school or friends), and overall (overall well-being) in the past week. Clients place a mark on a 10 cm line corresponding to each domain. The more to the left end indicates a higher level of distress, and the more to the right end indicates a lower level of distress. Similar to

the SRS, a centimeter ruler (or web-based system) is used to measure the distance between the left end and clients' marks, which indicates the score on each item. The four items total score from 0 to 40, with a higher score indicating less distress.

The reliability and validity of the ORS have been verified in many studies (Duncan, 2012). The Chinese version of the ORS has good internal consistency ($\alpha = .84$), test-retest reliability ($r = .60$), and moderate concurrent validity with the longer Outcome Questionnaire, similar to other psychometric studies of the ORS (She et al., 2017).

Statistical Analyses

Descriptive statistics, internal consistency (α), test-retest reliability (Pearson's r), and tests of concurrent validity (Pearson's r) and predictive validity (linear regression) were performed with SPSS 21.0. Tests of construct validity and longitudinal invariance were performed with Mplus 7.0.

For the CFA analysis and longitudinal invariance, we used indicators of the comparative fit index (CFI), the Tucker Lewis index (TLI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR) to evaluate model fit. The values of RMSEA below .06, SRMR below .08, and CFI and TLI above .95 were considered as a good fit (Hu & Bentler, 1999).

Given the SRS data was not a completely normal distribution, the parameter estimates for longitudinal invariance analysis were obtained using the maximum likelihood Method (MLM). ▵CFI and ▵RMSEA values were used to assess the model fit with increased constraints. ▵CFI was less than .01 and ▵RMSEA was lower than .015, indicating a good fit (Chen, 2007).

Results

Reliability

Internal consistency was evaluated with Cronbach's α coefficient. Based on data collected in the first four sessions, the alpha values ranged from .92 to .96 in the university student sample, and from .94 to .97 in the outpatient sample. Pearson correlations were used to calculate test-retest reliability over four sessions based on correlations between two successive total SRS scores. Tables 1 and 2 present the test-retest reliability in both the university and outpatient samples.

Concurrent Validity

The concurrent validity was assessed with Pearson correlations between the SRS and the WAQ. The SRS total scores were strongly correlated with WAQ total scores (Table 3) at each of the first four sessions.

Table 1 Test-retest reliability of SRS between four sessions in university sample

	1st-2nd		2nd-3rd		3rd-4th	
	n	*r*	*n*	*r*	*n*	*r*
SRS	278	.69***	229	.78***	114	.77***

Abbreviation: SRS, Session Rating Scale

*** $p < .001$

The decline in *n* from the intake value is due to missing data or clients who did not attend more sessions (the same below)

Predictive Validity

To test the predictive value of the SRS, we utilized linear regression to examine if early SRS total scores (the third ratings) could predict treatment outcome (final ORS). We selected the SRS rating at the third time point because the therapeutic alliance tends to stabilize from the third session on (Fluckiger et al., 2018), and most previous studies also used the third alliance score to predict treatment outcomes (Janse et al., 2014; Moggia et al., 2020). Thus, using the SRS rating at the third time point would allow us to compare the results with those of other studies. The initial symptom severity (baseline ORS total score) and treatment length (the number of sessions) were controlled in the analysis. SRS total score at third session significantly predicted post-ORS in both university sample ($R^2 = .097$, $\beta = .42$, $p < .001$) and outpatient sample ($R^2 = .084$, $\beta = .39$, $p < .01$).

Construct Validity

Quirk et al. (2013) found that the SRS indicated one latent factor structure. To examine this result, we conducted a confirmatory factor analysis (CFA) in both the university and outpatient samples with Mplus 7.0. Table 4 presents the fit indices for the single-factor model using the data at session one. The results showed that the one-factor model had a good fit with the data in both groups. Factor loadings on the four SRS items were in the range of .74 ~ .96 in the college sample .77 ~ .92 in the outpatient sample.

Table 2 Test-retest reliability of SRS between four sessions in outpatient sample

	1st-2nd		2nd-3rd		3rd-4th	
	n	*r*	*n*	*r*	*n*	*r*
SRS	145	.54***	107	.52***	74	.66***

Abbreviation: SRS, Session Rating Scale

*** $p < .001$

Table 3 The total correlations between SRS and WAQ

Session 1 (n=108)	Session 2 (n=70)	Session 3 (n=53)	Session 4 (n=42)
.70***	.64***	.64***	.69***

Abbreviation: SRS, Session Rating Scale; WAQ, Working Alliance Questionnaire

*** $p < .001$

Longitudinal Measurement Invariance

Because the sample size decreased over time, we used the first six sessions to examine longitudinal invariance. Using listwise deletion, at the 6th session 95 cases were remaining in the university sample, and 50 cases remaining in the outpatient sample. This is consistent with previous findings that clients usually have a small number of visits in a naturalistic setting (Anker et al., 2010; Youn et al., 2020). Given that a small sample size may result in inaccurate evaluation, we used the larger group (95 cases in the university sample) to conduct the longitudinal invariance analysis (Falkenstrom et al., 2015). This sample size was close to the recommend sample size (n = 100) to conduct the longitudinal measurement invariance in the previous study (Falkenstrom et al., 2015).

