Efficacy of the web-based PaarBalance program on relationship satisfaction, depression and anxiety – A randomized controlled trial

Alina Keller [a,1], Anna Babl [b,1,*], Thomas Berger [b], Ludwig Schindler [a]

[a] *Department of Clinical Psychology and Psychotherapy, University of Bamberg, Germany*
[b] *Department of Clinical Psychology and Psychotherapy, University of Bern, Switzerland*

ARTICLE INFO

ABSTRACT

Objective: Although relationship distress is strongly associated with mental health problems, poorer social functioning and lower quality of life, only a minority of distressed couples engage in effective couples therapy. Common barriers are the financial burden, fear of being stigmatized, long waitlists and logistical concerns, such as the difficulty in scheduling appointments. Therefore, more accessible help for relationship distress is needed, such as internet-based interventions.
Method: This study evaluates the efficacy of the German web-based PaarBalance program, an 18-sessions online program for couples and individuals in an intimate relationship. Participants with relationship distress recruited via the internet had access to the unguided self-help program for twelve weeks. A total of 117 individuals ($N = 60$ participated as couples, $N = 57$ participated without a partner) were randomly assigned to begin the intervention immediately or to a 12-week waitlist control group. The primary outcome was relationship satisfaction. Secondary outcomes included symptoms of depression and anxiety.
Results: The intervention group showed significant improvement in relationship satisfaction (Cohen's $d = 0.77$) compared with the waitlist control group. Small to medium effect sizes in favor of the intervention group, but no statistically significant differences were found regarding depression ($d = 0.43$) and anxiety ($d = 0.45$).
Conclusion: PaarBalance seems to be an effective self-guided intervention to improve relationship satisfaction in people with relationship problems.

1. Introduction

Humans predominantly live in social relationships. For most people, especially a romantic relationship or marriage is a major life goal (e.g., Diener et al., 2000). Although divorce rates have increased over the last years, most individuals still decide to get married (Bodenmann, 2016). The main conclusion of the 75-year longitudinal Harvard grant study was that the warmth of close relationships had the most positive influence on life satisfaction (Vaillant, 2002). In a meta-analysis, Proulx et al. (2007) analyzed 93 studies evaluating the effect of marital quality on personal well-being. The results suggest that the higher the levels of marital quality, the higher the levels of personal well-being. In addition,

empirical evidence suggests that marriage is related to a reduced risk of developing mental disorders (e.g., Scott et al., 2010; Whisman, 2019). Even mortality rates seem to decrease based on relationship stability (Kiecolt-Glaser and Newton, 2001). In sum, past research demonstrates the importance of stable, intimate relationships.

Accordingly, empirical evidence suggests marital distress to be a risk factor for the development of mental disorders (e.g., Whisman, 2019). A longitudinal study evaluating the effects of marital distress on symptoms of depression and generalized anxiety disorder proposed a positive association between marital discord and psychopathological symptoms (Whisman et al., 2018). Evidence also exists for the association between relationship distress and physical health problems. For example, hostile

behaviors in relationships are correlated with alterations of cardiovascular activity, stress-related hormones, and dysregulation of the immune system (Robles and Kiecolt-Glaser, 2003). Low relationship quality not only influences couples themselves but also the people around them. For instance, marital conflicts negatively influence child development (Cummings and Davies, 2010). Taken together, the effects of low marital quality are expansive.

Therefore, several interventions treating relationship distress have previously been developed. Behavioral couple therapy is the therapeutic approach that presents most empirical evidence and builds the foundation of the PaarBalance program (e.g., Baucom et al., 2019). The effect of behavioral couple therapy on relationship distress is well approved (e.g., Shadish and Baldwin, 2005). A review suggests that behavioral couple therapy effectively influences the relationship for 70% of the couples (Lebow et al., 2012). Moreover, behavioral couple therapy's important role in the treatment of concomitant mental disorders has been demonstrated (e.g., Baucom et al., 2020).

Although couple therapy is effective, distressed couples frequently consult therapeutic help too late or not at all. Barriers to face-to-face therapy could be treatment costs, fear of being stigmatized, long waitlists and logistical concerns, such as the difficulty in scheduling appointments (e.g., Amichai-Hamburger et al., 2014). Internet-delivered interventions overcome some of these boundaries. They are often accessible at low cost compared to face-to-face therapy (e.g., Georgia Salivar et al., 2018), more anonymous, and available at any time and any place. Moreover, populations with limited access to face-to-face therapy can benefit. Thus, internet-based programs for couples are an enrichment for couple interventions (Cicilia et al., 2014).

A number of trials and reviews demonstrating the efficacy and effectiveness of online interventions for different psychopathologies (e.g., Andersson et al., 2019; Andrews et al., 2018; Berger et al., 2011; Cuijpers et al., 2010). Several studies suggest that online couple therapy is effective in increasing relationship satisfaction in American populations. For instance, studies report that the computer-based version of the Prevention and Relationship Enhancement Program (ePREP) effectively teaches and improves relationship skills (e.g., Braithwaite and Fincham, 2007, 2009; Loew et al., 2012). PREP is a psychoeducative relationship program that focuses on functional communication (Markman et al., 2009). There is one existing translation of Integrative Behavioral Couple Therapy into an online treatment in English language, the OurRelationship program (OR; Doss et al., 2013). Doss et al. (2016) compared OR to a waitlist control group in a controlled trial. Couples in the OR-condition showed significantly more improvement in relationship satisfaction, individual functioning and symptoms of depression and anxiety than couples in the control group. These positive effects were stable at one-year-follow-up (Doss et al., 2019). In another clinical trial, it was shown that both ePREP and OR improved relationship quality for low-income couples (Doss et al., 2020). Significant differences in the effectiveness of the two treatments were found for verbal conflict (significant decrease in the OR group), but not for satisfaction, breakup potential, intimacy and intimate partner violence. To summarize, research suggests a high potential for web-based couple interventions. In German language, some scientifically substantiated and promising online programs for couples exist: besides PaarBalance those are e.g., Paarlife (Bodenmann, 2020) or Theratalk (Beer, 2020). However, so far no studies have been published examining the effects of online couple therapy in German-speaking populations.

