Do high-risk couples profit more or less from couple relationship education programs than low-risk couples? Room for improvement and vulnerability effects

Conradi, H.J.; Noordhof, A.; Dingemanse, P.; Kamphuis, J.H.

[Link to publication](#)

ORIGINAL ARTICLE

FAMILY PROCESS

Do high-risk couples profit more or less from couple relationship education programs than low-risk couples? Room for improvement and vulnerability effects

Henk Jan Conradi[1] | **Arjen Noordhof[1]** | **Pieter Dingemanse[2]** | **Jan H. Kamphuis[1]**

[1]Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands

[2]Division of Affective Disorders, Mental Health Care Altrecht, Utrecht, The Netherlands

Correspondence

Henk Jan Conradi, University of Amsterdam, Postbus 15933, 1001 NK Amsterdam, The Netherlands.
Email: h.j.conradi@uva.nl

Abstract

In recent years it has been discussed whether high-risk couples benefit *more* from Couple Relationship Education programs (CREs) than low-risk couples due to larger room for improvement, or profit *less* due to greater vulnerability. Pertinent response prediction studies yielded inconclusive results. Careful review suggests this may be due to: statistical handling (not disentangling room for improvement and vulnerability effects), time frame analyzed (not disentangling opposing effects during intervention and follow-up), sampling, and selection of risk factors. We used an analytic strategy that maximized odds for replicability and tested two hypotheses: (1) *room for improvement*: pre-intervention relationship dissatisfaction predicts gain in satisfaction during intervention, and decline during follow up, and (2) *vulnerability*: when adjusted for room for improvement (pre-intervention relationship dissatisfaction), risk factors show negative or negligible, but no positive associations with gain in satisfaction. Actor-Partner Interdependence Modeling (APIM) was employed in 79 self-referred (SR) couples and 50 clinician-referred (CR) couples who had completed the 'Hold me Tight' program, a CRE based on Emotionally Focused Couple Therapy. Our findings supported both: (1) the room for improvement hypothesis, with pre-intervention dissatisfaction predicting more gain during intervention (both samples) and decline during follow-up (SR sample, for the CR sample the effect was negligible), and (2) the vulnerability hypothesis, as several negative, but *no* positive effects of risk factors were observed during intervention and follow-up. Specific risk

factors did not replicate between samples. To promote replicable results in future research, we advocate disentangling room for improvement and vulnerability effects, separately testing effects during intervention and follow-up, purposeful sampling, and studying a large set of risk factors including partner variables.

KEYWORDS
couple relationship education programs, high-risk couples, response prediction, room for improvement effects, vulnerability effects

INTRODUCTION

Couple Relationship Education (CRE) programs were originally reserved for couples at an early stage of relationship distress to prevent problems from becoming chronic and harder to treat (Bradford et al., 2015). In part because of US government funding of CRE programs for high-risk groups, increasing numbers of more severely relationally distressed couples have attended CRE programs (Bradford et al., 2015). This diversity of participating couples triggered a debate whether more severely distressed couples profit more rather than less from CRE programs than non- or mildly distressed couples. Distressed couples may profit *more* from CRE programs because they have more severe problems that can be targeted and changed (see a review by Halford & Bodenmann, 2013). This 'room for improvement' hypothesis states that couples who report more relational dissatisfaction at the start of an intervention will show more change during intervention. The opposite position, however, states that distressed couples profit *less* from CRE programs because they have vulnerabilities that cannot be sufficiently modified by these interventions. This 'vulnerability' hypothesis dictates that couples who report higher scores on risk factors are more vulnerable at the start of an intervention and will show less change (see a meta-analysis by Arnold & Beelmann, 2019).

These hypotheses are not mutually exclusive: distressed couples may initially profit more from CREs due to room for improvement, but over time profit less due to unyielding vulnerability. In fact, multiple scenarios are conceivable: (1) neither of both effects may be present, (2) one effect may be stronger than the other, or (3) both effects may cancel each other out. The current literature does not consistently support either position as the effects of risk factors have been found to be positive (Allen et al., 2012; Amato, 2014; Cowan et al., 2014; Dupree et al., 2016; Halford et al., 2001; Job et al., 2017; Mitchell et al., 2015; Petch et al., 2012; Roberson et al., 2020; Williamson et al., 2015), negative (Dupree et al., 2016; Halford & Wilson, 2009; Job et al., 2017; Williamson et al., 2015) or nonsignificant (Allen et al., 2012; Barton et al., 2014, 2017; Carlson et al., 2017; Cowan et al., 2014; Dupree et al., 2016; Gordon et al., 2019; Gubits et al., 2014; Halford & Wilson, 2009; Hsueh et al., 2012; Job et al., 2017; Kennedy, 2017; Lundquist et al., 2014; Petch et al., 2012; Stanley et al., 2014; Van Widenfelt et al., 1996; Williamson et al., 2015). We conjectured that these inconsistencies may result from methodological differences and that a more consistent pattern of outcomes might be found after taking these into account. Therefore, we closely examined previous studies in terms of: (1) sampling, (2) the time frame during which response is predicted, (3) statistical analyses conducted, and (4) the specific risk factors included.

Accordingly, we conducted an informal review of 26 CRE studies (included in the reference list marked with an '*' and/or a '+') that utilized relationship satisfaction as outcome variable. A complete overview of characteristics of the included studies is presented in Table S1. Overall, this review clearly demonstrates that lack of replicability is a major concern for this line of research. More specifically, the response prediction studies included a total of 33 risk factors (Table S2), of which only six were consistently nonsignificant. The large majority were either included in only one study (20 risk factors), and are therefore still open to replication, or were inconsistent over studies (seven risk factors). Hence, no single *significant* risk factor was consistently replicated.

Upon closer inspection, it became evident that the set of reviewed studies markedly differed in several important design and analytical respects that may well contribute to the inconsistent findings and difficulties in replication. First, sampling differences between studies were observed in terms of severity of relational distress and supposedly in terms of underlying vulnerabilities. Most of the reviewed studies (18 out of 26) concerned nondistressed samples, whereas only two studies examined distressed samples (the remaining six studies did not provide sufficient information in this respect). This is relevant, as in samples that include predominantly relationally nondistressed couples (e.g., Allen et al., 2012; Petch et al., 2012) a ceiling effect may preclude substantial room for improvement. If a couple is already functioning rather well, a CRE program will not easily improve their relationship.