Longitudinal invariance analyses were conducted with each successive step including more constraints. In the first step, a baseline model was estimated without any constraints to test whether the SRS has equal factor structure across sessions—configural invariance. The results indicated a good fit to the data (see Table 5), supporting testing for a more restrictive levels of invariance. The second step was to test whether the SRS has equal factor loadings across sessions—metric invariance. The results showed a non-significant deterioration compared with the configural model (△CFI = .002, △RMSEA < .001). The third step was to examine the scalar invariance, that is whether the SRS has equal item intercepts across sessions. The results indicated a non-significant deterioration compared with the metric model (△CFI = .002, △RMSEA< .001). Thus, full scalar invariance was supported. The fourth step was to evaluate the error variance model. However, results indicated there was a significant deterioration in model fit compared with the scalar model (△CFI = .063, △RMSEA = .039). We then terminated the analysis.

Table 4 One-factor model fit statistics for SRS at session one in college and outpatient samples

	χ^2	df	CFI	TLI	SRMR	RMSEA [90%CI]
College sample	0.83	2	1.00	1.00	.00	.00[.00 .08]
Outpatient sample	2.73	2	.99	.99	.01	.04[.00 .14]

Abbreviations: χ^2, chi-square value; df = degrees of freedom; CI, confidence interval; CFI, comparative fit Index; TLI, Tucker–Lewis index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; CI, confidence interval

College sample, N_1 = 365; Outpatient sample, N_2 = 231

Cut-off Scores

Given that the SRS cut-off score was made for clinical purpose (alliance feedback and discussion), we only determined the SRS cut-off value in the feedback condition. There were 396 clients in the feedback (feedback arm in the RCT plus outpatient sample) who provided SRS ratings throughout therapy, creating a total of 2309 data points. In line with prior work (Cazauvieilh et al., 2020; Moggia et al., 2020), the cut-off value for SRS was evaluated by calculating the 25th percentile of all scores at all administrations in the feedback condition (n = 2309). The SRS score at the 25th percentile was 30 (29.90). This result was far below the cut-off score of 36 that has been reported in Western countries (Cazauvieilh et al., 2020; Moggia et al., 2020). Given the importance of the working alliance, as well as clients' tendency to give high ratings on alliance measures (Duncan, 2014), we considered using the 50th percentile (a more rigorous standard) to obtain the cut-off value for the Chinese SRS (Duncan & Reese, 2015). The SRS score at the 50th percentile was 34 (34.30), which was comparable to the Dutch normative sample (34) (Janse et al., 2014), but still lower than the cut-off score (36) from studies in the United States (Duncan & Reese, 2015), Spain (Moggia et al., 2020) and France (Cazauvieilh et al., 2020).

Discussion

This study tested the psychometric properties and longitudinal measurement invariance of a Chinese version of the Session Rating Scale (Duncan et al., 2003), a measure of the therapist-client working alliance in therapy. We found that the SRS had adequate reliability and validity in the Chinese context, a culture different from the western countries in which the scale was developed and tested. The results were similar in both a sample of college counseling clients and in a sample of community clinic outpatients. This is the first study to test the psychometric properties of SRS outside of the western countries, as well as the first attempt to establish the longitudinal invariance of SRS.

Table 5 The longitudinal measurement invariance analyses for the first six sessions($n = 95$)

Model	$S\text{-}Bx^2$	df	$\triangle\chi^2$	$\triangle df$	p	CFI	TLI	RMSEA [90% CI]
Configural invariance	254.304	177				.969	.952	.068[.048, .086]
Metric invariance	272.341	189	18.037	12	>.05	.967	.952	.068[.049, .086]
Scalar invariance	292.723	204	20.382	15	>.05	.965	.952	.068[.049, .084]
Error variance invariance	469.838	224	177.115	20	<.05	.902	.879	.107[.094, .121]