In the current study, we therefore aim to evaluate the efficacy of the web-based German PaarBalance program on relationship satisfaction, depression, and anxiety in a randomized controlled trial (RCT). We hypothesize that the self-help program will have a significant effect on increasing relationship satisfaction and decreasing depression and anxiety compared to the waitlist condition. Further, we predict a positive change in quality of life and work functioning in the intervention group compared to the control group.

2. Material and methods

This RCT compared an immediate intervention group with a waitlist control group. The final sample size, study design, outcome measures, and inclusion and exclusion criteria were determined before data collection. The protocol of the study was approved by the institutional review board of XXX (blinded for review). The study is registered in the German register of clinical studies (Deutsches Register Klinischer Studien, DRKS).

2.1. Recruitment and selection of participants

Participants were recruited via internet blogs and websites, which cover relationship or parenthood topics. They were neither paid nor received other incentives to participate except for the free use of PaarBalance. If interested, people wrote an e-mail to receive further information about the study and the requirements for participation. For instance, to obtain informed consent, it was pointed out that there are two different groups named Early Starter (intervention group) and Later Starter (control group) they would be assigned to. In the next step, potential participants had to specify whether they would like to participate alone or together with their partner. A subsequent screening questionnaire (included in the first measurement) determined eligibility for the study.

Participants had to be at least 18 years old, have private internet access, and a good knowledge of written and spoken German. Moreover, criteria for inclusion were being in a heterosexual relationship (to ensure homogeneity of the sample) and being in a relationship for at least six months. The relationship satisfaction, measured with a partnership questionnaire (PFB-K; Hahlweg, 2016) had to be 18 or lower ($T = 50$) to be in a distressed range. Criteria for exclusion were being in the process of separation, currently participating in couples counseling or individual therapy, or taking psychiatric medication. People who reported violence in their relationship or a suicidal tendency were also excluded and given information regarding appropriate treatment facilities. If one partner was not eligible, the other could also not participate in the study. However, both were given free access to the PaarBalance program without participation in the present study.

A total of 192 individuals applied for participation, 117 met the inclusion criteria. A stratified randomization procedure was applied such that a balanced distribution of couples or individuals to the two treatment arms was ensured. The allocation lists were made using a computerized random number generator and were unknown to the investigators. The main reason for exclusion was a too high level of relationship satisfaction (PFB-K raw score > 18), which was the case for 52 participants (see Fig. 1).

Post-hoc power analyses revealed that the sample of 117 participants provided 98.6% of the power for relationship satisfaction and 97.5% for quality of life but only 67.5% for anxiety, 63.9% for depression, and 50.3% for work functioning at the Type I error probability of 0.05.

2.2. Participants

Thirty couples and 57 individuals participated in the study (39 males, 78 females). The mean age was 40.4 years ($SD = 10.6$; *range* = 19–71). 72.0% were married, 25.0% were never married, 2.0% were engaged and 2.0% divorced. On average, couples had been together for 13.2 years ($SD = 9.4$; *range* = 1–41). Of the 117 participants, 105 (89.7%) had children, 12 (10.3%) did not have children. The majority of the sample had a university degree as the highest level of education (78.0%). Most participants were employed (73.0%), some self-employed (10.0%), the rest of the sample were students (7.0%), homemakers (14.0%), unemployed (3.0%), or retired (1.0%). The majority of the sample was German (91.0%) and Austrian (6.0%) or had other nationalities (3.0%).
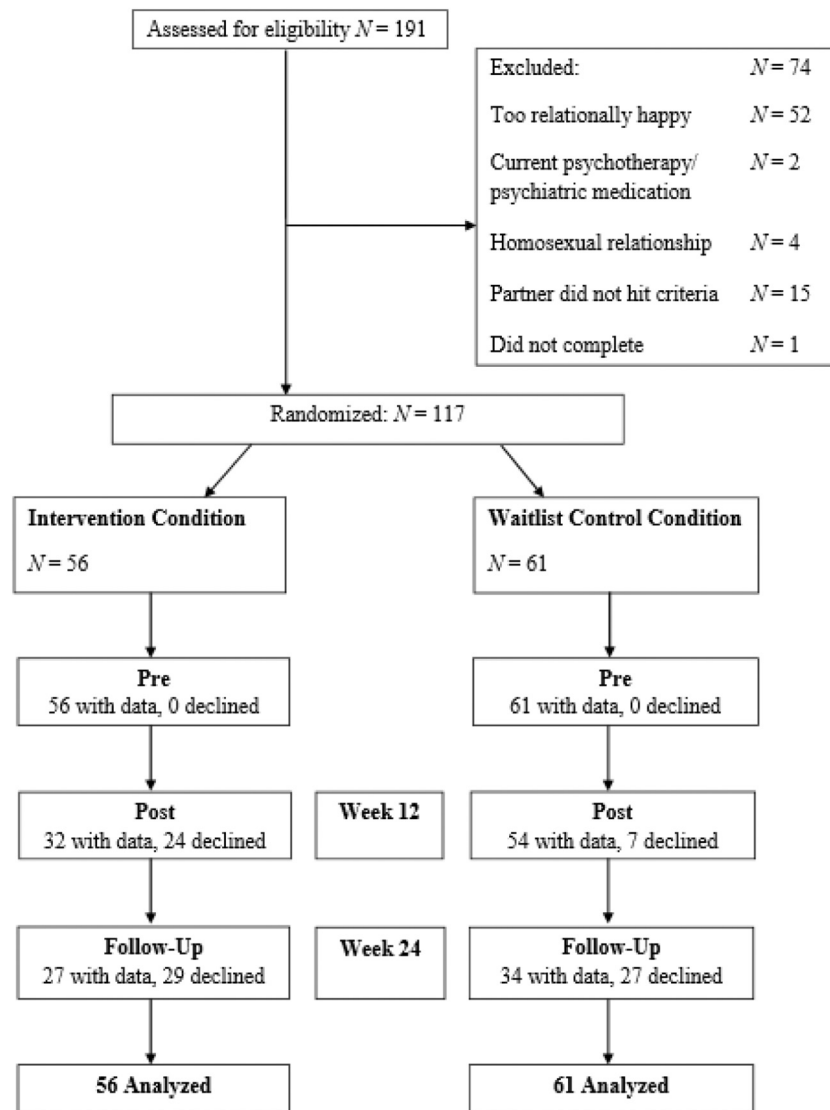
Fig. 1. CONSORT diagram.