Second, studies varied in the time period over which predictions were made. Positive effects associated with room for improvement are most likely to occur during intervention. Couples with acute distress may be susceptible to the provision of hope (remoralization) offered by a sound intervention rationale and, especially in distressed couples, starting an intervention together may signal enhanced commitment. Some of such benefit may well be short-lived and fade during follow-up if it does not co-occur with more substantial changes in underlying causes of problems (i.e., vulnerabilities). Unfortunately, of the studies reporting risk factor analyses that we reviewed, the vast majority (16 out of 21) did not disentangle the intervention and follow-up periods, but instead collapsed both into one analysis (e.g., Barton et al., 2017; Halford et al., 2001). Collapsing these time periods produces a confounding of room for improvement effects that are likely to be positive during intervention and negative during follow-up, and negative vulnerability effects. The 'net effect' of both then presumably depends on the nature of the sample, intensity of the intervention, and duration of the follow-up, which probably results in more inconsistent findings and difficulties in replication.

Third, response prediction studies varied considerably in statistical analyses. Of particular importance is whether they control for room for improvement effects in their risk factor analyses (e.g., Carlson et al., 2017; Job et al., 2017) or not (e.g., Amato, 2014; Lundquist et al., 2014). Controlling for room for improvement effects means including pre-intervention assessment of the outcome, in our study relationship dissatisfaction, as a predictor of change of satisfaction during both intervention and follow-up. Statistically this means including pre-intervention dissatisfaction as moderator, which is a direct test of the 'room for improvement' hypothesis, as it assesses whether pre-intervention dissatisfied couples profit more (or less) from an intervention. After such controlling for room for improvement, it can be tested whether other risk factors (e.g., attachment, demographics, etc.) show incremental predictive value indicating 'pure' vulnerability effects. In our literature review we therefore: (1) examined room for improvement effects by reviewing the results of studies that separately tested for dissatisfaction as moderator, and (2) compared results of studies examining effects of risk factors that did control for room for improvement (dissatisfaction as moderator) with studies that did not.

Separate analyses of room for improvement effects were reported by 12 of the 26 reviewed studies. Seven studies examined this during intervention and reported near consistent positive effects, meaning that pre-intervention dissatisfaction was associated with more gain. Unfortunately, no study reported an analysis of room for improvement effects during follow-up. Finally, the six studies that collapsed the intervention and follow-up time frames found inconsistent results (three found positive effects, one only positive effects for females, and two nonsignificant effects). This suggests that the summation of room for improvement effects, positive during intervention and negative during follow-up, may have produced inconsistent results across studies.

Importantly, almost all studies (18 out of 21) failed to control for room for improvement effects (i.e., they did not include pre-intervention dissatisfaction as moderator) when conducting risk factor analyses (one study was unclear in this respect). Accordingly, mixed findings were observed across intervention and intervention plus follow-up: risk factors were associated with positive, negative, or non-significant effects (Table S2). Two studies that did control for room for improvement effects in their risk factor analyses (Carlson et al., 2017; Job et al., 2017) found negative and non-significant

effects, but, importantly, *no* positive effects of high risk (Table S2). These findings founded our expectation that *not* controlling for room for improvement when conducting risk factor analyses results in negative, negligible or positive effects, whereas controlling for it leaves risk factors to have only negative or negligible effects.

A second important statistical issue is whether univariate or multivariate analyses were conducted. Of the reviewed studies reporting risk factor analyses the majority reported multivariate analyses (12 out of 21). By definition, multivariate analyses examine multiple risk factors simultaneously. As is widely recognized, the results of these procedures are quite dependent on the specific set of predictors included in the analyses, which is amplified by the particular characteristics of the selected sample. Indeed, no replications were notable in the multivariate analyses we reviewed.

Fourth, concerning the selection of risk factors included in response prediction studies, we note two further limitations of the extant research base: no consistency of the selection of predictors across studies limiting the possibility of replication (Table S2), and omission of partner effects. Studies were commonly restricted to actor effects, or the prediction of response to intervention within one individual. Of the 21 studies that reported risk factor analyses, only two studies (Halford & Wilson, 2009; Job et al., 2017) examined partner effects, or the effect of one partner on the other partner. This omission is unfortunate given that couples' dynamics are dyadic by nature (Levenson et al., 2015), and given the widespread clinical recognition of the importance of partner effects. For example, application of secure or insecure attachment strategies by one partner may have the effect on the other partner of feeling validated, or attacked or neglected. Such systemic effects are usually discussed between partners during intervention and used as vehicle to enlarge reciprocal understanding. Therefore, it seems important that scientific research catches up with clinical practice and includes partner effects.

In sum, due to variability in sampling, time frame analyzed, statistical handling, and selection of risk factors, response prediction studies fail to replicate, which in turn contributes to the trap of post hoc interpretations. Hence, we explicitly aimed to move toward more general testable and replicable ideas and hypotheses.

The present empirical study: two broad hypotheses in two samples of different vulnerability

In the current study we analyzed data from an effectiveness study on the Hold me Tight (HmT) intervention (Conradi, Dingemanse, et al., 2018), a CRE program (Johnson, 2008) based on Emotionally Focused Couple Therapy (EFCT). This study involved two independent samples including the same set of pre-intervention risk factors and the same time intervals for measuring relationship satisfaction during intervention and follow-up, but different levels of vulnerability. For this reason the data allow for testing of replication, which is a core focus of the present study. Based on our review, we tested two broad a priori hypotheses.

1. *Room for improvement*. Pre-intervention dissatisfaction predicts:
   Hypothesis 1a: more gain in satisfaction during the intervention;
   Hypothesis 1b: more decline during follow-up;
   Hypothesis 1c: from this it follows that when aggregating positive effects during intervention (1a) and negative effects during follow-up (1b) mixed results (gain, decline, or negligible effects) can be expected during intervention plus follow-up.[1]

---

[1]Hypotheses 1c and 2b–d all involve predictions of 'mixed results': negative, positive, and negligible effects. These clearly are not the way Popper would have liked hypotheses to be. In fact, none of the hypotheses can be tested by themselves as they will always be true. However, they systematically specify the broad hypotheses 1 and 2, and can be evaluated by the comparison between results: the broad hypotheses are supported if mixed results disappear when controlling for moderation of pre-intervention dissatisfaction and distinguishing between intervention and follow-up.