Abbreviations: χ^2, chi-square value; *df*, degrees of freedom; CI, confidence interval;
CFI, comparative fit Index; TLI, Tucker–Lewis index; RMSEA, root mean square error of approximation; CI, confidence interval

The results showed that SRS had strong internal consistency ($\alpha = .92 \sim .97$) in both outpatient and university samples, consistent with previous studies indicating an average alpha value of .92 (see Sparks & Duncan, 2018 for a review). The test-retest reliability of the SRS was also adequate in both samples. In the present study, the test-retest reliability was slightly lower ($r = .52 \sim .66$) in the outpatient sample. Given that the SRS is intended to measure variability in client experience of the alliance, we would expect changes in session to session ratings and therefore lower test-retest reliability as well as some variation across samples. Overall, the test-retest reliability found in this study was comparable to that found in a study in the US (with an overall r of .64) (Duncan et al., 2003) and two other studies in the Netherlands (with an overall r of .51) (Hafkenscheid et al., 2010; Janse et al., 2014).

Results also indicated high concurrent validity of the SRS with another established measure of the therapist-client alliance in therapy, the Working Alliance Questionnaire. Our results indicated that the SRS was strongly correlated with the WAQ ($r = .64 \sim .70$). These results were comparable to previous findings, indicating moderate to strong concurrent validity of SRS (Duncan et al., 2003; Janse et al., 2014). This suggests that the SRS and WAQ may largely measure a similar construct. Given the SRS is more user-friendly (Duncan, 2012), it could be considered as an alternative alliance measure to the longer alliance questionnaires in clinical practice.

The predictive validity of the SRS was also supported in the present study. The SRS total score at session three explained 9.7% and 8.4% of the outcome variance in college and outpatient samples, respectively. This is consistent with studies indicating that an early alliance predicts treatment outcome (Anker et al., 2010; Duncan et al., 2003; Martin et al., 2000). The predictive value for SRS in this study is larger than Spanish's study ($R^2 = .04$) (Moggia et al., 2020) and Dutch's study ($R^2 = .02$) (Janse et al., 2014). One possible reason is that the current study used ORS as an outcome measure, while the previous studies assessed treatment outcome with other symptoms-specific questionnaires (e.g., SCL-90, CORE-OM) (Janse et al., 2014; Moggia et al., 2020). Future studies can use different outcome measures to examine the predictive validity of SRS.

In addition, our study demonstrated a one-factor structure for SRS in both university and outpatient samples. These results support the assumption proposed by Duncan et al. (2003) that SRS should be regarded as a global measure of the alliance. This provides evidence for the construct validity of the SRS.

Moreover, the findings also provide the first evidence of the longitudinal invariance of the one-factor solution of the SRS over the first six sessions, with regard to factor structure, factor loadings, and item intercepts. The results did not provide evidence of error invariance (strict equivalence), which is a more difficult standard to meet. However, strict equivalence is not a prerequisite for testing mean differences, which instead requires metric invariance (equal factor loadings) and scalar invariance (equal item intercepts) (Putnick & Bornstein, 2016). The current results suggest that clients' interpretation of the SRS items remains the same over time, and the SRS as an alliance measure can be applied to make mean comparisons in brief longitudinal studies. This is consistent with the position that even brief alliance measures can have good psychometric properties and longitudinal invariance (Falkenstrom et al., 2015; Hukkelberg & Ogden, 2016). Now longitudinal invariance of another brief alliance measure, the SRS, has been established.

Although the SRS had a cut-off score of 36 based on the 25th percentile in the original US sample, a meaningful cut-off may differ from culture to culture (Janse et al., 2014). Later analysis on a larger US database also reported a cut off score of 36, but based on the 50th percentile (Duncan & Reese, 2015) and a cautious clinical perspective. Taking into account potential cultural differences, in the present study a relatively stricter standard (the 50th percentile) was also used to identify a cut-off score of 34 for the Chinese SRS. Interestingly, while it is believed that East Asians may inflate their SRS ratings because they are part of a collectivist culture (e.g., impression management) (Lalwani et al., 2006), the cut-off score for SRS in the Chinese sample (34) fell below most of the cut-off scores from Western countries, even when a more rigorous

standard was used. This implies that Chinese clients may use a different standard comparing with Westerners to rate the items on the SRS. When applied clinically, we advise Chinese therapists to discuss the alliance if the SRS total score is less than 34 or if any of the four items has a rating below 8.5.