2.3. Procedures and description of conditions

The recruitment started in May 2018 and ended in June 2019. After meeting the criteria and being randomly assigned (T1), the intervention group got access to the PaarBalance program. Every two weeks, participants in the intervention group received a short online questionnaire based on self-reported data to determine their current status in the program (e.g., number of sessions and homework) to assess their adherence. Participants who did not respond to the questionnaire received two reminder e-mails. During the 12-weeks treatment period, the waitlist control group did not receive access to the program. After the 12 weeks, all participants received the post-assessment questionnaires (T2) and the waitlist control group got access to the PaarBalance program for the following 12 weeks. Similar to the intervention group, they also received a short questionnaire every two weeks during their treatment period. After 12 weeks, again, all participants received the next and last assessment (T3). This was a 12-weeks-follow-up-measurement for the intervention group and a post-II-measurement for the waitlist control group directly after treatment.

2.4. Description of intervention

The Paarbalance program starts with the 64-item PaarBalance relationship questionnaire (Schindler et al., 2018). Directly after completing the questionnaire, clients receive the results of the questionnaire, that is, a relationship profile with the strengths and weaknesses of their relationship. The questionnaire is supposed to help the client identify difficulties in the intimate relationship. After finishing the questionnaire, participants get access to the 18 sessions PaarBalance program. Table 1 depicts their content. As mentioned above, the sessions are based on Integrative Behavior Therapy for couples (e.g., Baucom et al., 2019; Christensen and Glynn, 2019).

Every session includes a psychoeducative video with information about the respective topic. The video is followed by two exercises and three homework options related to the content of the video. Having finished the exercises, the client has to pick at least one homework option and enter a date until intending to complete it. The program reminds the client of the homework when logging into the program. The sessions have a predetermined order as they build upon each other. Once sessions are completed, they can be repeated as often as desired. Before and after every session, the client receives a standardized e-mail to increase the motivation of participation in the program. The content of the program can be shared with the partner, if desired. Gastner et al. (2018) recommend clients to complete one or two sessions each week to leave enough time for practice and homework. In addition to the sessions, the program offers features such as measuring the satisfaction of the client

Table 1
Content of the 18 sessions of the PaarBalance program.

Session number	Content
1	Sensitization for positive aspects of the relationship
2	Encouraging acceptance
3	Practice in helpful attributions
4	Activating resources and positive thinking
5	Training in reciprocity
6	Articulation of wishes versus acceptance and tolerance
7	Decreasing destructive communication
8	Communication training part 1: speaker skills
9	Communication training part 1: listener skills
10	Guidance for constructive conflict resolution
11	Development of self-care
12	Strengthening the bond
13	Helpful attitude towards sexuality and eroticism
14	Strategies for enriching sexuality
15	Handling jealousy
16	The importance of a committed relationship
17	Stress management
18	Relationship history and goals

with himself, with the partner, and with the relationship before every session. This process measure is illustrated to the user in a graph, allowing to monitor change over the 12 weeks. Furthermore, the relationship timeline can be used to record important events in the relationship. The client can also have a look at the quintessence of the sessions.

2.5. Measures

The primary outcome measure of the present study is relationship satisfaction. The secondary outcome measures of the study are depression, anxiety, quality of life, work functioning and satisfaction with the program.

2.5.1. Relationship satisfaction

Relationship satisfaction was measured by the total value of the short version of the PFB (Hahlweg, 2016), the PFB-K (Kliem et al., 2012). The PFB-K includes 10 items and has three subscales: *conflict behavior*, *tenderness* and *commonality/communication* and one extra item measuring the global relationship satisfaction. The PFB has shown good reliability ($\alpha = 0.84$; Kliem et al., 2012). The total value is composed of the three subscale values and ranges from 0 to 27 (= highest level of relationship satisfaction) measured on a 4-point Likert scale. The PFB has shown good discriminative and predictive validity (Hahlweg, 2016). In the present sample *Cronbach's alpha* was 0.67 (*conflict behavior $\alpha = 0.75$; tenderness $\alpha = 0.68$; commonality/communication $\alpha = 0.39$*).

2.5.2. Depression

Symptoms of depression were measured with the Patient Health Questionnaire (PHQ-9; German version: Löwe et al., 2002), a self-report measure with nine questions measuring characteristic symptoms of depression, such as feeling hopeless, having no motivation or interest, and suicidal thoughts on a 4-point Likert scale. The internal consistency, sensitivity and specificity for identifying depressive disorders of the PHQ-9 is good (e.g., Gräfe et al., 2004; Kroenke et al., 2010). PHQ-9 values between 10 and 27 indicate moderate to severe symptoms of depression. In the present sample *Cronbach's alpha* was 0.75.

2.5.3. Anxiety

With the Generalized Anxiety Disorder Scale (GAD-7; German version: Löwe et al., 2002) symptoms such as feeling anxious, nervous, being restless and worrying were measured on a 4-point Likert scale. The GAD-7 has shown good reliability and validity for measuring anxiety in the general population (Löwe et al., 2008). GAD-7 values between 5 and 21 indicate moderate to serve symptoms of anxiety. In the present sample *Cronbach's alpha* was 0.83.

2.5.4. Work functioning and quality of life

Quality of life and functioning were measured with two items: (1) "How would you describe your quality of life?" (Likert scale from 0 to 5: very low, low, rather low, rather high, high, very high; from the WHO (World Health Organization) Quality of Life BREF; (The WHOQOL Group, 1998); (2) "How would you rate your functioning? Please describe your work functioning or - if you are not working - your functioning at home" (Likert scale from 0 to 5: very low, low, rather low, rather high, high, very high).

2.5.5. Participants' satisfaction with the program

Right after participants had finished the program (T2 for intervention group, T3 for the waitlist control group), they were asked if the PaarBalance program helped them to increase their relationship satisfaction, if the most important relationship topics were included in the program, if they experienced pleasure completing the program, if they felt motivated enough, what aspect of the program they found most helpful, how they would describe their total satisfaction with the program, and if they would recommend it to friends and family.

