

University of Groningen

Reducing ceiling effects in the Working Alliance Inventory-Rehabilitation Dutch Version

Paap, Davy; Schepers, Melvin; Dijkstra, Pieter U

Published in:
Disability and Rehabilitation

DOI:
[10.1080/09638288.2018.1563833](https://doi.org/10.1080/09638288.2018.1563833)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Paap, D., Schepers, M., & Dijkstra, P. U. (2020). Reducing ceiling effects in the Working Alliance Inventory-Rehabilitation Dutch Version. *Disability and Rehabilitation*, 42(20), 2944-2950.
<https://doi.org/10.1080/09638288.2018.1563833>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.






Reducing ceiling effects in the Working Alliance Inventory-Rehabilitation Dutch Version

Davy Paap, Melvin Schepers & Pieter U. Dijkstra


To cite this article: Davy Paap, Melvin Schepers & Pieter U. Dijkstra (2019): Reducing ceiling effects in the Working Alliance Inventory-Rehabilitation Dutch Version, Disability and Rehabilitation, DOI: [10.1080/09638288.2018.1563833](https://doi.org/10.1080/09638288.2018.1563833)

To link to this article: <https://doi.org/10.1080/09638288.2018.1563833>

 View supplementary material 

 Published online: 24 Jan 2019.

 Submit your article to this journal 

 Article views: 13

 View Crossmark data 

Reducing ceiling effects in the Working Alliance Inventory-Rehabilitation Dutch Version

Davy Paap^{a,b*}, Melvin Schepers^{c,d*} and Pieter U. Dijkstra^{a,e}

^aDepartment of Rehabilitation Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; ^bPhysiotherapy Practise, Inter-Fysio, Groningen, The Netherlands; ^cDepartment of Clinical Health Science, Faculty of Medicine, University of Utrecht, Utrecht, The Netherlands; ^dPhysiotherapy Practise, PrengerHoekman, Winschoten, The Netherlands; ^eDepartment of Oral and Maxillofacial Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

ABSTRACT

Purpose: To reduce ceiling effects on domain scores (Task, Goal, and Bond) of the Working Alliance Inventory-Rehabilitation Dutch Version by changing response scales and using Visual Analogue Scales (VASs).

Methods: Clients, who had at least three treatment sessions prior, randomly received one of the three versions of the WAI-Rehabilitation Dutch Version, using items with a balanced Likert scale, Positive-Packed Likert scale or VAS. Primary outcome was percentage of ceiling effects in total- and domain scores, secondary outcomes were construct validity and internal consistency of the three versions.

Results: One hundred and seventy-six clients randomly received a set of questionnaires (one of the three versions of the WAI-Rehabilitation Dutch Version, Session Rating Scale (SRS) and Helping Alliance Questionnaire (HAQ-II)); 152 participants (mean age 51.5 ± 16.3 , 106 women) returned the questionnaires. No ceiling effects were present in the total scores of all versions. Significantly fewer ceiling effects were found in the VAS-Version (Goal: 8.0%, Bond: 7.7%) compared to the original (Goal: 18.0%, Bond: 29.8%) and Positive-Packed Version (Goal: 27.1%, Bond: 29.8%). Spearman's correlations between VAS-Version, SRS and HAQ-II ranged 0.747–0.845.

Conclusions: Visual Analogue Scales effectively reduced ceiling effects on domain scores of the WAI-Rehabilitation Dutch Version, while maintaining validity.

ARTICLE HISTORY

Received 19 September 2017
Revised 21 December 2018
Accepted 22 December 2018

KEYWORDS

Rehabilitation; therapeutic alliance; ceiling effect; response scale; reliability

► IMPLICATIONS FOR REHABILITATION

- Visual Analogue Scales effectively reduced ceiling effects on domain scores of the Working Alliance Inventory-Rehabilitation Dutch Version, while maintaining construct validity.
- The Working Alliance Inventory version with Visual Analogue Scales can be used in rehabilitation.



Introduction

The therapeutic alliance reflects the relationship between the (rehabilitation) client and the (rehabilitation) professional and includes three aspects: (1) agreement between client and professional on the goal(s) of treatment; (2) agreement between client and professional about the tasks to achieve the proposed goal(s); and (3) the quality of the bond between client and professional [1]. The therapeutic alliance is a negotiated and collaborative characteristic of the relationship, enabling clients to achieve their desired treatment goals [2]. The therapeutic alliance is derived from Freud's theory of transference and countertransference [3], and is the essential ingredient in promoting therapeutic change independent of the treatment modality [1].


Therapeutic alliance has been studied extensively in different areas of psychotherapy, with studies finding a positive association with satisfaction, quality of life [4], psychology well-being [5] and symptom improvement [6,7]. Growing evidence suggests that these associations exist also within rehabilitation [8]. In the

rehabilitation context, a strong therapeutic alliance contributes to higher client satisfaction as well as reduction of pain and disability in clients with chronic diseases [9,10]. Nonetheless, the therapeutic alliance has not been investigated systematically in rehabilitation, as evidenced by the lack of consensus regarding measurement instruments used [8,11–13].

The Working Alliance Inventory (WAI) is a valid instrument, commonly used to measure therapeutic alliance in psychotherapy [13]. To measure the therapeutic alliance within rehabilitation, the Working Alliance Inventory Rehabilitation Dutch Version (WAI-ReD) was recently developed [14]. The WAI-ReD is scored on a balanced five-point Likert scale and has similar clinimetric properties as the short version of the WAI [14,15]. However, the WAI-ReD has ceiling effects across all domain scores ranging between 16% and 33% and maximum scores occurred in 9% of the total scores [14]. In the Brazilian version of the WAI, intended for use in rehabilitation, ceiling effects occurred in 26% of the total scores [16]. A systematic review revealed that WAI scores are high in

CONTACT Davy Paap  d.paap@rug.nl  Department of Rehabilitation Medicine, University of Groningen, Hanzeplein 1, P.O. Box 30001, Groningen 9713 GZ, The Netherlands

*Both the authors contributed equally to this work.