Several limitations of this study should be noted. First, the use of only the first six sessions in the longitudinal invariance analysis limits the conclusions that may be drawn. This number of sessions is smaller than in a similar study that tested the longitudinal measurement invariance of the 6-item Session Alliance Inventory (SAI) over ten sessions (Falkenstrom et al., 2015). Nevertheless, given that in our samples the average number of sessions was typically small in both the college and outpatient samples (the mean number of sessions was 4.69 and 4.36, respectively), six sessions may represent the majority of sessions in the course of therapy. In addition, only the college sample was examined for invariance analysis in this study. Future studies can conduct a similar analysis in outpatient samples.

Another key limitation was that the university sample contained both feedback and no-feedback groups, while the outpatient sample included only a feedback group. One concern is that this mixture of clients in the university sample might have lowered the reliability of the SRS, because feedback tends to facilitate the development of an alliance (Janse et al., 2014). However, the SRS demonstrated adequate psychometric properties in both university and outpatient samples. This suggests that the SRS is a reliable clinical tool even when ratings are discussed from week to week in therapy.

Conclusion

With these limitations in mind, this study confirmed the SRS as a brief measure of the client-therapist alliance measure as valid in both western and non-western cultures. The SRS demonstrated a good "fit" in China, perhaps because therapy there is still largely based on western models that emphasize the client-therapist alliance (Duan et al., 2015). In addition, this is the first attempt to empirically test the longitudinal measurement invariance of the SRS. These promising findings provide contributions to future research and clinical practice. It is our hope that more clinical settings in China can now realize the benefits of alliance monitoring.

Funding The present study was supported by the National Social Science Foundation of China (18BSH129), the Shanghai Philosophy and Social Science Planning Project (2017BSH004), the Shanghai Pujiang Program of China (12PJ037), the Research Project of Shanghai Science and Technology Commission (20DZ2260300), the Science and Technology Innovation Action Plan of Shanghai Science and Technology Commission (20DZ2304400).

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Informed Consent Informed consent (written) was obtained from all individual participants included in the study.

Conflict of Interest Barry L. Duncan is a co-holder of the copyright of the Partners for Change Outcome Management System (PCOMS) instruments (the Outcome and Session Rating Scales). The measures are free for individuals but Duncan receives royalties from licenses issued to groups and organizations. In addition, the web-based application of PCOMS, BetterOutcomesNow.com, is a commercial product and he receives profits based on sales.

Research Involving Human Participants All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments. This research was also approved by all clinical sites where the study was conducted.

References

Anker, M. G., Owen, J., Duncan, B. L., & Sparks, J. A. (2010). The alliance in couple therapy: Partner influence, early change, and alliance patterns in a naturalistic sample. *Journal of Consulting and Clinical Psychology, 78*(5), 635–645. https://doi.org/10.1037/a0020051.

Campbell, A., & Hemsley, S. (2009). Outcome rating scale and session rating scale in psychological practice: Clinical utility of ultra-brief measures. *Clinical Psychologist, 13*(1), 1–9. https://doi.org/10.1080/13284200802676391.

Cazauvieilh, C., Gana, K., Miller, S. D., & Quintard, B. (2020). Validation of the French versions of two brief, clinician-friendly outcome monitoring tools: The ORS and SRS. *Current Psychology*. https://doi.org/10.1007/s12144-020-00992-x.

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.

Crits-Christoph, P., Gibbons, M. B. C., Hamilton, J., Ring-Kurtz, S., & Gallop, R. (2011). The dependability of alliance assessments: The alliance-outcome correlation is larger than you might think. *Journal of Consulting and Clinical Psychology, 79*(3), 267–278. https://doi.org/10.1037/a0023668.

Duan, C., Hill, C. E., Jiang, G., Hu, B., Lei, Y., Chen, J., & Yu, L. (2015). The counselor perspective on the use of directives in counseling in China: Are directives different in China as in the United States? *Counseling Psychology Quarterly, 28*(1), 57–77. https://doi.org/10.1080/09515070.2014.965659.

Duncan, B. L. (2012). The Partners for Change Outcome Management System (PCOMS): The heart and soul of change project. *Canadian Psychology-Psychologie Canadienne, 53*(2), 93–104. https://doi.org/10.1037/a0027762.

Duncan, B. L. (2014). *On becoming a better therapist: Evidence-based practice one client at a time (2nd ed)*. American Psychological Association.

Duncan, B. L., & Reese, R. J. (2015). The partners for change outcome management system (PCOMS) revisiting the client's frame of

reference. *Psychotherapy, 52*(4), 391–401. https://doi.org/10.1037/pst0000026.

Duncan, B., & Sparks, J. (2019). *The Partners for Change Outcome Management System: An integrated elearning manual for everything PCOMS*. Author.