2.6. Statistical analysis

All data analyses were done using R (R Core Team, 2017) and the package lme4 (Bates et al., 2015). Hierarchical linear modeling (HLM, Raudenbush and Bryk, 2002) addresses the dependency of data presented in longitudinal studies, due to repeated-measures (here: assessment of outcome measurements at baseline and after 12 and 24 weeks) nested within clients. HLM can handle missing data and mimic an intent-to-treat approach by including all participants into the analysis who completed outcome measures at least once.

We first ran *two-level fully unconditional models* with relationship satisfaction, depression, anxiety, quality of life, and level of functioning as outcome variables. Based on these five unconditional models, we calculated intraclass correlation coefficients to establish the variance explained by client effects (level 2). Results indicated that HLM is necessary.

To establish whether there was a significant modification in relationship satisfaction, depression, anxiety, quality of life, and functioning in the overall sample of the study, *time-as–only-predictor models* were calculated. Due to the design of the study, the time variable was fixed. Further, it was split into the first phase of the treatment (T1 to T2), where half of the sample received access to PaarBalance while the other half waited, and the second treatment phase (T2 to T3), where the waitlist control group was also given the active intervention. The time variable was defined as weeks and centered at week 12. This implies that the intercept of the model is interpreted as the estimated outcome variables at week 12, while the time slope is defined as the weekly rate of change in weeks 1–12 for T1 to T2 and weeks 12–24 for T2 to T3.

Based on these results we ran four *conditional models* with time in weeks as the only level-1 predictor as a fixed coefficient and centered at week 12, and with treatment condition (PaarBalance vs. waitlist) as level-2 predictor of both the intercept (estimated value of outcome variables in week 12) and the linear slope (weekly change in outcome variables between weeks 1 and 12 as well as 12 and 24) of the dependent variables.

Furthermore, we calculated effect sizes using Cohen's d by subtracting the means of the two subgroups and dividing the result of the subtraction by the pooled standard deviation with the weights for the sample sizes of both subgroups.

3. Results

3.1. Preliminary analyses

The results for the *time-as-only-predictor models* can be found in Table 1 of the appendix. These models significantly improved the model

fits compared to the *fully unconditional* models for relationship satisfaction ($\chi^2(5) = 27.28$, $p < .001$), depression ($\chi^2(5) = 10.58$, $p < .01$), anxiety ($\chi^2(5) = 23.76$, $p < .001$), and quality of life ($\chi^2(5) = 16.46$, $p < .001$). This was not the case for work functioning ($\chi^2(5) = 1.93$, $p = .38$).

3.2. Missing data and drop-outs

By the post survey, 42.7% of the intervention group and 11.5% of the control group had missing data due to no response or incompletion. By the 3-month follow-up, the intervention group had 51.8% and the control group 44.3% of missing data (see CONSORT diagram). Six participants (five from the intervention group, one from the waitlist control group) dropped out between the first and the second assessment. Two participants in the intervention group indicated their personal situation, two others indicated that they had not gotten round and one other indicated personal and professional stress as reasons for the drop-out. The participant in the waitlist control group declared that he did not want to participate anymore. Another three participants from the waitlist control group dropped out between the second and the third assessment. One indicated health issues, one a death in family and one that he did not want to participate anymore as the reasons for the drop-out. There were no significant differences between completers and dropouts neither on any measured patient characteristic at pretreatment nor on program engagement (all *ps* > 0.15).

3.3. Main results – treatment condition as predictor of outcome

We found significant effects of treatment modality on the estimated scores of relationship satisfaction at week 12 ($Y_{00} = 2.12$, $SE = 0.84$, *95% CI* [0.47, 3.77], $t(218) = 2.54$, $p < .05$) and the weekly rate of change in relationship satisfaction from week 1 to 12 ($Y_{10} = 0.24$, $SE = 0.06$, *95% CI* [−0.01, 0.03], $t(163) = 3.80$, $p < .001$) and week 12 to week 24 ($Y_{20} = -0.22$, $SE = 0.07$, *95% CI* [−0.01, 0.03], $t(155) = -3.06$, $p < .01$). Results revealed that clients who received PaarBalance in the first treatment phase had a 2.12 units higher estimated score in relationship satisfaction at week 12 than clients who were in the waitlist condition. In treatment phase one, the active intervention was related to an increase of 0.24 weekly units more in relationship satisfaction when compared to the waitlist control group. In treatment phase two, the waitlist group also received access to PaarBalance and improved by 0.16 weekly units in relationship satisfaction, which is 0.13 weekly units more than during their own waiting period (0.03). The intervention group maintained their improvements over follow-up (−0.07 weekly units) when compared to their weekly improvement by 0.26 units in relationship satisfaction over the intervention period. Additionally, there was a significant effect of treatment modality on the estimated scores of quality of life at week 12 ($Y_{00} = 0.37$, $SE = 0.18$, *95% CI* [0.02, 0.72], $t(227) = 2.07$, $p < .05$) and the weekly rate of change in quality of life from week 1 to 12 ($Y_{10} = 0.03$, $SE = 0.01$, *95% CI* [−0.01, 0.03], $t(156) = 2.20$, $p < .05$). Results indicate that clients in the intervention group presented a 0.37 higher estimated score in quality of life at week 12 than the clients without active treatment. Between weeks 1 and 12, this group increased by 0.03 units per week more, than the control group. The same models analyzing the effects of treatment modality on depression and anxiety did not reveal significant results.

Table 2 shows the mean and standard deviation (SD) of the outcome variables distinguished between conditions and assessment points as well as the between-group effect sizes at post-treatment.

3.4. Treatment adherence

On average, participants performed 7.2 ($SD = 2.7$) out of 18 sessions. 97% completed at least one session, 49.1% completed at least half of the program, and 21.2% completed most of the program. 4.3% of the participants involved their partner in the program via e-mail, 31.7% involved their partners during conversation, 7.7% showed or talked

Table 2
Estimated means of primary and secondary outcome measure and between-group effect sizes.