2. *Vulnerability*. When controlled for room for improvement effects (including pre-intervention dissatisfaction as moderator of change) risk factors predict:

> *Hypothesis 2a*: negative effects (less gain during intervention or more decline during follow-up) or negligible effects, but no positive effects during any time frame.

By contrast, when room for improvement is *not* controlled for (pre-intervention dissatisfaction is not included as moderator), risk factors predict:

> *Hypothesis 2b*: mixed results (more gain, less gain, or negligible effects) during intervention due to the summation of negative vulnerability effects (2a) and positive room for improvement effects (1a);
>
> *Hypothesis 2c*: negligible effects or decline during follow-up due to the summation of negative vulnerability effects (2a) and negative room for improvement effects (1b);
>
> *Hypothesis 2d*: mixed results (more gain, less gain, or negligible effects) during intervention plus follow-up due to the summation of effects (2b+2c).

METHOD

Participants and procedure

Heterosexual couples suffering from relational distress were included in the study. Couples who met criteria for individual DSM-IV axis-I (syndromes) or axis-II (personality disorders) diagnoses were excluded. Relational distress was determined by experienced clinicians via unstructured clinical interviews. Self-referred (SR) couples were recruited by advertising, the EFT website, and private practices in the EFT network. Clinician-referred (CR) couples were transferred by clinicians from the specialized mental health organization 'Altrecht' in Utrecht, the Netherlands. Partners were asked to sign an informed consent before entering the study. The study was approved by the Ethics Review Board of the University of Amsterdam (2011-CP-1826), and was registered in The Netherlands Trial Register (NTR4482). The study ran from August 2011 to August 2015.

The SR and CR samples consisted of 79 and 50 couples, respectively. CR couples were slightly younger than SR couples (41.1 vs. 44.8 years of age) and reported significantly less favorable on most predictors. Specifically, they had lower rates of high educational attainment (45.4% vs. 74.7%), were less likely to be married (49% vs. 82.2%), or to have children (49% vs. 86.2%), and their relationship was of shorter duration (13.5 vs. 17.8 years). On psychological measures, CR couples reported lower Self-disclosure, and higher attachment Avoidance and Anxiety (indicating highly insecure attachment compared to norms obtained from a representative national population sample; Conradi et al., 2018). On the other predictors, including motivation to engage in HmT, no differences were found between SR and CR couples (Table S3). Furthermore, pairwise comparisons of the highest and lowest scoring groups on each of the predictors strongly suggest that both clinical and nonclinical ranges were represented in our samples (Table S3).

Intervention

Hold me Tight (Johnson, 2008) consists of eight weekly 2-h sessions delivered by pairs of formally licensed therapists to groups consisting of on average 5.86 couples. HmT is a CRE program, as defined by Markman and Rhoades (2012), by its format as a workshop for groups of couples in which psychoeducation on relationships is provided combined with skills training through exercises and homework assignments. HmT may be applied as a preventive or as a relationship enrichment course (Johnson, 2008). Hence, HmT was designed to enhance relationship functioning or to reduce moderate levels of distress, but was not developed to tackle severe problems like current affairs or grave mental health problems. Intervention integrity checks showed that adherence to the core of the

protocol was excellent. On average couples completed 90% of the curriculum, which is equivalent to 14.4 h of attendance.

Measures

(Dis)satisfaction with the relationship (dissatisfaction as predictor and satisfaction as dependent variable) was measured with the corresponding 10-item subscale of the Dyadic Adjustment Scale (DAS; Spanier, 1976). We left out the Cohesion, Affection and Consensus subscales to avoid overlap with the set of predictor variables and because we more frequently administered the Satisfaction scale than the full DAS. An example item is 'In general, how often do you think that things between you and your partner are going well?' Cronbach's alphas were satisfactory ranging from 0.82 to 0.91 across sexes and SR and CR samples.

Education was determined by a general questionnaire. Low education is seen as an indicator for being 'socially disadvantaged' (Halford & Bodenmann, 2013).

Degree of relationship establishment was operationalized by relationship duration, marital status, having children and relationship commitment/stability.

Relationship duration, determined by general questionnaire, is a marker for investment in the relationship. Longer duration presumably means larger investments making separation less attractive (Atkins et al., 2005).

Marital status (general questionnaire) is a potential risk factor as cohabitation is associated with lower relationship quality than being married (Stanley et al., 2006).

Having children (general questionnaire) is considered a constraint to separation (Stanley & Markman, 1992).

Relationship commitment/stability was assessed with two items of the Relational Interaction Satisfaction Scale (RISS; Buunk, 1990). Lower levels suggest worse outcomes (Stanley et al., 2010). The items, for example, 'I am considering leaving my partner', are scored on a Likert scale ranging from 1 (never) to 5 (always). Cronbach's alphas were adequate ranging from 0.70 to 0.84 across sexes and SR and CR samples.

Attachment in romantic relationships was measured with the Experiences in Close Relationships questionnaire (ECR; Brennan et al., 1998). Insecure attachment is a well-established vulnerability factor (Conradi et al., 2011). The ECR comprises two subscales of 18 items each: *Anxiety* about rejection and abandonment or the expectation of being perceived by partners as unlovable, for example, 'I worry about being abandoned,' and *Avoidance* of intimacy or the expectation of inaccessibility and unresponsiveness of partners to one's attachment needs, for example, 'I find it difficult to allow myself to depend on my romantic partner'. Items are scored on a 7-point Likert scale ranging from 1 (disagree strongly) through 4 (neutral/mixed) to 7 (agree strongly). The Dutch ECR is reliable and valid (Conradi et al., 2006). Cronbach's alphas were consistently high ranging from 0.85 to 0.91 across sexes and SR and CR samples.

Secure bond with the partner, the main HmT target, was assessed with the 15 item Accessibility, Responsiveness, Emotional Engagement questionnaire (ARE; Johnson, 2008). Items tap attachment-related aspects of the bond, for example, 'If I need connection and comfort, he/she will be there for me' and are rated from 1 (disagree) to 5 (agree). Cronbach's alphas were excellent ranging from 0.94 to 0.96 across sexes and SR and CR samples.