 Supplemental data for this article can be accessed [here](#).

© 2019 Informa UK Limited, trading as Taylor & Francis Group

most studies, suggesting possible ceiling effects. However, these effects have not explicitly been reported in clinimetric studies [8].

Ceiling effects are present when more than 15% of clients achieve the highest possible score and these clients' scores cannot be distinguished from each other [17,18]. Ceiling effects reduce accurate interpretation of data [18]. These effects may also indicate that the response scale of an instrument is not comprehensive [17]. Moreover, ceiling effects affect responsiveness of an instrument because the highest scores cannot elevate [19,20]. Therefore, it is important to adjust the instrument or measurement procedures when ceiling effects occur [11].

The high therapeutic alliance scores measured with WAI-ReD should be interpreted carefully [14]. While these scores may suggest that the majority of the clients had a strong therapeutic alliance with their professional, biases such as social desirability and/or response tendencies, may also affect WAI-ReD scores. Technically, ceiling effects are an instrumental issue, whereby the therapeutic alliance as measured by the WAI-ReD is not sensitive enough to discriminate high therapeutic alliance scores among clients [21].

To address ceiling effects, high-end scale labels can be expanded by adding an extra option between the last two response options, or by adding an even "better"/"more positive" label to the far end of the spectrum [21]. This expansion is based on the idea that "average" is not in the middle of a rating scale [21,22], and is relevant when responses are expected to be mostly above average. By moving the average label toward the negative end of the rating scale, it creates more room to "pack" the rating scale with mostly above average labels [21]. A positive labelled five-point Likert scale generally prevents ceiling effects better than a balanced Likert scale [18,21–23]. In addition, a five-point Likert scale is less likely to produce higher means as compared to scales with fewer or more response options [24,25].

Another strategy used to reduce ceiling effects is to replace a balanced Likert scale with Visual Analogue Scale (VAS) [26–28]. When responding to a discrete Likert-scaled item, the client chooses one of the given options, but when answering to a continuous VAS item the client indicates a position between the lower and upper endpoints, which best represents the clients' opinion. When measuring treatment satisfaction, a VAS prevents ceiling effects more effectively than a Likert Scale [26].

At present, no studies have investigated methods to reduce the ceiling effects of the WAI-ReD or the original WAI version. Therefore, the aim of this study was to modify the WAI-ReD response scales by changing labels, utilizing VAS, and to analyse the ceiling effects of three versions of the WAI-ReD. The primary outcome of this study was to determine the percentage of ceiling effects on the total- and domain scores of the WAI-ReD and the two modified versions. Secondary outcomes sought in this study were the differences in total- and domain scores among the three versions of the WAI-ReD, and construct validity and internal consistency of three versions of the WAI-ReD.

We tested the following four hypotheses based on previous research: (1) a modified WAI-ReD with positive-packed Likert Scales (WAI-ReD^{PP}) and a modified version with VAS (WAI-ReD^{VAS}) will result in a lower percentage of ceiling effects on the total- and domain scores compared to the original balanced Likert scales of the WAI-ReD [14]; (2) the WAI-ReD^{PP} and WAI-ReD^{VAS} will result in lower mean scores on the total- and domain scores compared to the balanced Likert scale; (3) the strength of the correlations between the scores of the Helping Alliance Questionnaire (HAQ-II)/Session Rating Scale (SRS) and the modified versions of the WAI-ReD are expected to be ≥ 0.600 ; and (4) Cronbach's

alphas of the total and domains scores of the three different versions of the WAI-ReD are expected to be ≥ 0.700 [14].

Methods

Participants and recruitment

Participants were included if they had received at least three prior treatment sessions with their rehabilitation professional, were 18 years or older and had sufficient knowledge of the Dutch language to complete the questionnaires correctly. Participants were excluded if they experienced aphasia or were unable to read or write.

Clients from Department of Rehabilitation of the University Medical Center Groningen (UMCG) were recruited by rehabilitation professionals (hand therapists, physiotherapist, psychologists, psychomotor therapists, and speech therapists), between 19 December 2016 and 20 March 2017. During the intended recruitment period, the target sample could not be obtained; therefore, additional participants were recruited from physiotherapy practices in the area of Groningen. At those sites, the recruitment period was between 1 February 2017 and 20 March 2017. Participants recruited by physiotherapists in private practices were added to the group of participants recruited by physiotherapists within the department of rehabilitation of the UMCG.

Measures

Two additional versions of the WAI-ReD were constructed [18,21–28]. The WAI-ReD (original balanced Likert scale) scores were compared to WAI-ReD^{PP} and WAI-ReD^{VAS} scores (Supplementary material 1).