Measure	Pre-treatment	Post-treatment (estimated)	Follow-up (estimated)	Between-group effect sizes at post-treatment (estimated means)
	Mean	Mean (SE)	Mean	Cohen's d (95% CI)
Relationship satisfaction				
Treatment	12.45	15.62 (2.78)	14.82	0.77
Control	13.20	13.50 (2.74)	15.38	(0.39–1.14)
Depression				
Treatment	6.25	5.17 (2.40)	4.69	−0.43
Control	6.16	6.29 (2.83)	4.92	(−0.79 - -0.06)
Anxiety				
Treatment	5.84	4.39 (2.58)	3.72	−0.45
Control	5.84	5.60 (2.84)	3.95	(−0.81 - -0.08)
Quality of life				
Treatment	3.00	3.35 (0.50)	3.40	0.73
Control	3.02	2.98 (0.52)	3.43	(0.35–1.09)
Functioning				
Treatment	3.48	3.48 (0.31)	3.47	0.36
Control	3.43	3.36 (0.35)	3.64	(−0.00 - -0.73)

about a coaching video to the partner, and 11.1% reported that they used other ways to involve their partners.

3.5. Client satisfaction

Descriptive results of participants' evaluation of the program are presented in Table 3. Seventy-five percent of the intervention group and 63.9% of the control group reported an overall satisfaction between satisfied and very satisfied.

4. Discussion

4.1. Effects on relationship satisfaction

Past research has highlighted the importance of relationship satisfaction for mental and physical health. It has also suggested a high potential of web-based interventions. The primary aim of this study was to investigate the effects of the German PaarBalance program on relationship satisfaction. Consistent with our hypothesis, PaarBalance significantly improved relationship satisfaction in individuals and couples compared to the waitlist control condition ($d = 0.77$). This result is in accordance with findings by Doss et al. (2016), revealing significant positive effects of the OR program on relationship satisfaction ($d = 0.69$). Likewise, in Doss et al. (2020) an effect size of $d = 0.72$ for OR and $d = 0.61$ for ePREP was reported. Overall, the results underline the effectiveness of online couple interventions.

In the present study the waitlist control group showed no change in relationship satisfaction over the waiting period. This is in accordance

Table 3
Participants' perceived satisfaction with the program.

Item	Intervention group Mean (SD)	Control group Mean (SD)
Perceived Overall satisfaction	2.94 (1.16)	2.86 (0.88)
Perceived increase of relationship satisfaction	2.47 (0.92)	2.34 (0.91)
Perceived satisfaction with the content of the program	2.59 (0.98)	2.71 (0.83)
Perceived pleasure performing the program	2.88 (0.94)	2.69 (0.96)
Perceived motivation due to the program	2.56 (0.98)	2.49 (1.12)
Recommendation to friends and family	2.78 (1.16)	2.60 (1.22)

with past research indicating no spontaneous remission concerning relationship distress (Baucom et al., 2003). Once they received access to the program, this group also exhibited significant improvement in relationship satisfaction. Thus, PaarBalance might offer new and effective treatment options, especially to couples with little time, e.g. because of young children, when one or both partners have to work during normal business hours, for those who are in long-distance relationships, and/or live in underserved areas. In our case, the program even seemed to be attractive to couples with lower levels of relationship distress, as most subjects were excluded from the present study due to the high levels of relationship satisfaction, see Fig. 1.

4.2. Effects on depression, anxiety, quality of life and work functioning

In line with past research (e.g., Doss et al., 2016), we further examined the effects of the PaarBalance program on depression, anxiety, quality of life and work functioning. Effects on depression and anxiety were not significant. One reason explaining the nonsignificant effects on depression and anxiety in the present study may be the low severity of depression and anxiety symptoms at baseline leaving few room for improvement over the course of the intervention. Doss et al. (2016) found significant effects with respect to depression and anxiety at post-treatment. Anxiety was measured with the GAD-7 in both studies. Doss et al. (2016) reported baseline scores of 6.73 for men and 9.75 for women. In the present study, the sample presented an overall mean of 5.84 at treatment begin. This indicates higher levels of anxiety at pretreatment in the study of Doss et al. (2016). Symptoms of depression were measured with different instruments in the two studies and are therefore not directly comparable. Statistical analyses revealed a significant effect of PaarBalance on quality of life ($d = 0.73$) but no significant effects on work functioning. However, these results should be interpreted with caution, as quality of life and functioning were both measured using one item only and theoretical concerns have been raised about measuring constructs with a single-item. Besides, values of functioning were already relatively high at baseline (on average between rather high and high) and thus a ceiling effect may account for the nonsignificant results. Further, post-hoc power analyses revealed that the lack of significance for individual functioning versus relationship functioning could be due to the sample size as 117 participants provided high power for relationship satisfaction and quality of life but not for anxiety, depression, and work functioning thus impairing the probability of finding statistically significant differences when there are true differences.

We did not compute subgroup analyses accounting for initial problems of depression, anxiety, and functioning, because the sample size would have been too small to achieve meaningful results. Future studies with larger samples may further investigate whether depression, anxiety, and work functioning can be fostered by means of the PaarBalance program.

4.3. Participants' evaluation of the program

Overall, participants reported that they were satisfied with the program, enjoyed participation and most of them would recommend it to friends and family. Further, they rated their perceived increase in relationship satisfaction, satisfaction with the program's content, and motivation induced by the program between partially satisfied and satisfied on average. This demonstrates that the PaarBalance program was well received by users and supports the usefulness of the program in treating relationship distress. Doss et al. (2016) reported a satisfaction rate of 94%, while in the present study, 75% of the intervention group and 63.9% control group were satisfied with the program. Considering that the present study did not offer any additional phone calls or video-based contacts from the study team, this is a satisfactory result. In contrast to our study design, the participants in Doss et al. (2016) received four phone calls or videoconferences (about one hour each)

with a coach, supporting them in effective program usage. Similarly, in the study by Doss et al. (2020), participants were in contact with their coach every week. In the present study, we decided to make the program use as naturalistic as possible. The biweekly questionnaires to ensure adherence to PaarBalance was the only difference between normal program usage and usage as a participant in our study.

4.4. Limitations and future directions

Although PaarBalance has paved the way for online web-based relationship coaching in German-speaking countries, the present study comes with several limitations. First, this study has been compromised by the rather high drop-out-rate in the intervention group. Second, the study was underpowered to detect small to medium effect sizes found in secondary outcome measures such as depression and anxiety. This restricts the generalizability of the results. Third, the majority of participants (78%) had a university degree as their highest level of education. This also limits the generalizability of our findings to less educated groups. Future research is needed to evaluate the effectiveness of PaarBalance in a more diverse sample. Fourth, Cronbachs Alpha of the total value of the PFB was not as high as expected ($\alpha = 0.67$). As relationship satisfaction is the primary outcome measure, a more reliable measure would be preferable.