Self-disclosure to the partner, regarded a prerequisite for the development of secure attachment in HmT, was measured with the adapted Self-Disclosure Index (SDI; Miller et al., 1983). The seven items, for example, 'I disclose my most intimate thoughts and feelings to my partner' are rated from 1 (totally disagree) to 5 (totally agree). Cronbach's alphas were high ranging from 0.85 to 0.93 across sexes and SR and CR samples.

Psychological complaints, a marker of vulnerability and a source of stress in relationships (Halford & Bodenmann, 2013), were measured by the General Health Questionnaire (GHQ; Koeter

& Ormel, 1991). The 12 items, for example, 'Have you recently lost much sleep over worry?,' are scored on a scale with values depending on the actual item, for example, from 1 (not at all) to 4 (much more than usually). Cronbach's alphas were high ranging from 0.89 to 0.94 across sexes and SR and CR samples.

Analytic strategy and statistical analyses

As stated above, our analytic strategy was specifically designed to yield replicable findings. First, to test for sampling effects we included two independent samples of couples (i.e., SR and CR). Second, to distinguish different effects at different time-periods we tested our expectations separately during intervention, during follow-up, and during intervention plus follow-up. Third, we included pre-intervention actor relationship Dissatisfaction as a predictor of change over time in all analyses to distinguish room for improvement and vulnerability effects. Fourth, we ran separate analyses for each risk factor instead of a full multivariate analysis with all risk factors. Fifth, we included a broad set of risk factors to enhance the chance of replication with previous and future studies. Finally, to fill the gap in studies on partner effects we examined partner effects as well as actor effects. Altogether, we think that this research design offers ample opportunity for rejection of our hypotheses, as for each risk factor at each time frame and in each sample a result may be obtained that is inconsistent with our hypotheses. Given this large number of tests, setting $\alpha < 0.05$ (two-tailed) even further increases the probability of rejecting the main hypotheses. Therefore, should our hypotheses not be rejected we think we have ground for expecting replicability of results.

To analyze both actor and partner effects of the predictors we applied an Actor-Partner-Interdependence Model (APIM) framework (Cook & Kenny, 2005). The specific APIM analyses warrant elaborate description. Specifically, we used multilevel modeling as implemented by Linear Mixed Models in SPSS (Bryk & Raudenbush, 1987) with couple as the unit of analysis. Dependency of partner scores was accounted for by including a random intercept at the couples' level, and dependency between the repeated measurements of Satisfaction by inclusion of an autoregressive model (AR1). An additional benefit of a Linear Mixed Model is the possibility of dealing with partially missing data.

For each predictor separately we computed a model including: the fixed main effects of sex, time, and both the actor and partner predictor at pre-intervention and the five 2-way interactions (sex*time, sex*actor predictor, sex*partner predictor, time*actor predictor, and time*partner predictor). In analyses controlling for room for improvement pre-intervention Dissatisfaction was included as a covariate along with its two-way interactions with sex and time. In analyses without this control, these covariates were left out. Because relationship Duration, Marital status, and having Children are predictors at the couples' level, no partner effects were included in these models. Dependent variables were Satisfaction during HmT (assessed prior to sessions 1, 3, and 5 and immediately after the final session 8), during follow-up (session 8 and at follow-up-1 and -2, i.e., 2.75 and 14.27 weeks after session 8, respectively), and during HmT plus follow-up (all measurement points). Inspection of the residuals of Satisfaction revealed normal distributions. Analyses were based on the intention-to-treat sample.

There is no consensus concerning methods for power estimation of models with three levels, that is, couples (level 3), individual partners (level 2) and repeated assessments (level 1). Based on simulation studies, Maas and Hox (2005) recommend a sample size of ≥50 at level 2. Hence, we assume that our sample sizes would be sufficient, with 158 partners in the SR sample and 100 partners in the CR sample.

In the main analyses we did *not* dichotomize continuous variables. However, we post hoc calculated effect sizes based on dichotomized variables to illustrate the direction and strength of significant effects. This was done by splitting each sample into halves with a high-risk subgroup (scoring less favorably) and a low-risk subgroup (scoring more favorably). We computed Cohen's d effect sizes for these high- and low-risk groups based on the estimated marginal means and the pooled SDs of the corresponding raw means (Feingold, 2009). Effect sizes can be interpreted as small when $d = 0.2$,

moderate when $d = 0.5$, and large when $d = 0.8$ (Cohen, 1977). We interpreted differences in effect size of high- and low-risk groups with $d < 0.1$ as negligible.

RESULTS

Preliminary analyses

Pre-intervention predictor assessments were missing in 3.1%–4.3% of the cases, depending on the specific risk factor. Regarding the dependent variable, Satisfaction, nonresponse was higher, that is, 3.5% at pre-intervention, 10.9% at session 3, 20.5% at session 5 (partly due to postal failure), 13.6% postintervention, 17.8% at follow-up 1, and 19.4% at follow-up 2. To evaluate whether missingness was random, we examined if nonresponse at follow-up-2 affected the course of Satisfaction by running a linear mixed model as described above adding (non-)response to follow-up-2 and its interactions with time and sex as fixed factors. Because time*(non-)response at follow-up-2 was significant in both samples, we adjusted for nonresponse in each prediction model described above by adding the main effect of nonresponse at follow-up-2 along with its interaction with time.

Next, we plotted the course of Satisfaction, the dependent variable (Figure S1). Both SR and CR couples showed change during HmT at postintervention, with $d = 0.32$ in SR couples and $d = 0.29$ in CR couples. From postintervention to follow-up SR couples remained stable with $d = 0.02$ resulting in a total gain from pre-intervention to follow-up of $d = 0.34$. CR couples largely reversed from postintervention to follow-up with $d = -0.17$, resulting in a small total gain from pre-intervention to follow-up of $d = 0.11$.

Room for improvement: Effects of pre-intervention dissatisfaction on the course of satisfaction

Pre-intervention actor Dissatisfaction was included as a moderator to test the room for improvement hypothesis. Actor Dissatisfaction showed significant two-way interactions with time in both samples, across all three time periods. As expected (hypothesis 1a), during intervention, Dissatisfaction was a significant predictor of change in both samples. Post hoc analysis indicated that in more dissatisfied groups more gain was obtained than in less dissatisfied groups ('↑' in Table 1).