WAI-ReD with a balanced Likert scale. The WAI-ReD is a questionnaire for measuring therapeutic alliance between a client and rehabilitation professional, consisting of 12 items [14]. Items are responded to based on a five-point Likert scale. The scale is a balanced Equal-Interval Rating scale with labels; 1: Never, 2: Sometimes, 3: Often, 4: Very often; 5: Always. The scores of items of the WAI-ReD are added to calculate domain scores: Task (items 1, 2, 10, and 12), Goal (items 4, 6, 8, and 11), and Bond (items 3, 5, 7, and 9). The sum of the three domain scores reflects the therapeutic alliance. Confirmatory factor analysis of the WAI-ReD resulted in an acceptable fit for a model with three factors [14]. Internal consistency of the WAI-ReD domains, expressed as Cronbach's alpha (α), ranged between 0.804 and 0.927 [14]. Construct validity was determined through correlations, Pearson's rho, between the WAI-ReD and similar validated instruments for measuring therapeutic alliance including the SRS and the HAQ-II. Validity evidence was strong (ranging between 0.698 and 0.734) [14].

Modified WAI-ReD^{PP}. The WAI-ReD^{PP} items are identical to the original WAI-ReD items with the exception of scale options. The label "Never" was removed and the label "Almost Always" was added. The positive-packed response scale includes: 1: Sometimes, 2: Often, 3: Very often, 4: Almost always, and 5: Always [18,21–23].

Modified WAI-ReD^{VAS}. The items of the WAI-ReD^{VAS} are identical to the original WAI-ReD with the exception of the response option. The VAS is a horizontal line, 100 mm in length, anchored by worded descriptors at each end [26–28]. Clients were asked to draw a vertical line across the horizontal line that best represented their opinion. For this study, a VAS was used with the following labelled anchors: "Sometimes" (0) and "Always" (100) [18,21–23].

Table 1. Rehabilitation client characteristics and rehabilitation professionals by WAI-ReD version.

Variable	Overall	WAI-ReD	WAI-ReD ^{VAS}	WAI-ReD ^{PP}
Participants (<i>n</i>)	152	52	52	48
Female (<i>n</i> (%)) ^a	106 (70)	34 (65)	39 (75)	33 (69)
Mean age (SD) ^b	51.5 (16.3)	50.6 (16.5)	52.4 (13.7)	51.4 (18.9)
Health condition (<i>n</i> (%))				
Orthopaedic	96 (63.2)	29 (55.8)	35 (67.3)	32 (66.7)
Neurological	18 (11.8)	8 (15.4)	5 (9.6)	5 (10.4)
Psychosomatic	13 (8.6)	8 (15.4)	5 (9.6)	5 (10.4)
Cardiovascular	5 (3.3)	1 (1.9)	3 (5.8)	1 (2.1)
Autoimmune	7 (4.6)	3 (5.8)	2 (3.8)	2 (4.2)
Other	7 (4.6)	4 (7.7)	2 (3.8)	1 (2.1)
Unknown	6 (3.9)	2 (3.8)	0 (0.0)	4 (8.3)
Clients per discipline (<i>n</i>)				
Hand therapist	40	13	14	13
Psychologist	18	6	6	6
Psychomotor therapist	7	3	3	1
Physiotherapist	84	27	29	28
Speech therapist	3	3	0	0
Complete cases (<i>n</i>)	117	38	46	33

WAI-ReD: Working Alliance Inventory-Rehabilitation Dutch Version; ^{PP}: with positive packed labels; ^{VAS}: with Visual Analogue Scales; SD: standard deviation; *n*: number; %: column percentages; complete cases: participants who fully completed one of the versions of the WAI-ReD, Session Rating Scale and Helping Alliance Questionnaire-II.

^aMissing values gender: WAI-ReD *n* = 2, WAI-ReD^{VAS} *n* = 1, WAI-ReD^{PP} *n* = 3.

^bMissing values age: WAI-ReD *n* = 1, WAI-ReD^{VAS} *n* = 0, WAI-ReD^{PP} *n* = 1.

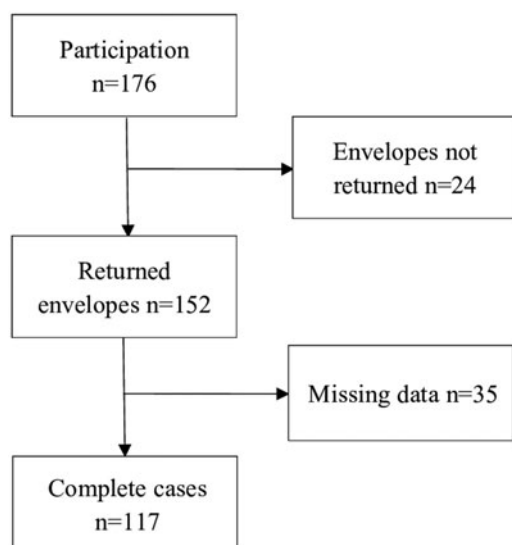


Figure 1. Flowchart data-collection.

SRS. The SRS is a self-report 4-item questionnaire that measures the strength of the therapeutic alliance [29]. Each item is rated on a VAS. Internal consistency (α) of the Dutch version was 0.92 [30]. In this study, the internal consistency (α) was 0.92. The correlation between the SRS and the HAQ-II was 0.48 [29].

HAQ-II. The HAQ-II is a widely used 19-item questionnaire that measures the strength of the therapeutic alliance [31]. Each item is rated on a six-point Likert scale. The HAQ-II showed good internal consistency, with a coefficient α of 0.92. In this study, the internal consistency (α) was 0.89. The HAQ-II demonstrated high convergence validity with the California Psychotherapy Alliance Scale ($r = 0.59-0.69$) [31].

