



## Working alliance with an avatar: How far can we go with internet interventions?



Eva Heim<sup>a,\*</sup>, Alexander Rötger<sup>a</sup>, Noah Lorenz<sup>b</sup>, Andreas Maercker<sup>a</sup>

<sup>a</sup> Department of Psychology, University of Zurich, Switzerland

<sup>b</sup> Faculty of Medicine, University of Leipzig, Germany

### ARTICLE INFO

#### Keywords:

Insomnia  
Internet intervention  
Avatar  
Working alliance

### ABSTRACT

**Objective:** To examine the working alliance between users and an avatar and users' treatment expectations in an unguided Internet intervention for the treatment of insomnia.

**Methods:** The sample included participants from the treatment condition ( $N = 29$ ) of a randomised controlled trial. The *task* and *goal* subscales of the Working Alliance Inventory Short Revised (WAI-SR) were applied in week three. Five items of the Bern Post-Session Report and one question about the extent to which users had missed a human therapist were administered after each session. Treatment expectations were measured with the Credibility Expectancy Questionnaire (CEQ), and the Insomnia Severity Index (ISI) was used as the primary outcome measure.

**Results:** The mean scores for the WAI-SR *task* and *goal* subscales were relatively high ( $M = 3.24$ ,  $SD = 0.79$ ;  $M = 3.16$ ,  $SD = 0.91$ , respectively). The mean score of the five Bern Post-Session Report items remained stable over time, but some users increasingly indicated that they missed a real therapist over the course of the intervention, with a strong linear effect ( $t(87) = 3.16$ ,  $p < 0.01$ ). ISI chance score was predicted by the mean score of the Bern Post-Session Report ( $b = -0.383$ ,  $t(21.80) = -2.97$ ,  $p < 0.01$ ), missing a human therapist ( $b = -0.013$ ,  $t(20.47) = -2.72$ ,  $p = 0.01$ ) and the CEQ ( $b = 0.18$ ,  $t(19.03) = -2.69$ ,  $p = 0.01$ ), but not by WAI-SR *task* and *goal* subscales.

**Conclusions:** Results indicate that users established a working alliance with the avatar. The affective bond remained stable over time, but towards the end of the intervention some users indicated that they missed having a human therapist. Affective bond and missing a real therapist predicted symptom change.

### 1. Introduction

Internet-delivered interventions have proved effective for the treatment of mental health conditions (Andersson et al., 2014), and for promoting healthy behaviours (Hou et al., 2014). The amount of therapeutic support offered in these interventions varies, ranging from unguided self-help to e-mail-based programmes with high therapist investment. It seems that in general there is a curvilinear relationship between the amount of personal support provided to the user and effectiveness of the programme: Guided self-help programmes tend to be more effective than unguided self-help interventions, but this relationship levels off above a certain amount of weekly support time (Titov, 2011).

Insomnia research is one field in which there has been rapid development in Internet-delivered interventions over the past decade (Zachariae et al., 2016). A variety of Internet-delivered self-help tools have been developed and tested, with promising results. One meta-

analysis (Zachariae et al., 2016) identified eleven published randomised controlled trials, all of which used cognitive behavioural therapy (CBT). The global effect size was high for self-reported insomnia severity (Hedge's  $g = 1.09$ , 95% CI 0.74–1.45) and satisfactory for a broad range of secondary outcomes, such as sleep onset latency, wake after sleep onset, and total sleep time. In this meta-analysis, the effect of Internet-based treatment on insomnia severity was positively moderated by personal support provided to users. Similar results have been reported in meta-analyses of Internet-based treatment for other disorders such as depression (e.g. Johansson and Andersson, 2012).

Most surprisingly, two of the most effective interventions were fully-automated self-help programmes with no additional human support (Espie et al., 2012; Ritterband et al., 2009). One of these programmes (Espie et al., 2012) used an automated, virtual sleep coach that guided the user through the programme by giving feedback on data from the sleep diary, the current sleep status and progress made since the start of the intervention. Information delivery was governed by an algorithm of

\* Corresponding authors at: Binzmühlestrasse 14, 8050 Zurich, Switzerland.  
E-mail address: [e.heim@psychologie.uzh.ch](mailto:e.heim@psychologie.uzh.ch) (E. Heim).

sufficient complexity to ensure that the feedback and advice was tailored. The second intervention (Ritterband et al., 2009) did not use an animated sleep coach, but it delivered tailored feedback on the weekly self-reports of insomnia variables and also included a broad range of interactive features such as vignettes, quizzes and short games. A recent study of this intervention found a high effect size ( $d = 2.32$ ) for change between baseline and 12-month follow-up in the treatment condition (Ritterband et al., 2017).

Thus it seems that to some extent therapist contact can be replaced with technological features that simulate human interaction. One of the important issues in this context is users' perception of the working alliance with the automated programme. Bordin (1979) distinguished three aspects of the working alliance in face-to-face therapy: the affective bond between patient and therapist, and agreement between patient and therapist on the tasks and goals of therapy. The Working Alliance Inventory (WAI, Horvath and Greenberg, 1989) is based on Bordin's conceptualisation of working alliance and is widely used in psychotherapy research. In recent years the WAI has been used in studies of Internet-based interventions (e.g. Andersson et al., 2012; Knaevelsrud and Maercker, 2007; Preschl et al., 2011). Berger (2017) summarised this evidence and differentiated between Internet-based interventions on the basis of the amount of support they provide.