Pre-intervention actor Dissatisfaction also predicted change during follow-up (hypothesis 1b). In the SR sample, consistent with our hypothesis, post hoc analyses indicated that in the more dissatisfied group more decline was found than in less dissatisfied group ('↓' in Table 1). However, in the CR sample the difference in effect-size was negligible ('≈' in Table 1).

Finally, during intervention plus follow-up we anticipated mixed results due to the summation of positive (1a) and negative (1b) effects (hypothesis 1c). In both our samples this summation turned out to be net more gain for more dissatisfied groups, due to larger effect sizes for the positive effects.

Vulnerability: effects of risk factors on the course of satisfaction while controlling for pre-intervention dissatisfaction as moderator

We tested the vulnerability hypothesis (2a) that, after controlling for room for improvement (including pre-intervention Dissatisfaction as moderator), all risk factors would show negligible or less gain, but not more gain across all time periods. During intervention, findings across samples were in support of this hypothesis. In the SR sample five risk factors showed significant two-way interactions with time; post hoc analyses revealed that three predictors, not having Children, lower partner Secure bond and lower partner Commitment/stability, indicated less gain in Satisfaction ('↓' in Table 1),

whereas for Marital status and partner Dissatisfaction the difference at the end of intervention between low-risk and high-risk groups was negligible ('≈' in Table 1). In the CR sample two risk factors showed significant two-way interactions with time during intervention. Post hoc analyses showed that high actor Anxiety was associated with substantially less gain during intervention, whereas the difference between low and high actor Education was negligible.

During follow-up, results were again in support of hypothesis 2a across samples. In the SR sample no risk factors significantly predicted change. In the CR sample significant two-way interactions with time were found for two risk factors. Post hoc analyses indicated that low partner Secure bond and low partner Self-disclosure were associated with more decline in Satisfaction.

During the full period of intervention plus follow-up, findings were again in line with hypothesis 2a across samples. In the SR sample four risk factors showed significant two-way interactions with time. Post hoc analyses indicated that high-risk was associated with less gain for Marital status (cohabitation), not having Children, low partner Commitment/stability, and high partner Dissatisfaction. In the CR sample three risk factors showed significant two-way interactions with time. Post hoc analyses indicated that low actor Education, high actor Anxiety, and low partner Self-disclosure were associated with small decline in Satisfaction, whereas the low-risk groups showed no or small improvement.

In sum, all findings were in support of hypothesis 2a: when room for improvement is controlled for, risk factors consistently predict negligible effects or less improvement (i.e., less gain during intervention or more decline of benefit during follow-up) but no positive effects during any time frame.

Vulnerability revisited: effects of risk factors on the course of satisfaction without controlling for pre-intervention dissatisfaction as moderator

We then conducted the same analyses without controlling for room for improvement (not including dissatisfaction as a moderator), which is in line with the modal analytic strategy in extant CRE research. As these analyses did *not* disentangle room for improvement and vulnerability effects, this resulted, as predicted, in very discrepant findings (Table S4). These analyses were not intended to reliably identify individual risk factors and we, therefore, do not discuss these findings in detail.

During intervention, we anticipated mixed results (more gain, less gain or negligible effects) due to the confounding of positive room for improvement effects and negative vulnerability effects (hypothesis 2b). Such an inconsistent pattern was exactly what we found in both samples. Less favorable scores on risk factors were significantly associated with: *more* gain ('↑' in Table S4) for three risk factors in the SR sample and two risk factors in the CR sample, with *less* gain ('↓' in Table S4) for one risk factor in the SR sample and one risk factor in the CR sample, and was nonsignificant for the other risk factors.

During follow-up, consistent with hypothesis 2c, we observed five risk factors in the SR sample and seven in the CR sample to be associated with more decline, presumably due to the summation of negative vulnerability and negative room for improvement effects. This was the case in both samples with only one exception in the CR sample.

During intervention plus follow-up, we observed, consistent with hypothesis 2d, mixed results for risk factors: more gain (two risk factors in the CR sample), less gain (two risk factors in the SR sample, and three in the CR sample), or negligible effects (three risk factors in the SR sample), due to the summation of positive and/or negative effects during intervention (2b) and follow-up (2c).

DISCUSSION

This study examined whether risk factors have positive or negative effects on gain from HmT. We proposed that two potentially confounding effects, room for improvement and vulnerability, may occur,

TABLE 1 Room for improvement (RFI) and vulnerability effects on the course of satisfaction controlled for RFI (pre-intervention dissatisfaction)