Procedures

This study was approved by the local medical ethical review committee (METC2016.b12). Prior to participant recruitment, rehabilitation professionals were informed verbally and in writing about

the research project. During the first treatment session, within the recruitment period, professionals informed clients about the research. Professionals informed clients that they were blinded from participant scores. For those clients who met the inclusion criteria and chose to participate in the study, written informed consent was obtained. The professionals were asked to record how many clients met inclusion criteria and how many agreed to participate. After the third treatment session, and after signing the informed consent form, participants were given a sealed and opaque envelope containing one of the three versions of the WAI-ReD, the SRS, the HAQ-II, and a form to document gender, age, treatment reason and the name of the rehabilitation professional. Participants randomly received one of the three versions of the WAI-ReD. Stratified block randomization was applied. Each block consisted of three envelopes in a random sequence. Stratification was based on the rehabilitation discipline, because significant differences have been found between different rehabilitation disciplines [14].

To calculate the sample size, α was set at 0.05 and power was set at 0.80. The percentage of ceiling effects in Bond scores of the WAI-ReD was 33% [6]. The percentage of ceiling effects of the domain Bond for the modified versions of the WAI-ReD was estimated to be 8% [23], resulting in a sample size of 40 participants for each of the three groups. A similar study amongst rehabilitation clients within the Department of Rehabilitation of the UMCG showed less than 10% missing values [14]. To compensate for missing data, we aimed for a sample of 132 participants.

Statistical analysis

All statistical analyses were performed using SPSS software version 24 (SPSS Inc., Chicago, IL). The Shapiro-Wilk test and QQ-plots showed that scores of the different versions of the WAI-ReD were not normally distributed. A Chi-square test was used to analyse differences in ceiling effects between the three versions of the WAI-ReD and differences in the number of complete cases between the three versions. Complete cases were operationalized as clients who fully completed the SRS, HAQ-2 and one of the versions of the WAI-ReD.

Table 2. Percentage of ceiling effects and median (interquartile range) on the total- and domain scores per WAI-ReD version.

Ceiling effects (%)	WAI-ReD	WAI-ReD ^{VAS}	WAI-ReD ^{PP}	T.St	df	<i>p</i> Value	ES
Total score	6.4%	4.1%	10.9%	1.717	2	0.424	0.110 ^a
Domain score							
Task	12.0%	10.2%	14.9%	0.496	2	0.780	0.058 ^a
Goal	18.0%	8.0%	27.1%	6.168	2	0.046	0.204 ^a
Bond	29.8%	7.7%	29.8%	9.550	2	0.008	0.256 ^a
	<u>Median (IQR)</u>	<u>Median (IQR)</u>	<u>Median (IQR)</u>				
Total score	925 (825;1100)	1048 (931;1137)	975 (819;1125)	5.341	2	0.069	0.125 ^b
Domain score							
Task	300 (244;350)	337 (307;376)	324 (250;375)	7.763	2	0.021	0.170 ^b
Goal	325 (275;375)	364 (327;388)	350 (281;400)	4.654	2	0.098	0.106 ^b
Bond	300 (275;400)	340 (303;378)	325 (275;400)	0.821	2	0.663	0.000 ^b

WAI-ReD: Working Alliance Inventory-Rehabilitation Dutch Version; ^{PP}: with positive packed labels; ^{VAS}: with Visual Analogue Scales; *p*: probability; IQR: inter quartile range; %: column percentages; T.St: test statistic; df: degrees of freedom; ES: effect size, for Chi-square test Cramer's *V* and for the Kruskal-Wallis test *r*.

^aBased on Chi-square test.

^bBased on the Kruskal-Wallis test.

Table 3. Correlations (Spearman's ρ) of WAI-ReD versions with Session Rating Scale and Helping Alliance Questionnaire 2 and Cronbach's α of the total score and domain scores per WAI-ReD version.

	WAI-ReD	WAI-ReD ^{VAS}	WAI-ReD ^{PP}
	<u>ρ</u>	<u>ρ</u>	<u>ρ</u>
SRS	0.552*	0.845*	0.404
HAQ-2	0.638*	0.747*	0.647*
	<u>α</u>	<u>α</u>	<u>α</u>
Total score	0.870	0.928	0.898
Domain score			
Task	0.865	0.764	0.706
Goal	0.661	0.915	0.773
Bond	0.715	0.903	0.790

WAI-ReD: Working Alliance Inventory-Rehabilitation Dutch Version; ^{PP}: with positive packed labels; ^{VAS}: with Visual Analogue Scales; SRS: Session Rating Scale; HAQ-2: Helping Alliance Questionnaire 2.

*Significant at 0.01 level.

Likert scaled items were converted to a 0–100 scale to enable comparison with VAS scores using the formula $((\text{Likert scale score} - \text{min}) / (\text{max} - \text{min})) \times 100$, where "min" is the lowest and "max" the highest possible score on the Likert scale [27,32]. The highest possible outcome on the domain- and total scores of the WAI-ReD are respectively "400" and "1200". The Kruskal-Wallis test was used to analyse differences in scores between the three versions of the WAI-ReD, and between clients of different rehabilitation professionals. The Mann-Whitney *U* test was used to analyse differences in scores between participants recruited by physiotherapists in the UMCG and those recruited in physiotherapy practices.

To determine construct validity, Spearman's ρ was calculated for outcomes of the SRS, HAQ-II and the scores of the versions of the WAI-ReD. Cronbach's α of the total scores and domains scores of the three versions of the WAI-ReD were calculated to determine internal consistency. A probability (*p*)-value of ≤ 0.05 was considered statistically significant.

Complete case analyses and a multiple imputation analyses for each version of the WAI-ReD were used to create two different datasets. For the complete case method, all collected data were analysed. The multiple imputation method was added due to missing values. A separate multiple imputation method was used for each version of the WAI-ReD to prevent variation in results from the different versions of the WAI-ReD. Both datasets were analysed and results were compared.