Interventions with high therapist investment, such as Interapy (Lange et al., 2003; Lange et al., 2001), are characterised by manualised e-mail exchange between a clinician and a patient. Knaevelsrud and Maercker (2007) showed that a good working alliance could be established in Interapy-based treatment for post-traumatic stress disorder. They reported that the working alliance improved over the course of treatment and that quality of working alliance at the end of treatment correlated with treatment outcome. Similarly, Preschl et al. (2011) found a correlation between working alliance and therapy outcome in an Interapy-based treatment for depression, but working alliance did not predict the residual gain score.

In guided self-help contact between therapist and patient is limited to short e-mails or telephone conversations. Andersson et al. (2012) analysed WAI data from three studies of guided self-help interventions for major depression, generalised anxiety disorder and social anxiety. They found that in all three samples WAI ratings were comparable with those for face-to-face treatments; however working alliance was uncorrelated with treatment outcome. Jasper et al. (2014) compared WAI scores for group-based and guided Internet-based cognitive behaviour therapy for tinnitus (Jasper et al., 2014). In the Internet condition WAI scores on all three subscales were lower than in the group condition in the second and fifth weeks of treatment, but comparable in week nine. Treatment outcome was correlated with the *task* subscale in the Internet-based condition and with the *bond* subscale in group therapy.

Berger et al. (2014) argued that the main component of a guided Internet-based intervention is the programme itself, not the therapist providing guidance. They used adapted versions of the WAI *goal* and *task* subscales to measure users' agreement with the *program* in a guided, Internet-based treatment for several anxiety disorders. Comparing a tailored and a standardised version of the intervention, they found that the adapted WAI subscales correlated with change score in the tailored condition, but not in the standardised condition. Finally, Meyer et al. (2015) looked at alliance ratings in an unguided intervention for depression. In line with Berger et al. (2014) they acknowledged that the "alliance" between a programme user and a software is not equivalent to the alliance between a patient and a human therapist. They therefore administered an adapted version of the Helping Alliance Questionnaire (HAQ; Alexander and Luborsky, 1986) to assess the extent to which users felt that the programme was helpful, seemed to view problems in the same way as them, and seemed to share their goals. The authors found that early HAQ scores (i.e. three weeks after the start of treatment) predicted treatment response after three months.

The research reviewed above suggests that it is possible for users to

establish a working alliance even with fully automated programmes. This is by no means a new insight; in a qualitative analysis of self-help books, Richardson et al. (2010) showed that it is possible to foster a working alliance with a self-help intervention by formulating texts in such a way that the user perceives that there is an understanding clinician behind the text material. In Internet-based interventions, this effect can be intensified by using an avatar. Working alliance with an avatar is a special case: although formally the alliance is established with a programme and not a human therapist - as Berger et al. (2014) and Meyer et al. (2015) observed - an avatar nonetheless simulates human interactional behaviour and this may have an impact on working alliance. To the best of our knowledge this is the first study to examine working alliance in the special case of and intervention with an avatar.

An automated intervention with an element of simulated human interaction might be perceived as more credible and enhance patients' expectations of treatment outcome, both crucial aspects of psychotherapy (Greenberg et al., 2006). The most frequently used measure, the Credibility Expectancy Questionnaire (CEQ, Devilly and Borkovec, 2000), encompasses both components. Boettcher et al. (2013) examined CEQ ratings amongst participants in a randomised controlled trial of an unguided intervention for social anxiety. Positive expectations (i.e. high scores on the CEQ) were associated with improvement in symptoms and also predicted treatment adherence. El Alaoui et al. (2016) looked at the effect of CEQ scores on symptom change in a large sample of adult patients ( $N = 1738$ ) who had been treated for depression as part of routine care in an Internet psychiatry clinic. They found that higher CEQ scores were associated with faster recovery and lower levels of depression at the end of treatment. Thus perceived treatment credibility expectation of success seem to play an important role in Internet-based interventions as well as face-to-face therapy.

A final point to note is that working alliance is a dynamic process, and research on face-to-face psychotherapy has focused on monitoring it over time (Flückiger et al., 2010). Current evidence shows that there is a relationship between good working alliance at the moment-to-moment level and therapeutic change (Mellado et al., 2017). In face-to-face settings the therapist has an immediate influence on the working alliance, whereas in unguided self-help the therapeutic process is fully automated, making it more likely that users will be "lost" over the course of the treatment, with potential consequences for treatment outcome. For this reason, measurements of working alliance at the beginning of the intervention does not provide the full picture. Tracking alliance over time should reveal more detailed information about the course of the therapeutic process.

This study looked at users' perceived working alliance with an avatar in an unguided self-help programme for the treatment of insomnia, and their expectations of treatment success. Working alliance was measured using the *task* and *goal* subscales of the German version of the short, revised WAI (WAI-SR; Munder et al., 2010). The *bond* subscale was not used because the items are formulated in a way that is not suitable for assessing a patient's relationship with an avatar (e.g. 'I feel that the therapist cares about me even when I do things that he/she does not approve of'). To get a general impression of the strength of the working alliance we compared mean WAI-SR subscale scores for the avatar-based intervention with those of the outpatient sample used in the German validation study (Munder et al., 2010), as well as those reported in a study of a guided Internet intervention for tinnitus (Jasper et al., 2014).