Duncan, B. L., Miller, S. D., Sparks, J. A., Claud, D. A., Reynolds, L. R., Brown, J., & Johnson, L. D. (2003). The session rating scale: Preliminary psychometric properties of a "working" alliance measure. *Journal of Brief Therapy, 3*(1), 3–12.

Falkenstrom, 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.

Fluckiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy, 55*, 316–340. https://doi.org/10.1037/pst0000172.

Fluckiger, 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*. Advance online publication*, 67*, 706–711. https://doi.org/10.1037/cou0000424.

Hafkenscheid, A., Duncan, B. L., & Miller, S. D. (2010). The outcome and session rating scales: A cross-cultural examination of the psychometric properties of the Dutch translation. *Journal of Brief Therapy, 7*(1), 1–12.

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.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*(3), 117–144. https://doi.org/10.1080/03610739208253916.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

Hukkelberg, S. S., & Ogden, T. (2016). The short working Alliance inventory in parent training: Factor structure and longitudinal invariance. *Psychotherapy Research, 26*(6), 719–726. https://doi.org/10.1080/10503307.2015.1119328.

Janse, P., Boezen-Hilberdink, L., van Dijk, M. K., Verbraak, M., & Hutschemaekers, G. J. M. (2014). Measuring feedback from clients. *European Journal of Psychological Assessment, 30*(2), 86–92. https://doi.org/10.1027/1015-5759/a000172.

Lalwani, A. K., Shavitt, S., & Johnson, T. (2006). What is the relation between cultural orientation and socially desirable responding? *Journal of Personality and Social Psychology, 90*(1), 165–178. https://doi.org/10.1037/0022-3514.90.1.165.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68*(3), 438–450. https://doi.org/10.1037/0022-006x.68.3.438.

Miller, S. D., Duncan, B., Brown, J., Sparks, J., & Claud, D. (2003). The outcome rating scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of Brief Therapy, 2*(2), 91–100.

Moggia, D., Niño-Robles, N., Miller, S, D., & Feixas, G. (2020). Psychometric properties of the session rating scale 3.0 in a Spanish clinical sample. *British Journal of Guidance & Counselling*, 1–12. https://doi.org/10.1080/03069885.2020.1778635.

Murphy, M. G., Rakes, S., & Harris, R. M. (2020). The psychometric properties of the session rating scale: A narrative review. *Journal of Evidence-Based Social Work, 17*, 279–299. https://doi.org/10.1080/26408066.2020.1729281.

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.

Quirk, K., Miller, S., Duncan, B., & Owen, J. (2013). Group session rating scale: Preliminary psychometrics in substance abuse group interventions. *Counselling and Psychotherapy Research, 13*(3), 194–200. https://doi.org/10.1080/14733145.2012.744425.

Reese, R. J., Norsworthy, L. A., & Rowlands, S. R. (2009). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy: Theory, Research, Practice. Training, 46*(4), 418–431. https://doi.org/10.1037/a0017901.

Sharf, J., Primavera, L. H., & Diener, M. J. (2010). Dropout and therapeutic alliance: A meta-analysis of adult individual psychotherapy. *Psychotherapy, 47*(4), 637–645. https://doi.org/10.1037/a0021175.

She, Z., Sun, Q.-W., & Jiang, G.-R. (2017). Reliability and validity of Chinese version of outcome rating scale. *Chinese Journal of Clinical Psychology, 25*(2), 272–275. https://doi.org/10.16128/j.cnki.1005-3611.2017.02.017.

She, Z., Duncan, B. L., Reese, R. J., Sun, Q. W., Shi, Y. W., Jiang, G. R., et al. (2018). Client feedback in China: A randomized clinical trial in a college counseling center. *Journal of Counseling Psychology, 65*(6), 727–737. https://doi.org/10.1037/cou0000300.

Sparks, J. A., & Duncan, B. L. (2018). The Partners for Change Outcome Management System: A both/and system for collaborative practice. *Family Process., 57*(3), 800–816. https://doi.org/10.1111/famp.12345.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002.

Youn, S. J., Castonguay, L. G., McAleavey, A. A., Nordberg, S. S., Hayes, J. A., & Locke, B. D. (2020). Sensitivity to change of the counseling center assessment of psychological Symptoms-34. *Measurement and Evaluation in Counseling and Development, 53*(2), 75–88. https://doi.org/10.1080/07481756.2019.1691459.

Zhu, X., & Jiang, G. (2011). Development of the working alliance questionnaire. *Chinese Journal of Clinical Psychology, 19*, 449–454 CNKI:SUN:ZLCY.0.2011–04-008.