Another limitation is treatment adherence which was measured by means of self-report. On average, participants reported that they completed 7 out of 18 sessions. This is a common difficulty concerning unguided web-based interventions (Titov et al., 2013). In comparison with guided internet interventions which include weekly human support, unguided interventions with no human support tend to be associated with higher dropout rates (Melville et al., 2010), lower adherence (Christensen et al., 2009), and lower effects (Karyotaki et al., 2017). Future research could investigate adherence to and the efficacy of PaarBalance, when delivered in a guided format. Currently, the PaarBalance progam is tested in a blended format, in which the program is combined with couple counseling. Moreover, due to a limited budget we were not able to assess adherence by means of computerized data (e.g. completed lessons, number of log-ins, duration of individual lessons, etc.) but used self-reported data only. Computerized models may allow for predictions of program usage and may help to improve the program. For instance, it would be enriching to know how long participants on average need for completion of one session, how many sessions they perform per week, whether they repeat sessions, whether participants perform all 18 sessions and when they stop using the program. Interestingly, it would also allow to examine individual differences in program suitability and effectiveness.

Finally, the study strongly relies on self-reported data. However, when it comes to couple data explicit measures such as self-report data can be misleading (e.g. Farue et al., 2018; Joel et al., 2017). Future research should consider combining self-report data with implicit observational data by external observers.

5. Conclusions

To conclude, in the present study the PaarBalance program was effective in significantly increasing relationship satisfaction and quality of life in distressed couples. The program might offer a new evidence-based treatment opportunity especially for couples with few time, long-distance relationships and/or couples in underserved areas. Internet-based treatments for couples with relationship distress hold promise, however, more research is needed to substantiate these claims.

CRediT authorship contribution statement

LS, TB and AK designed the study. AB contributed significantly to the design and was responsible for the data analysis via HLM. AK wrote the first draft of the manuscript, AB, TB and LS revised and finalized the

manuscript. All authors read and approved the final manuscript.

Declaration of competing interest

Ludwig Schindler was involved in developing the PaarBalance program and is shareholder of the PaarBalance GmbH. The authors declare that they have no other conflicts of interest.

Appendix A. Time-as-only predictor models

Table 1
Effects of time on outcome variables.

Dependent variables	Estimated score at post treatment			Intervention period			Follow-up period		
	γ_{00}	SE	p	γ_{10}	SE	p	γ_{20}	SE	p
Relationship satisfaction	14.25	0.42	<0.001	0.12	0.03	<0.001	0.07	0.04	0.07
Depression	5.88	−0.39	<0.001	−0.03	−0.03	0.37	−0.03	0.03	0.37
Anxiety	5.15	0.39	<0.001	−0.06	−0.03	<0.05	−0.11	0.03	<0.01
Quality of life	3.12	−0.09	<0.001	0.009	0.007	0.22	0.02	0.008	<0.01
Work functioning	3.41	0.09	<0.001	−0.01	0.008	0.62	0.01	0.01	0.17

Appendix B. Data transparency statement

This manuscript and the material of this manuscript has not been submitted or published elsewhere. There are no previously published or in press works stemming from the same dataset.

Appendix C. Equations of HLM

C.1. Fully unconditional model

The equations for the fully unconditional models are as follows:
Level-1 Model

$$PFB/PHQ - 9/GAD - 7/QL/WF_{ij} = \beta_{0j} + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

At Level 1, the model estimated the scores of relationship satisfaction, depression, anxiety, quality of life and work functioning at moment i for client j by client j's average score across study time (β_{0j}). The random effect r_{ij} allowed client j to vary at time i from his average score across study time.

At Level 2, the average score for client j (β_{0j}) was predicted by the average sample mean score across study time (γ_{00}). The random effect u_{0j} allowed client j to vary from the average sample's score.

C.2. Unconditional time-as-only predictor model

The equations for the unconditional time-as-only-predictor model are as follows:
Level-1 Model

$$PFB/PHQ - 9/GAD - 7/QL/WF_{ij} = \beta_{0j} + \beta_{1j*(Linear\ Week\ Piece\ 1ij)} + \beta_{2j*(Linear\ Week\ Piece\ 2ij)} + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + u_{2j}$$

At level 1, the model estimated the scores of PFB/PHQ-9/GAD-7/QL/WF at moment i for client j, as a function of the level of the dependent variable at week 12 (β_0) and its rate of change during the first (β_1) and second 12 weeks (β_2) analyzed for client j. The random effect r_{ij} is the measurement error for client j at moment i.

At level 2, these coefficients dropped down to be predicted by the average of the dependent variable in week 12 (β_{00}) and the average change over the course of the first (β_{10}) and second 12 weeks (β_{20}) across all clients. The random effects u_{0j}, u_{1j}, and u_{2j} represent each client's deviation from (or variability around) the sample's average post-treatment score, piece 1 weekly change and piece 2 weekly change.

C.3. Conditional model with treatment modality as predictor

The equations for the conditional models are as follows:
Level-1 Model

$$PFB/PHQ-9/GAD-7/QL/WF_{ij} = \beta_{0j} + \beta_{1j*(\text{Linear Piece 1ij})} + \beta_{2j*(\text{Linear Piece 2ij})} + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(\text{treatment modality}_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}*(\text{treatment modality}_j) + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}*(\text{treatment modality}_j) + u_{2j}$$

At Level 1, PFB/PHQ-9/GAD-7/QL/WF$_{ij}$ is the score for client j at time i, which was predicted by the score in session 12 (β_{0j}) and the weekly rate of change during the first (β_{1j}; piece 1 slope) and second (β_{2j}; piece 2 slope) 12 weeks. The random effect r_{ij} is the measurement error for client j at moment i.