Results

Of the 176 clients who signed informed consent, 152 (response rate 86%) returned their envelopes of whom 95 were treated within the rehabilitation department of the UMCG and 57 in physiotherapy practices (see Table 1 and Figure 1).

Regarding missing values significantly more clients completed (no missing data) the WAI-ReD^{VAS} (89%) as compared to WAI-ReD (73%) and the WAI-ReD^{PP} (69%) (Chi-square test $\chi^2(2) = 6.149$; $p = 0.046$; $\text{phi}(\phi) = 0.201$). A significantly larger number of clients of the WAI-ReD^{VAS} group (98%) completed the SRS compared to WAI-ReD group (85%) and the WAI-ReD^{PP} group (77%) (Chi-square test $\chi^2(2) = 5.054$; $p = 0.080$; $\text{phi}(\phi) = 0.188$, Supplementary material 2).

Regarding primary outcome, significantly less ceiling effects were found in the WAI-ReD^{VAS} Goal scores (Chi-square test $\chi^2(2) = 6.168$; $p = 0.046$; $\text{phi}(\phi) = 0.204$), and Bond scores as compared to the WAI-ReD and the WAI-ReD^{PP} (Chi-square test $\chi^2(2) = 9.550$; $p = 0.008$; $\text{phi}(\phi) = 0.256$, Table 2). No significant differences between the total scores were found (Chi-square test $\chi^2(2) = 1.711$; $p = 0.424$; $\text{phi}(\phi) = 0.110$). For pairwise comparisons analyses, see Supplementary material 3.

Regarding the secondary outcomes, significantly higher scores were found in the WAI-ReD^{VAS} Task scores as compared to the WAI-ReD and the WAI-ReD^{PP} (the Kruskal-Wallis test $H(2) = 7.763$; $p = 0.021$; $r = 0.170$, Table 1). For pairwise comparisons analyses, see Supplementary material 3. Construct validity and internal consistency of the WAI-ReD versions are summarized in Table 3. The total- and domain scores of the WAI-ReD for the different disciplines did not differ significantly (the Kruskal-Wallis test; Total: $H(4) = 3.575$; $p = 0.467$; $r = 0.006$, Task: $H(4) = 6.168$; $p = 0.187$; $r = 0.073$, Bond: $H(4) = 3.514$; $p = 0.476$; $r = 0.005$, Goal: $H(4) = 3.687$; $p = 0.450$; $r = 0.011$, see Supplementary material 4 and Figure 2). Participants recruited in the UMCG by physiotherapists had significantly lower total- and domain scores compared to participants recruited in physiotherapy practices by physiotherapist (the Mann-Whitney *U* test; Total: $U = 2929.5$; $p = 0.013$; 95% CI = -15.3 ; -140.9 , Task: $U = 3142.5$; $p = 0.006$; 95% CI = -9.5 ; -60.4 , Bond: $U = 3137.0$; $p = 0.010$; 95% CI = -4.2 ; -50.9 , Goal: $U = 2925.5$; $p = 0.008$; 95% CI = -4.9 ; -47.0 , Supplementary material 5).

The multiple imputation analyses showed that the percentage of clients with the highest scores (ceiling effects) were lower for the WAI-ReD^{VAS} in the total score and domain scores. For the Bond scores, the difference was significant (Chi-square test $\chi^2(2) = 9.841$; $p = 0.007$; $\text{phi}(\phi) = 0.254$, Supplementary material 6). Additionally, the median scores of the WAI-ReD^{VAS} were

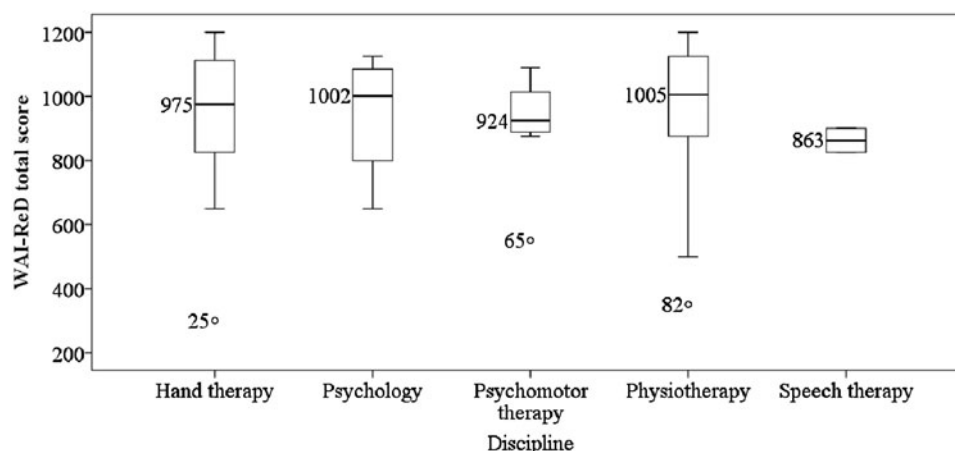


Figure 2. Boxplot total score WAI-ReD per discipline. The median scores per discipline of the total scores of the WAI-ReD did not differ significantly from each other (Kruskal–Wallis's test $p = 0.467$).

significantly higher than those of the other WAI-ReD versions for the total score (the Kruskal–Wallis test $H(2) = 6.196$; $p = 0.045$; $r = 0.133$).