We also used a process measure to assess the affective bond component of the working alliance. The Bern Post-Session Report (Flückiger et al., 2010) was developed to track various aspects of the therapeutic process and working alliance in face-to-face therapy and can be administered after each session. For reasons of parsimony and usability we selected the five items with the greatest face validity as indicators of the bond component in Bordin's (1979) conceptualisation of working alliance. As mentioned above, the content of self-help interventions can

be formulated to create the impression that there is an understanding clinician behind the intervention (Richardson et al., 2010). We wanted to determine whether users believe that the avatar is understanding and appreciates their efforts. We also wanted to find out whether users of the self-help intervention miss having a real therapist. If users do not miss having a human therapist, we can conclude that the working alliance they establish with the avatar is sufficient, but missing having a therapist would indicate dissatisfaction with the working alliance. Finally, we predicted that working alliance would be related to the credibility of the intervention.

## 2. Methods

### 2.1. Participants

All participants were recruited via a study website. The link to the webpage was disseminated via e-mail, newspapers, clinics and clinical practitioners in Switzerland, Austria and Germany. Individuals who expressed an interest in participating in the study received an e-mail with information about the study and a consent form. Upon receipt of their signed consent form, they were sent a link to the screening questionnaires consisting of self-report measures and socio-demographic variables.

The inclusion criteria were fluency in German, age of at least 18 years, access to a computer with Internet access, and a minimum score of 8 on the Insomnia Severity Index (ISI; Bastien et al., 2001). The exclusion criteria were working in shifts, ongoing psychological treatment, suicidality, psychotic disorder, a score of  $> 19$  on the Beck Depression Inventory (Beck et al., 1996), and a sleep disorder other than insomnia. People taking medication were only excluded if the dose was changed during the intervention or in the four weeks prior to it, or if there were indications of abusive intake.

### 2.2. Study design and procedure

The sample was drawn from the sample for a randomised controlled trial (Lorenz et al., 2017) of the effectiveness of an Internet-based CBT programme (i.e. *mementor somnium*). Ethical approval for the original randomised controlled trial was granted by the Zurich Ethics Committee, and the trial was registered at [clinicaltrials.gov](http://clinicaltrials.gov) (NCT02629913). The original study showed that *mementor somnium* was effective in reducing self-reported symptoms of insomnia, with a post-treatment between-group effect size of  $d = 0.96$ , and a within-group effect size of  $d = 1.64$  for pre- to post-treatment change in the treatment condition.

The original trial had two parallel arms: an intervention group and a waiting list control group. After potential participants had completed the screening questionnaires, their eligibility was assessed in a structured clinical interview administered by trained interviewers with at least a bachelor's degree in Psychology over the telephone. In total 337 people expressed an interest in the study, 119 registered as potential participants and completed the pre-study screening and 56 met the inclusion criteria. A total of 29 participants were assigned to the intervention group of whom 25 completed the post-treatment assessment. A flowchart of the recruitment process, randomisation, and pre-treatment, post-treatment and follow-up measures can be found in Lorenz et al. (2017). Participants were assigned to the intervention or waiting list condition in a 1:1 ratio using blocked randomisation (blocks of ten). Randomisation was conducted by an independent person who had no contact with the participants. The interviewers were blinded to the randomisation list. Sample size was calculated from a power analysis performed with G\*Power (Faul et al., 2009). We anticipated that the treatment effect would be of medium size ( $f = 0.5$ ) and that there would be 15% of attrition. Based on power of 80% and a significance level of  $p < 0.01$  we calculated that a minimum sample of 40 participants (20 in each group) was required.

The intervention lasted six weeks. All questionnaires were administered through the Internet platform. The post-treatment outcome measure (ISI score) was also obtained electronically, via a link to the online questionnaire that was sent to participants' e-mail address.

### 2.3. Measures

ISI score (Bastien et al., 2001) was the primary outcome measure, with ISI score at screening used as the baseline. The ISI is a self-report questionnaire consisting of seven items scored on a five-point Likert scale: higher scores indicate more severe symptoms and greater perceived impact of sleep difficulties. Items are summed up to yield a total score ranging from 0 to 28. The scoring categories are as follows, 0–7: normal; 8–14: sub-threshold insomnia; 15–21: moderate insomnia;  $> 21$ : severe clinical insomnia. The ISI has shown good reliability and validity (Morin et al., 2011) and has been validated for online use (Thorndike et al., 2011). In our sample the internal consistency (Cronbach's alpha) of the ISI was  $\alpha = 0.547$  at baseline and  $\alpha = 0.792$  at post-assessment.

The *task* and *goal* subscales of the German WAI-SR (Munder et al., 2010) were administered in the third week of treatment. Each subscale consists of four items that are rated using a five-point Likert scale ranging from one to five. The wording was slightly adapted: the word *therapy* was replaced with *training*, e.g. 'What I am doing in my training gives me new ways of looking at my problem'. The internal consistency was  $\alpha = 0.860$  for the *task* subscale and  $\alpha = 0.925$  for the *goal* subscale.

Five items out of the Bern Post-Session Report, patient version (Flückiger et al., 2010) were used to assess users' evaluation of the working alliance with the avatar after each session. The selected items captured the user's affective bond with the avatar, e.g. 'I think that my sleep coach understands my problems correctly'. The items were rated on a seven-point Likert scale ranging from zero to six. The five questions were administered after each session. Principal axis factoring was used to determine whether these five items could be grouped together. Internal consistency ranged from  $\alpha = 0.696$  to  $\alpha = 0.851$  over the course of the intervention.