	Self-referred sample						Clinician-referred sample					
	F; p			Effect size			F; p			Effect size		
Room for improvement (RFI)	HmT	FU	HmT+FU	HmT	FU	HmT+FU	HmT	FU	HmT+FU	HmT	FU	HmT+FU
Hypothesis (≈/↓/↑)[1]	1a ↑	1b ↓	1c ≈/↓/↑	1a ↑	1b ↓	1c ≈/↓/↑	1a ↑	1b ↓	1c ≈/↓/↑	1a ↑	1b ↓	1c ≈/↓/↑
Actor dissatisfaction	**F = 23.701; p <0.001**	**F = 8.130; p <0.001**	**F = 11.827; p <0.001**	HDis = 0.87 ↑; LDis = 0.02	HDis = −0.07 ↓; LDis = 0.57	HDis = 0.79 ↑; LDis = 0.59	**F = 10.366; p <0.001**	**F = 5.856; p = 0.004**	**F = 7.646; p <0.001**	HDis = 0.71 ↑; LDis = −0.19	HDis = −0.29 ≈; LDis = −0.22	HDis = 0.43 ↑; LDis = −0.41
Vulnerability (controlled for RFI)	HmT	FU	HmT+FU	HmT	FU	HmT+FU	HmT	FU	HmT+FU	HmT	FU	HmT+FU
Hypothesis (≈/↓/↑)[2]	2a ≈/↓	2a ≈/↓	2a ≈/↓	2a ≈/↓	2a ≈/↓	2a ≈/↓	2a ≈/↓	2a ≈/↓	2a ≈/↓	2a ≈/↓	2a ≈/↓	2a ≈/↓
Relationship duration	ns	ns	ns				ns	ns	ns			
Actor education	ns	ns	ns				**F = 3.159; p = 0.005**	ns	**F = 2.278; p = 0.015**	HR = 0.06 ≈; LR = 0.13	HR = −0.15; LR = −0.03	HR = −0.09 ↓; LR = 0.10
Partner education	ns	ns	ns				ns	ns	ns			
Marital status	**F = 3.264; p = 0.022**	ns	**F = 2.604; p = 0.025**	HR = 0.41 ≈; LR = 0.40	HR = −0.27; LR = 0.21	HR = 0.13 ↓; LR = 0.62	ns	ns	ns			
Having children	**F = 4.304; p = 0.005**	ns	**F = 2.949; p = 0.013**	HR = 0.13 ↓; LR = 0.52	HR = 0.09; LR = 0.07	HR = 0.22 ↓; LR = 0.58	ns	ns	ns			
Actor commitment/stability	ns	ns	ns				ns	ns	ns			
Partner commitment/stability	**F = 3.448; p = 0.017**	ns	**F = 3.535; p = 0.004**	HR = 0.35 ↓; LR = 0.54	HR = −0.01; LR = 0.17	HR = 0.33 ↓; LR = 0.71	ns	ns	ns			
Actor anxiety	ns	ns	ns				**F = 3.444; p = 0.018**	ns	**F = 2.913; p = 0.014**	HR = 0.00 ↓; LR = 0.50	HR = −0.27; LR = −0.28	HR = −0.27 ↓; LR = 0.22
Partner anxiety	ns	ns	ns				ns	ns	ns			
Actor avoidance	ns	ns	ns				ns	ns	ns			
Partner avoidance	ns	ns	ns				ns	ns	ns			
Actor secure bond	ns	ns	ns				ns	ns	ns			
Partner secure bond	**F = 3.104; p = 0.027**	ns	ns	HR = 0.31 ↓; LR = 0.58	HR = −0.02; LR = 0.16	HR = 0.29; LR = 0.73	ns	**F = 4.338; p = 0.015**	ns	HR = 0.27; LR = 0.21	HR = −0.51 ↓; LR = −0.13	HR = −0.23; LR = 0.08

TABLE 1 (Continued)

Vulnerability (controlled for RFI)	HmT	FU	HmT+FU	HmT	FU	HmT+FU	HmT	FU	HmT+FU	HmT	FU	HmT+FU
Hypothesis (≈/↓/↑)²				2a ≈/↓	2a ≈/↓	2a ≈/↓				2a ≈/↓	2a ≈/↓	2a ≈/↓
Actor self-disclosure	ns	ns	ns				ns	ns	ns			
Partner self-disclosure	ns	ns	ns				ns	*F* = 6.941; *p* = 0.002	*F* = 3.018; *p* = 0.012	HR = 0.31, LR = 0.15	HR = −0.52 ↓, LR = −0.10	HR = −0.21 ↓, LR = 0.04
Actor Psych. compl.	ns	ns	ns				ns	ns	ns			
Partner Psych. compl.	ns	ns	ns				ns	ns	ns			
Partner dissatisfaction	*F* = 4.458; *p* = 0.004	ns	*F* = 2.732; *p* = 0.019	HR = 0.41 ≈, LR = 0.44	HR = 0.08, LR = 0.20	HR = 0.49 ↓, LR = 0.64	ns	ns	ns			
Total				2× and 3× ↓	—	4× ↓				1×≈ and 1× ↓	2× ↓	3× ↓

Abbreviations: HDis, High Dissatisfaction; HR, High risk; LDis, Low Dissatisfaction; LR, Low risk; Psych. compl., Psychological complaints.

¹≈ means effect size HDis ≈ LDis (Δd < 0.1); ↓ means effect size HDis < LDis (Δd ≥ 0.1); ↑ means effect size HDis > LDis (Δd ≥ 0.1).

²≈ means effect size HR ≈ LR (Δd < 0.1); ↓ means effect size HR < LR (Δd ≥ 0.1); ↑ means effect size HR > LR (Δd ≥ 0.1).

with the ultimate results depending on sample, time frame of prediction, statistical handling, and specific predictors. Both our literature review and the present findings support a large positive room for improvement effect during intervention, such that higher pre-intervention relationship dissatisfaction predicts more improvement in relationship satisfaction. We expected this effect to decline at follow-up, which was indeed found for the SR sample, but not fully replicated in the CR sample in which the effect was negligibly small. Furthermore, and in line with the few prior studies that controlled for room for improvement, our present findings supported the vulnerability hypothesis across samples: when controlling for room for improvement all risk factors showed negative or negligible effects, but in no instance positive effects. We organized the discussion of our findings with an eye to the replication crisis that psychology is currently coping with (e.g., Tackett et al., 2017). Hence, for each of our major findings, we will discuss how likely we believe replication is, as well as under which conditions (e.g., sample and intervention-type).

Probably replicable: room for improvement hypothesis

Support for the room for improvement hypothesis was quite strong and robust over the two samples of our empirical study, and we, therefore, believe that these results are likely to be replicated in future studies. Specifically, high pre-intervention relationship dissatisfaction was consistently associated with larger improvement in satisfaction during intervention in both SR and CR samples (hypothesis 1a). Of note, this finding is also consistent with the review of CRE studies provided in the introduction (Table S2). In the SR sample, the room for improvement hypothesis was also supported during follow-up: high pre-intervention dissatisfaction was associated with a decline of satisfaction (hypothesis 1b). For the CR sample, however, this decline was negligibly small as post hoc analyses revealed that both high and low pre-intervention dissatisfied groups showed similar decline.

To explain this particular finding, it is important to consider sampling first. Individuals from the CR sample first received individual intervention for anxiety, mood, or personality disorders, and were in remission upon entering HmT. However, the mean level of attachment anxiety in this sample was found to be particularly high. It may well be that the degree of vulnerability of the CR sample was not well-matched with the intervention intensity, as HmT is a rather 'light' intervention. As a result, we conjecture that HmT may have sensitized underlying attachment-related vulnerabilities. In line with this, CR couples showed their peak gain (the point of highest relationship satisfaction), on average prior to session 5 (Figure S1). In this session couples start practicing the crucial secure attachment strategies. The relapse after session 5 may be interpreted as couples not being able to structurally change their core attachment problems. Of course, these are speculations in need of further research.