Discussion

Hypothesis one, stating that the WAI-ReD^{PP} and WAI-ReD^{VAS} will result in a lower percentage of ceiling effects on the total- and domain scores compared to the balanced Likert scales of the WAI-ReD, was confirmed for the WAI-ReD^{VAS} and rejected for the WAI-ReD^{PP}. Hypothesis one was confirmed for the WAI-ReD^{VAS} since significantly less ceiling effects were found in the WAI-ReD^{VAS} for the domains Bond and Goal compared to the WAI-ReD and WAI-ReD^{PP}. Ceiling effects were still present on the domains Bond and Goal of the WAI-ReD^{PP}. Ceiling effects were absent in the total scores of all three versions of the WAI-ReD. We expected less ceiling effects and lower means for the WAI-ReD^{PP} because the expansion of the high-end scale labels better differentiates “above-average” ratings from what is “outstanding” [18]. In addition, fewer ceiling effects and lower means were expected for the WAI-ReD^{VAS} because there are more scoring options on a continuous VAS [26–28]. The respondent can score any outcome between “0” and “100” and is less likely to choose the highest possible score, compared to a balanced five-point Likert scale.

Although previous studies provide some evidence suggesting the benefit of using rating scales loaded with positive labels, in the current study, the WAI-ReD^{PP} did not prevent ceiling effects [21–23]. One explanation for ceiling effects on the two domains of the WAI-ReD^{PP} is a tendency of participants to give social desirable answers [33]. Another possibility is that clients may have answered the questions without reading the Likert scale labels adequately or without giving a value to these labels. The latter may have occurred because the WAI-ReD and the positive-packed version look quite similar with both a five-point Likert scale. This similarity might also explain the similar outcomes regarding ceiling effects in the total- and individual domain scores.

Hypothesis two, stating that the WAI-ReD^{PP} and WAI-ReD^{VAS} will result in lower mean scores on the total- and domain scores compared to the balanced Likert scales was rejected. Scores on the modified versions were higher than those of the original version. However, these results were not significant. Previous studies support our findings showing that a VAS can prevent ceiling effects better than a balanced Likert scale [26–28]. However, these studies did not find significant differences in VAS scores as compared to the Likert scale scores [26–28]. We modified the VAS by

including more positive-packed anchors expecting the WAI-ReD^{VAS} to have significantly lower scores as compared to the WAI-ReD [16,17]. However, the results of our study showed that median scores are higher on the WAI-ReD^{VAS} as compared to the original version. This finding corresponds to results of a study analysing client satisfaction in which a positive-packed method did not reduce median scores [21].

The therapeutic alliance scores are high for the domain- and total scores of the WAI-ReD and the two modified versions. There is some evidence suggesting that rating scales with construct-related anchors (anchors that are specific to the items) may produce greater response variability than rating scales with generic anchors [34]. Furthermore, the order of the questions of the WAI-ReD may have influenced clients' answers to the other questions. The first items may give a positive frame for the following items in the questionnaire which may have influenced the total score of the WAI-ReD. Further research is needed to explore the use of construct-related rating scales and the influence of the order effect on therapeutic alliance scores of the WAI-ReD.

Another mechanism that might have influenced the therapeutic alliance scores is potential underreporting of ruptures in the treatment relationship due to a lack of awareness of ruptures or a discomfort to acknowledge them [35]. Ruptures are inevitable moments in the treatment process that offer important opportunities for the client and therapist to work through disagreement or discomfort in the therapeutic relationship. Moreover, hindering or disruptive rupture events are important components of the therapy process, with the resolution of these conflicts being a catalyst for change [36]. Underestimation of ruptures in the treatment relationship does not only occur in clients; in previous studies professionals also reported less ruptures as compared to independent observers [35]. To gain a better understanding of therapeutic alliance scores, we recommend that future research includes in depth or semi-structured interviews with clients are needed. A better understanding about the clients' perspective might provide more evidence for the construct validity of the WAI-ReD.

Hypothesis three, stating that the strength of the correlations between the scores of the HAQ-II/SRS and the modified versions of the WAI-ReD are expected to be ≥ 0.600 , was confirmed for the WAI-ReD^{VAS}, and rejected for the WAI-ReD^{PP}. Previous research demonstrated strong correlations between the WAI-ReD and HAQ-II/SRS within rehabilitation settings [14]. Confirming these hypotheses adds evidence to support the construct validity of the WAI-ReD^{VAS}. The finding that the correlation between the HAQ-II

and the WAI-ReD^{PP} was moderate and the WAI-ReD^{PP} and the SRS were not significantly correlated may be explained by the high percentage of missing values on the SRS.

Hypothesis four, stating that the Cronbach's alphas of the total and domains scores of the three different versions of the WAI-ReD are expected to be ≥ 0.700 , was confirmed. Thus, internal consistency of the total- and domain scores of both modified versions was high. In rehabilitation, similar internal consistencies for the WAI-ReD subscales and total score have been reported [14]. In previous studies within psychotherapy, similar internal consistency scores have been reported [15,37].

Our results support the use of WAI-ReD^{VAS} rather than the WAI-ReD^{PP} or the WAI-ReD, because VAS was more effective at preventing ceiling effects. Moreover, the VAS scores may have better precision and more sensitivity to detect changes as compared to Likert scales, simply because of finer gradations of levels of response [18]. However, what a specific change in VAS means for an individual client remains uncertain. The Likert response categories are labelled with words, and thus, they have better face validity and the changes are defined [27]. When ceiling effects are present, the Likert response categories of the WAI-ReD and WAI-ReD^{PP} are not sensitive enough to discriminate therapeutic alliance scores among clients.