A single item with a 0–100 point response scale was administered after every session to assess whether users missed having contact with a human therapist.

The CEQ (Devilly and Borkovec, 2000) measures the credibility of the treatment and expectations about treatment success and was developed specifically for use in clinical outcome studies. It comprises six items, e.g. 'How successfully do you think this treatment will be in reducing your symptoms?', or 'How logical does the therapy offered to you seem?' Responses to all items were given using a visual analogue scale and converted into a value between 0 and 100. Items can be divided into two subscales (treatment credibility and outcome expectancies), and a total score can be calculated. The CEQ was administered two weeks after the start of the intervention and the internal consistency for the full scale was  $\alpha = 0.969$ .

### 2.4. Treatment

The online intervention *mementor somnium* was developed in accordance with the German Society of Sleep Medicine's recommendations for the treatment of insomnia (Becker et al., 2009). It consists of six fully automated sessions of CBT, covering psychoeducation, sleep restriction, relaxation, sleep hygiene, cognitive restructuring and changes to sleep-related behaviours. The six treatment sessions are delivered chronologically, with a minimum of 48 h between sessions. Participants are guided through the intervention by an avatar who provides audio information that is supported by dynamic, interactive graphical content. Instructions for sleep restriction are based on calculations of accurate sleep windows that are based on the sleep diary data provided by the participant. Cognitive restructuring is covered in an automated dialogue with the avatar, which is used to identify

potential sleep-related worries and fears. A full description of the contents of the intervention can be found in Lorenz et al. (2017).

### 2.5. Statistical analyses

Only data from the treatment group in the original randomised controlled trial were used ( $N = 29$ ). All variables were normally distributed according to the one-sample Kolmogorov-Smirnov test, so inferential statistics were used despite the small sample size.

First, mean scores on the WAI-SR *task* and *goal* subscales were compared with those of an outpatient sample from the German validation study (Munder et al., 2010), and with data from a guided Internet intervention for tinnitus (Jasper et al., 2014). Effect sizes (Cohen's  $d$ ) were calculated, using the pooled standard deviation as the standardiser. The calculation of the pooled standard deviation was adjusted with weights for the unequal sample sizes.

Second, two separate linear mixed models were calculated to assess changes in the mean score of the Bern Post-Session Report items and the missing a real therapist variable during the treatment. Model degree (i.e., linear, quadratic, cubic) was tested initially. Thereafter, four nested models were tested against each other: fixed factors (subjects and time), random intercepts (subjects), random slopes (time), and a last model accounting for a lag-1 autocorrelation. These four models were tested for each score (i.e. the mean score of the Bern Post-Session Report and the question whether users had missed a real therapist). The linear mixed models were calculated using the R package “nlme” (Pinheiro et al., 2017).

Third, correlations between predictor variables and change scores were analysed using Pearson's correlations. And fourth, the predictive value of the working alliance and expectations about treatment outcome were analysed in separate regression models, using ISI pre- to post-treatment change score as the dependent variable. Multiple imputation ( $m = 15$ ) was used to handle missing data in the regression analyses.

### 3. Results

Descriptive statistics of the included measures are presented in Table 1. The comparison with published data on outpatients ( $N = 88$ ) from the German validation study (Munder et al., 2010) revealed effect sizes of  $d = 0.21$  (95% CI:  $-0.24-0.65$ ) for the *task* subscale, and

**Table 1**

Minimum, maximum, mean, and standard deviation for the ISI, WAI-SR; CEQ total score, Bern Post Session Report (five items, mean score), and single item *How much have you missed a real therapist?*

Measure	Assessment Time	N	Min	Max	Mean	SD
ISI	Pre-assessment	29	9.00	23.00	15.38	3.74
ISI	Post-assessment	25	1.00	21.00	7.80	4.94
CEQ	Week two	21	37.67	83.33	62.87	14.75
WAI-SR, <i>task</i>	Week three	25	1.75	4.75	3.24	0.79
WAI-SR, <i>goal</i>	Week three	25	1.00	4.50	3.16	0.91
Bern Post Session Report (total score)	Session 1	25	2.80	5.20	4.21	0.65
	Session 2	23	2.40	6.00	4.23	0.91
	Session 3	26	2.80	5.60	4.31	0.77
	Session 4	22	2.20	5.60	4.01	0.92
	Session 5	22	2.60	6.00	4.29	0.88
	Mean (S1–S5)	28	2.90	5.47	4.26	0.72
Missed a real therapist	Session 1	23	0	41.00	15.44	11.77
	Session 2	21	0	69.00	31.24	23.72
	Session 3	25	0	100.00	28.76	28.34
	Session 4	21	0	74.00	25.00	25.27
	Session 5	21	0	97.00	35.86	29.99
	Mean (S1–S5)	28	0	75.50	29.03	21.46

Note: ISI = Insomnia Severity Index; CEQ = Credibility Expectancy Questionnaire WAI-SR = Working Alliance Inventory Short Revised.

$d = 1.14$  (95% CI:  $0.67-1.61$ ) for the *goal* subscale. Thus in the case of the subscale *task*, the difference between the samples was small and the confidence interval included zero, indicating that there was practically no difference between the Internet intervention group and the reference group. In contrast there was a considerable difference between the samples in scores on the *goal* subscale. The comparison with data published by Jasper et al. (2014) is presented in Table 2. Participants who received the *mementor somnium* intervention had WAI-SR *task* and *goal* subscale scores comparable with those reported by Jasper et al.'s (2014) sample in week nine. In weeks two and five Jasper et al.'s (2014) sample reported much lower scores than participants receiving the *mementor somnium* intervention.