In sum, we think that hypothesis 1 (1a and 1b) is most likely to replicate in CRE studies in low distressed samples with matching intervention intensities. Our review of the empirical literature further underscores this hypothesis: the comprising study samples predominantly consisted of nondistressed couples, and room for improvement effects for CREs were consistently observed. We strongly recommend testing for room for improvement in other samples and intervention modalities as well, but are less confident with respect to replication under those conditions.

Probably replicable: vulnerability hypothesis

Consistent with hypothesis 2a, all risk factors that showed incremental predictive value after controlling for room for improvement effects were associated with either negligible or negative effects (less gain during intervention or more decline during follow-up). Importantly, across two independent samples with a large set of predictors no exceptions (i.e., positive effects) were found. We think this finding is probably replicable, although specific outcome may depend on the balance of distress and vulnerability within the sample and the intervention intensity. After controlling for room for improvement, significant risk factors are indicative of 'pure' vulnerability effects which constitute obstacles to inter-

vention or vulnerability to relapse. Our findings were fully in line with the two previous studies that controlled for dissatisfaction as moderator (Table S2). Interestingly, landmark response prediction studies concerning couple *therapy* have also controlled for room for improvement and almost consistently found risk factors to have negligible or negative effects (cf., Atkins et al., 2005; Baucom et al., 2009).

By contrast, analyses not controlling for pre-intervention dissatisfaction yielded divergent effects, namely: (1) positive effects of risk factors during intervention that are likely to result from positive room for improvement effects outweighing negative vulnerability effects, and (2) more risk factors that become significant predictors of decline during follow-up which is likely the result from the additional negative effect of room for improvement during follow-up on top of the already negative vulnerability effects. Most of these effects disappeared after controlling for dissatisfaction as moderator. We conjecture the same would happen if previous studies were to be re-analyzed this way: only negligible and negative effects related to vulnerability remain.

Probably not replicable: sample-specific risk factors

No specific indicators of vulnerability were identified that replicated across samples. In fact, most of the risk factors (72% in the SR and 78% in the CR sample) showed nonsignificant associations to outcome. These results are consistent with the notable absence of replication we found in the response prediction literature (Table S2). The only partial replication was found for low partner Secure bond, which was a predictor of less gain during intervention in the SR sample and stronger decline of satisfaction during follow-up in the CR sample.

In sum, risk factors did not replicate between the two samples we used, almost all significant findings in the extant literature failed to replicate, and the majority of selected risk factors was not significantly associated with outcome in the first place. These observations lead to a somewhat pessimistic conclusion: apart from relationship dissatisfaction itself, no specific single measures can be recommended for use in CRE prediction studies with satisfaction as outcome.

Maybe replicable: partner effects and commitment

As noted, the great majority of studies we reviewed did not take into account room for improvement, nor distinguished between intervention and follow-up. If our basic premise is correct, doing so would yield improved odds for finding meaningful replicable predictors. We believe that two (post hoc) tentative hypotheses can be derived for future research from previous prediction studies and our current findings. First, after controlling for pre-intervention dissatisfaction, the rarely examined partner variables became important indicators of vulnerability in both our samples, whereas actor variables mostly were nonsignificant. We anticipate replication of partner variables in general if pre-intervention outcome is adequately being controlled for. Second, in the SR sample several indicators of lower relationship establishment, that is, not being married, not having children, and low levels of partner commitment/stability, were associated with less gain. Interestingly, in this less vulnerable sample than the CR sample these factors represent relationship-related characteristics. In the CR sample a different set of risk factors, namely actor anxiety, partner secure bond, and partner self-disclosure featured as important vulnerabilities. These variables represent the main intervention targets of HmT, that is, making partners less anxious about rejection and more disclosing, to eventually repair and enhance the secure bond between them. CR couples scored significantly lower than the SR sample on these variables pre-intervention. We speculate that in this more vulnerable sample these factors may form an obstacle for a light intervention like HmT. Yet, replication over clinical samples often fails, probably as a result of the particular characteristics of each 'convenience sample'.

Recommendations for future research

In addition to formulating general rather than predictor-specific hypotheses, we would like to advocate some guidelines for study designs that may serve to improve the likelihood of achieving replicable results. Probably the most crucial guideline is to include multiple samples within one study as this provides much more opportunity to distinguish sample-dependent results from replicable results. To illustrate, we observed that no risk factors replicated across the SR and CR samples. As all variables in the current study were measured with identical instruments at identical time points in both samples, measurement variance can be ruled out as an alternative explanation for the inconsistencies. In our view, this makes sampling to be the most likely cause of the observed discrepancies. This implies that replications are most likely across comparable samples and interventions. Therefore, we recommend researchers to specify to which type of samples and interventions their hypotheses likely apply, and, when feasible, to use multiple samples in one study to test for replicability. Of course, failures of replication may inspire useful hypotheses about mechanisms (the "why" in one sample, and "why not" in the other).

Second, we advocate separate analyses for change during the intervention and follow-up time frames. As effects may well be diverge across time intervals (positive room for improvement effects may be more prominent during intervention and negative vulnerability effects more during follow-up), the aggregation of these periods increases the probability that they are positive in one study, negative in another, and nonsignificant in a third study.

Third, we strongly recommend to control for room for improvement effects in response prediction analyses. This guideline was indicated by our review of the literature and the markedly different findings we obtained by analyzing with and without controlling for pre-intervention dissatisfaction. Since we hypothesize that during intervention negative vulnerability effects will often be contaminated with, or masked by positive room for improvement effects, we believe vulnerability effects will become more prominent after controlling for room for improvement. Of note, this does not mean we completely rule out the possibility that risk factors may act as a positive predictor of change. This may be the case when the risk factor is modifiable by the intervention (e.g., attachment anxiety in EFCT; Dalgleish et al., 2015).

Fourth, we recommend univariate approaches instead of multivariate analyses. As shown in previous studies (e.g., Amato, 2014; Petch et al., 2012), multivariate approaches tend to include a study-specific set of predictors, which results in analyses that are not replicated in other studies.