There are several study limitations worth noting. First, Likert scaled items were converted to a 0–100 scale to enable comparisons with VAS scores. This conversion is based on the assumption that the Likert scale labels have equal distances, meriting a linear transformation. However, there is no certainty that participants perceived the distances of the different labels on the Likert scale and VAS equally. By converting Likert scales to VAS, interpretations of the labels can be distorted. Second, recruiting participants from physiotherapy practices may have increased the risk for selection bias, as significantly lower total scores on the WAI-ReD were found from participants recruited in the physiotherapy practices as compared to those recruited by physiotherapists in the UMCG. A possible explanation for this finding is that the number of treatments is different by location, severity or duration of complaints. Unfortunately, no specific data were collected about the number of treatments or severity of the complaints in this study. Additionally, no significant differences were found in WAI-ReD scores between different professionals, although clinically differences were substantial. In addition, the lack of significance may be related to sample size. Third, the WAI-ReD^{VAS} group completed the SRS questionnaires (with VAS) significantly more often than the other groups which may have occurred because the WAI-ReD^{VAS} group had an example that showed how to fill in a VAS, whereas the WAI-ReD and WAI-ReD^{PP} groups had no example. As a result, a sequence effect may have occurred. Finally, professionals inconsistently recorded eligible participants and therefore participation rate could not be calculated.

Theoretically, a study with an alternative design to analyse ceiling effects in which participants are asked to complete two versions of the questionnaire within a period of two weeks may seem stronger. The order of filling in two versions of the WAI-ReD (WAI-ReD and the WAI-ReD^{VAS} or WAI-ReD and the WAI-ReD^{PP}) should be determined randomly. This approach, allows for within client comparisons to be conducted instead of between client comparisons, and it also requires fewer clients. However, if differences in scores are found, it is unclear if this is due to the difference in answering options or an actual change in the construct of the WAI in that two-week period.

A strength of this study was the stratification of the different disciplines and the randomization of participants. In addition,

participants were aware that professionals were blinded, which may have decreased social desirability. Regarding missing values, complete cases analyses and multiple imputation analyses were conducted and results of both methods showed similar results, with the exception of differences in the percentage of ceiling effects for the domain Goal, and difference in median score for the domain Task. Nevertheless, the *p* values for both analyses were close to 0.05.

The WAI-ReD^{VAS} showed no ceiling effects on the total- and domain scores, and this measure's clinimetric properties were better than those of the original and Positive-Packed version. Therefore the WAI-ReD^{VAS} is the best version for preventing ceiling effects and improving responsiveness. Notably, the structural validity of the WAI-ReD^{VAS} may be different from the WAI-ReD given the difference in the distribution of the scores. However, the samples size for this study was too small to conduct a confirmatory factor analysis. Therefore, future researchers should consider investigating the structural validity of the WAI-ReD^{VAS}. In addition, no studies have investigated the reliability of the WAI-ReD or the WAI scores. Previous studies demonstrated that applying a VAS to other constructs has a moderate to good reliability [38,39]. A digital version of the VAS or using a slider may increase the reliability of the answering option [40].

Conclusions

This study was successful in reducing ceiling effects in the total and in domain scores of the WAI-ReD by modifying the WAI-ReD in WAI-ReD^{VAS}. The two modified versions WAI-ReD^{PP} and WAI-ReD^{VAS} did not result in lower mean scores on the total- and domain scores compared to WAI-ReD. The therapeutic alliance scores are high on the three versions of the WAI-ReD. Therefore, more research is needed to gain more insight in the mechanisms underlying these high scores. This study provides evidence of better clinimetric properties of the WAI-ReD^{VAS}; therefore, we recommend the use of WAI-ReD^{VAS} for measuring therapeutic alliance in rehabilitation.

Acknowledgements

Special thanks to all the participants and health professionals who have contributed to this study.

Disclosure statement

The authors report no declarations of interest.

References

- [1] Bordin E. The generalizability of the psychoanalytic concept of the working alliance. *Psychother: Theory Res Pract.* 1979; 16:252–260.
- [2] Bordin E. Theory and research on the therapeutic working alliance: new directions. In: Horvath A, Greenberg L, editors. *The working alliance: theory, research and practice.* New York: J. Wiley & Sons; 1994. p. 13–37.
- [3] Ardito RB, Rabellino D. Therapeutic alliance and outcome of psychotherapy: historical excursus, measurements, and prospects for research. *Front Psychol.* 2011;2:270.
- [4] Corso KA, Bryan CJ, Corso ML, et al. Therapeutic alliance and treatment outcome in the primary care behavioural health model. *Fam Syst Health.* 2012;30:87–100.