Within-group differences in session scores for the Bern Post-Session Report were examined using linear mixed models. There was no effect of time on the session means. The best model fit was achieved with random intercepts (subjects) and slopes (time); accounting for autocorrelation did not improve model fit. The effect of time on session means was non-significant when testing linear ( $t(87) = 0.00, p = 0.99$ ), quadratic ( $t(87) = -1.12, p = 0.26$ ), and cubic ( $t(87) = 1.03, p = 0.31$ ) relationships. With regard to temporal changes in missing having a real therapist, the best model fit was achieved with random intercepts (subjects) and slopes (time) and once again accounting for autocorrelation did not improve model fit. There were strong linear ( $t(87) = 3.16, p < 0.01$ ) and quadratic ( $t(87) = 3.42, p < 0.001$ ) effects, indicating that users increasingly missed having a real therapist.

Correlations between all study variables are displayed in Table 3. The two WAI-SR subscales (*task* and *goal*) did not correlate with ISI difference score. Session means for the Bern Post-Session Report and missing having a real therapist were highly correlated and CEQ was moderately correlated with ISI difference score. The strength of the correlation between missing having a real therapist and ISI difference score decreased over time, from  $r = 0.56$  at session one to  $r = 0.40$  at session five. The changes in Bern Post-Session Report scores followed the opposite pattern. The strength of the correlation between Bern Post-Session Report scores and ISI difference score increased from  $r = -0.47$  at session one to  $r = -0.59$  at session five. The WAI-SR subscales and the Bern Post-Session Report were both highly correlated with the CEQ.

Due to the high correlations between the predictor variables and the risk of multicollinearity three single regressions were conducted with ISI difference score as the dependent variable and the Bern Post-Session Report mean score, missing having a therapist and total CEQ total score as predictors whilst controlling for the pre-assessment ISI score. The results are displayed in Table 4. The coefficients for the three predictors achieved statistical significance, in other words they explained variance in ISI difference score independently of pre-assessment ISI score.

### 4. Discussion

This paper looked at the users' perceived working alliance with an avatar in an Internet-based intervention for the treatment of insomnia entitled *mementor somnium*. So far working alliance has been examined in Internet interventions with high therapist investment, such as Interapy, as well as in guided and unguided self-help programmes (Berger, 2017). Working alliance with an avatar is a special case, because the avatar simulates human interaction and may therefore enhance the illusion that there is an understanding clinician behind the text material (Richardson et al., 2010).

In this study, mean WAI-SR *task* subscale scores were comparable with ratings from an outpatient sample in the German validation study (Munder et al., 2010). In contrast, mean WAI-SR *goal* subscale scores were considerably lower than those reported in the German validation study. Moreover, participants using *mementor somnium* gave higher ratings on both WAI-SR subscales (in the third week of treatment) than the means obtained in weeks two and five of a guided Internet intervention for tinnitus (Jasper et al., 2014). These results indicate that a

**Table 2**  
Comparison of the WAI-SR subscales *task* and *goal* with means and standard deviations published by Jasper et al. (2014).

	Mementor somnium		Jasper et al. (2014) Week 2			Jasper et al. (2014) Week 5			Jasper et al. (2014) Week 9		
	M	SD	M	SD	Effect size Cohen's d [95% CI]	M	SD	Effect size Cohen's d [95% CI]	M	SD	Effect size Cohen's d [95% CI]
WAI-SR <i>task</i>	3.24	0.79	2.32	0.67	-1.28 [-1.83; -0.73]	2.70	0.78	-0.69 [-1.21; -0.17]	3.14	0.78	-0.13 [-0.63; 0.38]
WAI-SR <i>goal</i>	3.16	0.91	2.26	0.95	-0.96 [-1.50; -0.43]	2.38	1.32	-0.66 [-1.18; -0.15]	3.79	0.80	0.76 [0.22; 1.27]

working alliance might be established much earlier in interventions using an avatar than in guided self-help programmes. However, comparison of our sample's working alliance ratings (in week three) and those of Jasper et al.'s (2014) sample (in week nine) revealed that WAI-SR *task* scores were similar whilst the patients using *mementor somnium* reported lower WAI-SR *goal* scores.

We can only speculate about the reasons for this result. One possible explanation for the fact that *mementor somnium* users gave considerably lower scores on the *goal* subscale than on the *task* subscale relates to the wording of these two subscales. The *task* subscale captures agreement with the therapy or training, whereas the *goal* subscale captures agreement with the avatar. Our results may indicate that ultimately the working alliance is established with the programme as a whole rather than with the avatar (Berger et al., 2014; Meyer et al., 2015).

A process measure was administered after each session to assess users' affective bond with the avatar. This consisted of five items from the Bern Post-Session Report and a question about whether users had missed having a real therapist during the session. The session mean scores of the Bern Post Session-Report remained stable at a relatively high level (a mean score of 4 on a 0–6 scale) over the five sessions. These findings contrast with those of Jasper et al. (2014), who found that users' ratings of their affective bond with the therapist in a guided intervention for tinnitus were relatively low in weeks two and five but increased over time, achieving a level that was comparable with group therapy in week nine. Again, our results suggest that an affective bond might be established faster when using an avatar. However, there were linear and quadratic effects of time on responses to the question about missing having a real therapist. It seems that over the course of the intervention some of the users increasingly came to feel that they would appreciate the specific support provided by a human therapist.