At Level 2, the estimated scores for client j at session 12 (β_{0j}) were predicted by the sample's average score in session 12 (γ_{00}) and treatment modality (γ_{01}). The rate of change over the first 12 weeks (β_{1j}) for client j was predicted by the sample average rate of change over the first 12 weeks (γ_{10}) and treatment modality (γ_{11}). The weekly rate of change over the second 12 weeks (β_{2j}) for client j was predicted by the sample average weekly rate of change over the second 12 weeks (γ_{20}) and treatment modality (γ_{21}). The random effects u_{0j}, u_{1j}, and u_{2j} represent each client's deviation from (or variability around) the sample average post-treatment score, piece 1 weekly change and piece 2 weekly change.

References

Amichai-Hamburger, Y., Klomek, A.B., Friedman, D., Zuckerman, O., Shani-Sherman, T., 2014. The future of online therapy. Comput. Hum. Behav. 41, 288–294. https://doi.org/10.1016/j.chb.2014.09.016.

Andersson, G., Titov, N., Dear, B.F., Rozental, A., Carlbring, P., 2019. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 18, 20–28. https://doi.org/10.1002/wps.20610.

Andrews, G., Basu, A., Cuijpers, P., Craske, M.G., McEvoy, P., English, C.L., Newby, J.M., 2018. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated metaanalysis. Journal of Anxiety Disorders 55, 70–78. https://doi.org/10.1016/j.janxdis.2018.01.001.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Baucom, D.H., Hahlweg, K., Kuschel, A., 2003. Are waiting-list control groups needed in future marital therapy outcome research? Behav. Ther. 34, 179–188. https://doi.org/10.1016/S0005-7894(03)80012-6.

Baucom, D.H., Fischer, M.S., Hahlweg, K., & Epstein, N.B. (2019). Cognitive behavioral couple therapy. In B.H. Fiese (Ed.), APA Handbook of Contemporary Family Psychology (Volume 3, pp. 257–274). American Psychological Association.

Baucom, D.H., Fischer, M.S., Corrie, S., Worrellm, M., Boeding, S., 2020. Treating Relationship Distress and Psychopathology in Couples. Routledge, A cognitive-behavioural approach.

Beer, R. (2020, 16. October). Theratalk.de. Eheberatung, Paartherapie und Partnschaftstests online. Theratalk: https://www.theratalk.de/.

Berger, T., Hämmerli, K., Gubser, N., Anderrson, G., Caspar, F., 2011. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn. Behav. Ther. 40 (4), 251–266. https://doi.org/10.1080/16506073.2011.616531.

Bodenmann, G., 2016. Lehrbuch der Klinischen Paar-und Familienpsychologie, 2. Auflage. Hogrefe.

Bodenmann, G. (2020, 16. October). Online Training. Glücklich zu zweit trotz Alltagsstress. Paarlife: https://www.paarlife.ch/paarlife-online-training/.

Braithwaite, S., Fincham, F.D., 2007. ePREP: computer based prevention of relationship dysfunction, depression and anxiety. J. Soc. Clin. Psychol. 26 (5), 609–622. https://doi.org/10.1521/jscp.2007.26.5.609.

Braithwaite, S., Fincham, F.D., 2009. A randomized clinical trial of a computer based preventive intervention: replication and extension of ePREP. J. Fam. Psychol. 23 (1), 32–38. https://doi.org/10.1037/a0014061.

Christensen, A., Glynn, S., 2019. Integrative behavioral couple therapy. In: Fiese, B.H., Celano, M., Deater-Deckard, K., Jouriles, E.N., Whisman, M.A. (Eds.), APA handbooks in psychology®. APA handbook of contemporary family psychology: Family therapy and training (p. 275–290). American Psychological Association. https://doi.org/10.1037/0000101-017.

Christensen, H., Griffiths, K.M., Farrer, L., 2009. Adherence in internet interventions for anxiety and depression. J. Med. Internet Res. 11 (2), e13 https://doi.org/10.2196/jmir.1194.

Cicilia, L., Georgia, E., Doss, B., 2014. Incorporating internet-based interventions into couple therapy: available resources and recommended uses. Aust. N. Z. J. Fam. Ther. 35, 414–430. https://doi.org/10.1002/anzf.1077.

Cuijpers, P., Donker, T., van Straten, A., Li, J., Andersson, G., 2010. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol. Med. 40, 1943–1957. https://doi.org/10.1017/S0033291710000772.

Cummings, E.M., Davies, P.T., 2010. Marital Conflict and Children. An emotional security perspective, Guilford.

Diener, E., Gohm, C.L., Suh, E., Oishi, S., 2000. Similarity of the relationships between marital status and subjective well-being across cultures. J. Cross-Cult. Psychol. 31, 419–436. https://doi.org/10.1177/0022022100031004001.

Doss, B.D., Benson, L.A., Georgia, E.J., Christensen, A., 2013. Translation of integrative behavioral couple therapy to a web-based intervention. Fam. Process 52, 139–153. https://doi.org/10.1037/t02125-000.

Doss, B., Cicila, L., Georgia, E.J., Benson, L.A., Roddy, K., Nowlan, K., Christensen, A., 2016. A randomized controlled trial of the web-based ourrelationship program: effects on relationship and individual functioning. J. Consult. Clin. Psychol. 84 (4), 285–296. https://doi.org/10.1037/ccp0000063.

Doss, B., Roddy, K., Nowlan, K., Rothman, K., Christensen, A., 2019. Maintenance of gains in relationship and individual functioning following the Online OurRelationship program. Behav. Ther. 50, 73–86. https://doi.org/10.1016/j.beth.2018.03.011.

Doss, B.D., Knopp, K., Roddy, M.K., Rothman, K., Hatch, S.G., Rhoades, G.K., 2020. Online programs improve relationship functioning for distressed low-income couples: results from a nationwide randomized controlled trial. J. Consult. Clin. Psychol. 88 (4), 283–294. https://doi.org/10.1037/ccp0000479.

Farue, R., Righetti, F., Seibel, M., Hofmann, W., 2018. Speech is silver, nonverbal behavior is gold: how implicit partner evaluations affect dyadic interactions in close relationships. Psychol. Sci. 29 (11), 1731–1741. https://doi.org/10.1177/0956797618785899.

Gastner, J., Schindler, L., Metz, N., Zuber, A., 2018. Internetgestützte Prävention und Behandlung von Paarproblemen am Beispiels von PaarBalance. In: Christiansen, H., Ebert, D., Röhrle, E. (Eds.), Prävention und Gesundheitsförderung Bd. VI, DGVT Verlag.