- [5] Byrne M, Deane FP. Enhancing patient adherence: outcomes of medication alliance training on the therapeutic alliance, insight, adherence, and psychopathology with mental health patients. *Int J Ment Health Nurs*. 2011;20:284–295.
- [6] Graves TA, Tabri N, Thompson-Brenner H, et al. Meta-analysis of the relation between therapeutic alliance and treatment outcome in eating disorders. *Int J Eat Disord*. 2017;40:323–340.
- [7] Nienhuis JB, Owen J, Valentine JC, et al. Therapeutic alliance, empathy, and genuineness in individual adult psychotherapy: a meta-analytic review. *Psychother Res*. 2018;28:593–605.
- [8] Babatunde F, MacDermid J, MacIntyre N. Characteristics of therapeutic alliance in musculoskeletal physiotherapy and occupational therapy practice: a scoping review of the literature. *BMC Health Serv Res*. 2017;17:375.
- [9] Fuertes JN, Mislouack A, Bennett J, et al. The physician–patient working alliance. *Patient Educ Counsel*. 2007;66:29–36.
- [10] Ferreira PH, Ferreira ML, Maher CG, et al. The therapeutic alliance between clinicians and patients predicts outcome in chronic low back pain. *Phys Ther*. 2013;93:470–478.
- [11] Besley J, Kayes NM, McPherson KM. Assessing therapeutic relationships in physiotherapy: literature review. *N Z J Physiother*. 2011;39:81–91.
- [12] Kayes NM, McPherson KM. Human technologies in rehabilitation: ‘Who’ and ‘How’ we are with our clients. *Disabil Rehabil*. 2012;34:1907–1911.
- [13] Hall AM, Ferreira PH, Maher CG, et al. The influence of the therapist–patient relationship on treatment outcome in physical rehabilitation: a systematic review. *Phys Ther*. 2010;90:1099–1110.
- [14] Paap D, Schrier E, Dijkstra PU. Development and validation of the Working Alliance Inventory Dutch Version for use in rehabilitation setting. *Physiother Theory Pract*. 2018;7:1–12.
- [15] Hatcher RL, Gillaspay JA. Development and validation of a revised short version of the Working Alliance Inventory. *Psychother Res*. 2006;16:12–25.
- [16] Araujo AC, Filho RN, Oliveira CB, et al. Measurement properties of the Brazilian version of the Working Alliance Inventory (patient and therapist short-forms) and Session Rating Scale for low back pain. *BMR*. 2017;30:879–887.
- [17] De Vet HC, Terwee CB, Mokkink LB, et al. *Measurement in medicine: a practical guide*. Cambridge: Cambridge University Press; 2011. p. 231–235.
- [18] Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 3rd ed. New York: Oxford University Press; 2003. p. 116–117.
- [19] Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10:10–22.
- [20] Paap D, Dijkstra PU. Working alliance inventory-short form revised. *J Physiother*. 2017;63:118.
- [21] Masino C, Lam TC. Choice of rating scale labels: implication for minimizing patient satisfaction response ceiling effect in telemedicine surveys. *Telemed J E Health*. 2014;20:1150–1155.
- [22] Vita S, Coplin H, Feiereisel KB, et al. Decreasing the ceiling effect in assessing meeting quality at an academic professional meeting. *Teach Learn Med*. 2013;25:47–54.
- [23] Moret L, Nguyen JM, Pillet N, et al. Improvement of psychometric properties of a scale measuring inpatient satisfaction with care: a better response rate and a reduction of the ceiling effect. *BMC Health Serv Res*. 2007;7:197.
- [24] Garratt AM, Helgeland J, Gulbrandsen P. Five-point scales outperform 10-point scales in a randomized comparison of item scaling for the Patient Experiences Questionnaire. *J Clin Epidemiol*. 201;64:200–207.
- [25] Østerås N, Gulbrandsen P, Garratt A, et al. A randomised comparison of a four- and a five-point scale version of the Norwegian Function Assessment Scale. *Health Qual Life Outcomes*. 2008;15:14.
- [26] Voutilainen A, Pitkääho T, Kvist T, et al. How to ask about patient satisfaction? The visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale. *J Adv Nurs*. 2016;72:946–957.
- [27] Bolognese JA, Schnitzer TJ, Ehrich EW. Response relationship of VAS and Likert scales in osteoarthritis efficacy measurement. *Osteoarthr Cartil*. 2003;11:499–507.
- [28] Harland NJ, Dawkin MJ, Martin D. Relative utility of a visual analogue scale vs a six-point Likert scale in the measurement of global subject outcome in patients with low back pain receiving physiotherapy. *Physiotherapy*. 2015;101:50–54.
- [29] Duncan BL, Miller S, Sparks JA, et al. The Session Rating Scale: preliminary psychometric properties of a “working” alliance measure. *J Brief Ther*. 2003;3:3–12.
- [30] Hafkenscheid A. De Outcome rating scale (ORS) en de Session rating scale (SRS). *Tijdschr Psychother*. 2010;36:394–403.
- [31] Luborsky L, Barber JP, Siqueland L, et al. The revised Helping Alliance questionnaire (HAQ-II): psychometric properties. *J Psychother Pract Res*. 1996;5:260–271.
- [32] Hasson D, Arnetz BB. Validation and findings comparing VAS vs. Likert scales for psychosocial measurements. *J Health Educ*. 2005;8:178–192.
- [33] Van de Mortel TF. Faking it: social desirability response bias in self-report research. *Pain*. 2008;136:366–372.
- [34] Barge S, Gehlbach H. Using the theory of satisficing to evaluate the quality of survey data. *Res High Educ*. 2012;53:182–200.
- [35] Safran JD, Muran JC, Eubanks-Carter C. *Repairing alliance ruptures*. Psychotherapy (Chic). 2011;48:80.
- [36] Nockross JC. *Psychotherapy relationships that work: evidence based responsiveness*. 2nd ed. New York: Oxford University Press; 2011.
- [37] Munder T, Wilmers F, Leonhart R, et al. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. *Clin Psychol Psychother*. 2010;17:231–239.
- [38] Carlsson AM. Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale. *Pain*. 1983;16:87–101.
- [39] Kersten P, White PJ, Tennant A. Is the pain visual analogue scale linear and responsive to change? An exploration using Rasch analysis. *PLoS One*. 2014;9:e99485.
- [40] Cook C, Heath F, Russel LT, et al. Score reliability in Webor internet-based surveys: unnumbered graphic rating scales versus Likert-type scales. *Educ Psychol Meas*. 2001;61:697–706.