One possible explanation for the fact that some users increasingly missed having a human therapist relates to the therapeutic techniques used in each session. The first sessions of the *mementor somnium* intervention cover psychoeducation and behavioural tasks such as sleep restriction, whereas the later sessions encompass cognitive

**Table 3**  
Pearson correlation between change score for the ISI, working alliance, therapeutic relationship, credibility and expectancies.

		1.	2.	3.	4.	5.
1. ISI difference score		–				
2. WAI-SR, subscale <i>task</i>	Correlation coefficient (r)	-0.208	–			
	Significance (p)	0.319				
	(df)	(23)				
3. WAI-SR, subscale <i>goal</i>	Correlation coefficient (r)	-0.348	0.738	–		
	Significance (p)	0.088	0.000			
	(df)	(23)	(24)			
4. Bern Post Session Report (mean score)	Correlation coefficient (r)	-0.583	0.484	0.677	–	
	Significance (p)	0.001	0.012	0.000		
	(df)	(25)	(24)	(24)		
5. Missed a real therapist (mean score)	Correlation coefficient (r)	0.619	-0.533	-0.531	-0.628	–
	Significance (p)	0.001	0.005	0.005	0.000	
	(df)	(25)	(25)	(24)	(26)	
6. CEQ	Correlation coefficient (r)	-0.428	0.726	0.707	0.614	-0.613
	Significance (p)	0.047	0.000	0.000	0.002	0.002
	(df)	(20)	(20)	(20)	(21)	(21)

Note: ISI = Insomnia Severity Index; WAI-SR = Working Alliance Inventory Short Revised; CEQ = Credibility Expectancy Questionnaire.

**Table 4**  
Single regressions predicting the ISI change score.

	b	se	t(df)	p	R <sup>2</sup> [95% CI]
<b>ISI change score</b>					0.44 [0.15; 0.69]
ISI_pre	-0.54	0.25	-2.22(21.30)	0.038	
Missed therapist	0.13	0.05	2.74(20.47)	0.012	
<b>ISI change score</b>					0.46 [0.16; 0.71]
ISI_pre	-0.47	0.25	-1.89 (22.13)	0.072	
BPSR	-3.83	1.29	-2.97 (21.80)	0.007	
<b>ISI change score</b>					0.44 [0.14; 0.70]
ISI_pre	-0.61	0.24	-2.54(20.79)	0.019	
CEQ	-0.18	0.07	-2.69 (19.03)	0.014	

Note: ISI = Insomnia Severity Index; CEQ = Credibility Expectancy Questionnaire; BPSR = Bern Post Session Report.

restructuring and modification of safety behaviours through automated dialogue with the avatar, which is based on closed questions. It is possible that the users find this simulated dialogue too restricted and thus increasingly wish for a dialogue with a human therapist. Future research could explore which interventions can more easily be delivered via a automated self-help and which interventions require human support.

As expected, the credibility of the intervention and users' expectations about treatment success were highly correlated with measures of working alliance. Moreover, the mean Bern Post-Session Report score, the mean score on the question about missing a human therapist and CEQ score all explained variance in the ISI change score. In contrast scores on the WAI-SR *task* and *goal* subscales were not correlated with symptom change. Evidence about the relationship between working alliance and symptom change in guided and unguided Internet-based interventions is inconsistent (Andersson et al., 2012; Berger et al., 2014; Jasper et al., 2014; Meyer et al., 2015). In our study, the affective bond with the avatar and the question about missing a human therapist were more important for symptom change than agreement with the avatar

about the tasks and goals of the intervention. This contrasts with the findings of Jasper et al. (2014), who reported that in patients receiving an Internet-based intervention for tinnitus the WAI-SR task subscale was related to symptom change, whereas in those receiving group treatment the WAI-SR bond subscale was related to symptom change. More research is needed to improve understanding of how the different components of working alliance relate to symptom change.

Several limitations have to be taken into account when interpreting the results of this study. First, the small sample size did not allow for more sophisticated analyses. With a bigger sample it would have been possible to divide users into sub-samples and assess the course of their symptoms in relationship with their expectations, described by Boettcher et al. (2013). Second, the inclusion and exclusion criteria were very strict and we implemented a multi-stage screening procedure. This resulted in a highly selective sample that met clearly defined criteria and was highly motivated to engage with treatment. A more naturalistic study would be needed to test the efficiency of the programme in routine care. Third, the WAI-SR task and goal subscales were used due to the lack of a specific instrument for measuring working alliance with an avatar. The WAI-SR was developed for use in face-to-face treatments, and we adapted the wording slightly. A questionnaire designed specifically to measure working alliance with automated programmes and avatars might provide a clearer picture of how working alliance relates to symptom change (Andersson et al., 2012).

Despite these limitations, the study provides insight into automated intervention users' perspectives on working alliance with an avatar. Working alliance with an avatar is a special case, because the avatar simulates human interaction more realistically than a mere text-based self-help intervention. Our results indicate that using an avatar can enhance users' impression that there is an understanding clinician behind the intervention (Richardson et al., 2010) and that users develop and maintain a working alliance with the avatar, which might improve treatment effects. However, it is not clear from our results whether the working alliance was established with the avatar or with the programme as a whole. Further research is needed to replicate these first results on working alliance with an avatar.