Georgia Salivar, E.J., Rothman, K., Roddy, M.K., Doss, B.D., 2018. Relative cost effectiveness of in-person and internet interventions for relationship distress. Fam. Process 59 (1), 66–80. https://doi.org/10.1111/famp.12414.

Gräfe, K., Zipfel, S., Herzog, W., Löwe, B., 2004. Screening psychischer Störungen mit dem "Gesundheitsfragebogen für Patienten (PHQ-D)": Ergebnisse der deutschen Validierungsstudie. Diagnostica 50 (4), 171–181. https://doi.org/10.1026/00121924.50.4.171.

Hahlweg, K. (2016). FPD. Fragebogen zur Partnerschaftsdiagnostik (2nd ed.). Hogrefe.

Joel, S., Eastwick, P. W., & Finkel, E. J. (2017). Is romantic desire predictable? Machine learning applied to initial romantic attraction. Psychological Science, 28, 1478–1489. doi:10.1177%2F0956797617714580.

Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., Mackinnon, A., Meyer, B., Botella, C., Littlewood, E., Andersson, G., Christensen, H., Klein, J.P., Schröder, J., Breton-Lopez, J., Scheider, J., Griffiths, K., Farrer, K., Huibers, M.J.H., Phillips, R., Gilbody, S., Moritz, S., Berger, T., Pop, V., Spek, V., Cuijpers, P., 2017. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 74, 351–359. https://doi.org/10.1001/jamapsychiatry.2017.0044.

Kiecolt-Glaser, J.K., Newton, T.L., 2001. Marriage and health: his and hers. Psychol. Bull. 127, 472–503. https://doi.org/10.1037/0033-2909.127.4.472.

Kliem, S., Job, A.-K., Kröger, C., Bodenmann, G., Stöbel-Richter, Y., Hahlweg, K., Brähler, E., 2012. Entwicklung und Normierung einer Kurzform des Partnerschaftsfragebogens (PFB-K) an einer repräsentativen deutschen Stichprobe [Development and standardization of a short form of the Partnership Questionnaire (PFB-K) on a representative German sample]. Zeitschrift für Klinische Psychologie und Psychotherapie: Forschung und Praxis 41 (2), 81–89. https://doi.org/10.1026/16163443/a000135.

Kroenke, K., Spitzer, R.L., Williams, J.B., Löwe, B., 2010. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen. Hosp. Psychiatry 32 (4), 345–359. https://doi.org/10.1016/j.genhosppsych.2010.03.006.

Lebow, J. L., Chambers, A. L., Christensen, A., & Johnson, S. M. (2012). Research on the treatment of couple distress. J. Marital. Fam. Ther., 38, 145–168. doi:https://doi.org/10.1111/j.1752-0606 .2011.00249.

Loew, B., Rhoades, G., Markman, H., Stanley, S., Pacifici, C., White, L., Delaney, R., 2012. Internet delivery of PREP-based relationship education for at-risk couples.

Journal *of Couple & Relationship Therapy* 11 (4), 291–309. https://doi.org/10.1080/15332691.2012.718968.

Löwe, B., Spitzer, R.L. Zipfel, S., & Herzog, W. (2002). *PHQ-D. Gesundheitsfragebogen für Patienten (2nd ed.)*. Pfizer.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., Herzberg, P.Y., 2008. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med. Care 46 (3), 266–274. https://doi.org/10.1097/MLR.0b013e318160d093.

Markman, H., Rienks, S., Stanley, S.M., 2009. The prevention and relationship enhancement program (PREP). In: Reis, H.T., Sprecher, S. (Eds.), Encyclopedia of Human Relationships. Sage.

Melville, K.M., Casey, L., Kavanagh, D.J., 2010. Dropout from internet-based treatment for psychological disorders. Br. J. Clin. Psychol. 49 (4), 455–471. https://doi.org/10.1348/014466509X472138.

Proulx, C.M., Helms, H.M., Buehler, C., 2007. Marital quality and personal well-being: a meta-analysis. J. Marriage Fam. 69, 576–593. https://doi.org/10.1111/j.1741-3737.2007.00393.x.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, URL. http://R-project.org/.

Raudenbush, S.W., Bryk, A.S., 2002. Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage.

Robles, T.F., Kiecolt-Glaser, J.K., 2003. The physiology of marriage: pathways to health. Physiol. Behav. 79 (3), 409–416. https://doi.org/10.1016/s0031-9384(03)001604.

Schindler, L., Gastner, J., Henrich, G., Metz, N., 2018. Die Entwicklung des PaarBalance Partnerschaftstests. Beratung Aktuell 1, 3–24.

Scott, K.M., et al., 2010. Gender and the relationship between marital status and first onset of mood, anxiety and substance use disorders. Psychol. Med. 40, 1495–1505. https://doi.org/10.1017/S0033291709991942.

Shadish, W.R., Baldwin, S.A., 2005. Effects of behavioral marital therapy: a meta-analysis of randomized controlled trials. J. Consult. Clin. Psychol. 73 (1), 6–14. https://doi.org/10.1037/0022-006X.73.1.6.

The WHOQOL Group, 1998. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol. Med. 28, 551–558. https://doi.org/10.1017/S0033291798006667.

Titov, N., Dear, B.F., Johnston, L., Lorian, C., Zou, J., Wootton, B., Spence, J., McEvoy, P. M., Rapee, R.M., 2013. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: randomised controlled trial. PLoS One 8 (7), e62873. https://doi.org/10.1371/journal.pone.0062873.

Vaillant, G.E., 2002. Aging Well. Boston: Little: Brown & Company.

Whisman, M. A. (2019). Psychopathology and couple and family functioning. In B. H. Fiese, M.Celano, K. Deater-Deckard, E. N. Jouriles, & M. A. Whisman (Eds.), APA Handbook of Contemporary Family Psychology: Applications and Broad Impact of Family Psychology. Family therapy and training (p. 3–20). American Psychological Association.

Whisman, M.A., Robustelli, B.L., Labrecque, L.T., 2018. Specificity of the association between marital discord and longitudinal changes in symptoms of depression and generalized anxiety disorder in the Irish longitudinal study on ageing. Fam. Process 57 (3), 649–661. https://doi.org/10.1111/famp.12351.