## Conflict of interest

Alexander Rötger and Noah Lorenz have equity ownership in *mementor*, the company that has licensed the *mementor somium* program which was tested in this study.

## References

- Alexander, L.B., Luborsky, L., 1986. The Penn helping alliance scales. In: Greenberg, L.S. (Ed.), *The Psychotherapeutic Process: A Research Handbook*. Guilford Press, New York, NY, pp. 325–366.
- Andersson, G., Cuijpers, P., Carlbring, P., Riper, H., Hedman, E., 2014. Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. *World Psychiatry* 13 (3), 288–295. <http://dx.doi.org/10.1002/wps.20151>.
- Andersson, G., Paxling, B., Wiwe, M., Vermark, K., Felix, C.B., Lundborg, L., ... Carlbring, P., 2012. Therapeutic alliance in guided internet-delivered cognitive behavioural treatment of depression, generalized anxiety disorder and social anxiety disorder. *Behav. Res. Ther.* 50 (9), 544–550. <http://dx.doi.org/10.1016/j.brat.2012.05.003>.
- Bastien, C.H., Vallieres, A., Morin, C.M., 2001. Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep Med.* 2 (4), 297–307.
- Beck, A.T., Steer, R.A., Ball, R., Ranieri, W., 1996. Comparison of Beck depression inventories -IA and -II in psychiatric outpatients. *J. Pers. Assess.* 67 (3), 588–597. [http://dx.doi.org/10.1207/s15327752jpa6703\\_13](http://dx.doi.org/10.1207/s15327752jpa6703_13).
- Becker, H.F., Ficker, J., Fietze, I., Geisler, P., Happe, S., Hornyak, M., ... Wiater, A., 2009. S3-Leitlinie: Nicht erholsamer Schlaf. *Somnologie - Schlaforschung und Schlafmedizin* 13 (Suppl. 1), 1–160. <http://dx.doi.org/10.1007/s11818-009-0430-8>. (1).
- Berger, T., 2017. The therapeutic alliance in internet interventions: a narrative review and suggestions for future research. *Psychother. Res.* 27 (5), 511–524. <http://dx.doi.org/10.1080/10503307.2015.1119908>.
- Berger, T., Boettcher, J., Caspar, F., 2014. Internet-based guided self-help for several anxiety disorders: a randomized controlled trial comparing a tailored with a standardized disorder-specific approach. *Psychotherapy* 51 (2), 207–219.
- Boettcher, J., Renneberg, B., Berger, T., 2013. Patient expectations in internet-based self-help for social anxiety. *Cogn. Behav. Ther.* 42 (3), 203–214. <http://dx.doi.org/10.1080/16506073.2012.759615>.
- Bordin, E.S., 1979. The generalizability of the psychoanalytic concept of the working alliance. *Psychother. Theory Res. Pract.* 16 (3), 252–260. <http://dx.doi.org/10.1037/h0085885>.
- Devilly, G.J., Borkovec, T.D., 2000. Psychometric properties of the credibility/expectancy questionnaire. *J. Behav. Ther. Exp. Psychiatry* 31 (2), 73–86.
- El Alaoui, S., Ljótsson, B., Hedman, E., Svanborg, C., Kaldo, V., Lindfors, N., 2016. Predicting outcome in internet-based cognitive behaviour therapy for major depression: a large cohort study of adult patients in routine psychiatric care. *PLoS One* 11 (9), e0161191. <http://dx.doi.org/10.1371/journal.pone.0161191>.
- Espie, C.A., Kyle, S.D., Williams, C., Ong, J.C., Douglas, N.J., Hames, P., Brown, J.S., 2012. A randomized, placebo-controlled trial of online cognitive behavioral therapy for chronic insomnia disorder delivered via an automated media-rich web application. *Sleep* 35 (6), 769–781. <http://dx.doi.org/10.5665/sleep.1872>.
- Faul, F., Erdfelder, E., Buchner, A., Lang, A.-G., 2009. Statistical power analyses using G\*power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160.
- Flückiger, C., Regli, D., Zwahlen, D., Hostettler, S., Caspar, F., 2010. Der Berner Patienten- und Therapeutenstundenbogen 2000 [the Bern post session report, patient and therapist versions 2000]. *Z. Klin. Psychol. Psychother.* 39 (2), 71–79. <http://dx.doi.org/10.1026/1616-3443/a000015>.
- Greenberg, R.P., Constantino, M.J., Bruce, N., 2006. Are patient expectations still relevant for psychotherapy process and outcome? *Clin. Psychol. Rev.* 26 (6), 657–678. <http://dx.doi.org/10.1016/j.cpr.2005.03.002>.
- Horvath, A.O., Greenberg, L.S., 1989. Development and validation of the working alliance inventory. *J. Couns. Psychol.* 36. <http://dx.doi.org/10.1037/0022-0167.36.2.223>.
- Hou, S.-I., Charlery, S.-A.R., Roberson, K., 2014. Systematic literature review of internet interventions across health behaviors. *Health Psychol. Behav. Med.* 2 (1), 455–481. <http://dx.doi.org/10.1080/21642850.2014.895368>.
- Jasper, K., Weise, C., Conrad, I., Andersson, G., Hiller, W., Kleinstäuber, M., 2014. The working alliance in a randomized controlled trial comparing internet-based self-help and face-to-face cognitive behavior therapy for chronic tinnitus. *Int. Interv.* 1 (2), 49–57. <http://dx.doi.org/10.1016/j.invent.2014.04.002>.
- Johansson, R., Andersson, G., 2012. Internet-based psychological treatments for depression. *Expert. Rev. Neurother.* 12 (7), 861–870. <http://dx.doi.org/10.1586/ern.12.63>.
- Knaevelsrud, C., Maercker, A., 2007. Internet-based treatment for PTSD reduces distress and facilitates the development of a strong therapeutic alliance: a randomized controlled clinical trial. *BMC Psychiatry* 7, 13. <http://dx.doi.org/10.1186/1471-244x-7-13>.
- Lange, A., Rietdijk, D., Hudcovicova, M., van de Ven, J.P., Schrieken, B., Emmelkamp, P.M., 2003. Interapy: a controlled randomized trial of the standardized treatment of posttraumatic stress through the internet. *J. Consult. Clin. Psychol.* 71 (5), 901–909. <http://dx.doi.org/10.1037/0022-006x.71.5.901>.
- Lange, A., van de Ven, J.P., Schrieken, B., Emmelkamp, P.M., 2001. Interapy, treatment of posttraumatic stress through the internet: a controlled trial. *J. Behav. Ther. Exp. Psychiatry* 32 (2), 73–90.
- Lorenz, N., Heim, E., Roetger, A., Birrer, E., Maercker, A., 2017. Randomised controlled trial to test the efficacy of an unguided online intervention with automated feedback for the treatment of insomnia. (Manuscript submitted for publication).
- Mellado, A., Suárez, N., Altimir, C., Martínez, C., Pérez, J., Krause, M., Horvath, A., 2017. Disentangling the change-alliance relationship: observational assessment of the therapeutic alliance during change and stuck episodes. *Psychother. Res.* 27 (5), 595–607. <http://dx.doi.org/10.1080/10503307.2016.1147657>.
- Meyer, B., Bierbrodt, J., Schröder, J., Berger, T., Beevers, C.G., Weiss, M., ... Klein, J.P., 2015. Effects of an internet intervention (Deprexis) on severe depression symptoms: randomized controlled trial. *Int. Interv.* 2 (1), 48–59. <http://dx.doi.org/10.1016/j.invent.2014.12.003>.
- Morin, C.M., Belleville, G., Belanger, L., Ivers, H., 2011. The insomnia severity index: psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep* 34 (5), 601–608.
- Munder, T., Wilmers, F., Leonhart, R., Linster, H.W., Barth, J., 2010. Working alliance inventory-short revised (WAI-SR): psychometric properties in outpatients and inpatients. *Clin. Psychol. Psychother.* 17 (3), 231–239. <http://dx.doi.org/10.1002/cpp.658>.
- Pineiro, J., Bates, D., DeBrooy, S., Sarkar, D., R Core Team, 2017. nlme: Linear and Nonlinear Mixed Effects Models. R Package Version. vol. 3. pp. 1–131. <https://CRAN.R-project.org/package=nlme>.
- Preschl, B., Maercker, A., Wagner, B., 2011. The working alliance in a randomized controlled trial comparing online with face-to-face cognitive-behavioral therapy for depression. *BMC Psychiatry* 11 (1), 189. <http://dx.doi.org/10.1186/1471-244x-11-189>.
- Richardson, R., Richards, D.A., Barkham, M., 2010. Self-help books for people with depression: the role of the therapeutic relationship. *Behav. Cogn. Psychother.* 38 (1), 67–81. <http://dx.doi.org/10.1017/s1352465809990452>.
- Ritterband, L.M., Thorndike, F.P., Gonder-Frederick, L.A., Magee, J.C., Bailey, E.T., Saylor, D.K., Morin, C.M., 2009. Efficacy of an internet-based behavioral intervention for adults with insomnia. *Arch. Gen. Psychiatry* 66 (7), 692–698. <http://dx.doi.org/10.1001/archgenpsychiatry.2009.66>.
- Ritterband, L.M., Thorndike, F.P., Ingersoll, K.S., Lord, H.R., Gonder-Frederick, L.A., Frederick, C., ... Morin, C.M., 2017. Effect of a web-based cognitive behavior therapy for insomnia intervention with 1-year follow-up: a randomized clinical trial. *JAMA Psychiatry* 74 (1), 68–75. <http://dx.doi.org/10.1001/jamapsychiatry.2016.3249>.
- Thorndike, F.P., Ritterband, L.M., Saylor, D.K., Magee, J.C., Gonder-Frederick, L.A., Morin, C.M., 2011. Validation of the insomnia severity index as a web-based measure. *Behav. Sleep Med.* 9 (4), 216–223. <http://dx.doi.org/10.1080/15402002.2011.606766>.
- Titov, N., 2011. Internet-delivered psychotherapy for depression in adults. *Curr. Opin. Psychiatry* 24 (1), 18–23. <http://dx.doi.org/10.1097/YCO.0b013e32833ed18f>.
- Zachariae, R., Lyby, M.S., Ritterband, L.M., O'Toole, M.S., 2016. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia: a systematic review and meta-analysis of randomized controlled trials. *Sleep Med. Rev.* 30, 1–10. <http://dx.doi.org/10.1016/j.smrv.2015.10.004>.