Fifth, we advocate inclusion of a broad set of predictors rather than a priori selecting only one or a few risk factors. Accordingly, the likelihood of testing similar predictors and hence replicate effects across studies strongly increases. Furthermore, we propose the inclusion of partner effects as there are important clinical reasons to suspect these are relevant (Conradi et al., 2021).

Finally, the substantial room for improvement effects point to a crucial question for future research. Some aspects of room for improvement, regression to the mean, re-engagement, or new hope, may to some extent occur *without* intervention. Therefore, the estimated effects of room for improvement may mix-up actual change caused by intervention and change that would occur in a control condition as well. Interestingly, Doss et al. (2020) reported significant positive room for improvement effects during intervention but also for a waiting list group which did not significantly differ in magnitude. This may suggest that room for improvement captures a phenomenon *unrelated* to specific interventions. However, replication is needed across different samples and interventions.

Several limitations should be kept in mind when interpretating the present findings. First, although we included a follow-up of 3.5 months it remains unclear whether maintenance of intervention gains in SR couples will prove stable over longer periods of time. Second, as discussed above, we suspect findings regarding CR couples to be rather sample specific as is often the case with clinical convenience samples. Findings regarding SR couples, therefore, seem more representative for regular users of HmT and replication of these findings appear to be the most fruitful direction for future studies.

Finally, our analyses pertain only to satisfaction as outcome variable. Studying other variables that are of theoretical interest, such as attachment, communication, or affect, is recommended.

CONCLUSION AND CLINICAL IMPLICATIONS

In conclusion, in line with the room for improvement hypothesis, the current study established pre-intervention dissatisfaction as a predictor of gain in satisfaction during intervention (both samples) and decline during follow-up (SR sample, for the CR sample the effect was negligible). Moreover, room for improvement effects were shown to be confounded with negative vulnerability effects, implying that not disentangling both may result in erroneous interpretations and conclusions.

As the primary problem for the vulnerable couples appears to be decline of benefits the present findings bear clear clinical implications for the issue of selection for intervention and its effects on outcome (Kamphuis et al., 2021). First, CRE program trainers might want to assess the risk for decline prior to intervention. However, the current results and extant literature do not provide reason for optimism concerning consistency of risk factors across samples. In fact, monitoring during intervention and follow-up might be a better diagnostic strategy for indicating couples who need more intensive interventions. Second, while the initial steep gain in relationship satisfaction in high-risk couples following a CRE program may suggest to trainers that such programs work rather well in high-risk couples (and even better than in low-risk couples), we recommend caution. Making too much of short-lived positive room for improvement effects may end in unnecessary disappointment as benefit may be followed by decline due to persistent vulnerability. To use a metaphor: use of low dose interventions in severe couple distress may be like pumping new air into a flat tire without adequately fixing the puncture. Hence, it seems more efficient to match the intensity of the intervention to the degree of vulnerability observed in the couple. For some, CRE programs may suffice, whereas others need an enhanced CRE program or couple therapy. Although more research is needed, this conclusion fits the original aim of regular CRE programs, which was prevention of relationship problems from becoming chronic and harder to treat (Bradford et al., 2015).

ORCID

Henk Jan Conradi https://orcid.org/0000-0002-8013-0408

REFERENCES

References included in the literature review are marked with an '*' when reporting response prediction analyses and with a '+' to indicate studies reporting separate room for improvement analyses.

*Allen, E. S., Rhoades, G. K., Stanley, S. M., Loew, B., & Markman, H. J. (2012). The effects of marriage education for army couples with a history of infidelity. *Journal of Family Psychology*, *26*(1), 26–35. https://doi.org/10.1037/a0026742

*Amato, P. (2014). Does social and economic disadvantage moderate the effects of relationship education on unwed couples? An analysis of data from the 15-month building strong families evaluation. *Family Relations*, *63*, 343–355. https://doi.org/10.1111/fare.12069

Arnold, L. S., & Beelmann, A. (2019). The effects of relationship education in low-income couples: A meta-analysis of randomized-controlled evaluation studies. *Family Relations*, *68*(1), 22–38. https://doi.org/10.1111/fare.12325

Atkins, D. C., Berns, S. B., George, W. H., Doss, B. D., Gattis, K., & Christensen, A. (2005). Prediction of response to treatment in a randomized clinical trial of marital therapy. *Journal of Consulting and Clinical Psychology*, *73*(5), 893–903. https://doi.org/10.1037/0022-006X.73.5.893

*Barton, A. W., Beach, S. R. H., Lavner, J. A., Bryant, C. M., Kogan, S. M., & Brody, G. H. (2017). Is communication a mechanism of relationship education effects among rural African Americans? *Journal of Marriage and Family*, *79*(5), 1450–1461. https://doi.org/10.1111/jomf.12416

*Barton, A. W., Futris, T. G., & Bradley, R. C. (2014). Changes following premarital education for couples with differing degrees of future marital risk. *Journal of Marital and Family Therapy*, *40*(2), 165–177. https://doi.org/10.1111/jmft.12006

Baucom, B. R., Atkins, D. C., Simpson, L. E., & Christensen, A. (2009). Prediction of response to treatment in a randomized clinical trial of couple therapy: A 2-year follow-up. *Journal of Consulting and Clinical Psychology*, *77*(1), 160–173. https://doi.org/10.1037/a0014405

+Bodenmann, G., Hilpert, P., Nussbeck, F. W., & Bradbury, T. N. (2014). Enhancement of couples' communication and dyadic coping by a self-directed approach: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *82*(4), 580–591. https://doi.org/10.1037/a0036356

+Bradford, A. B., Drean, L., Adler-Baeder, F., Ketring, S. A., & Smith, T. A. (2017). It's about time! Examining received dosage and program duration as predictors of change among non-distressed and distressed married couple and relationship education participants. *Journal of Marital and Family Therapy*, *43*(3), 391–409. https://doi.org/10.1111/jmft.12209

Bradford, A. B., Hawkins, A. J., & Acker, J. (2015). If we build it, they will come: Exploring policy and practice implications of public support for couple and relationship education for lower income and relationally distressed couples. *Family Process*, *54*(4), 639–654. https://doi.org/10.1111/famp.12151

Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement of adult romantic attachment: An integrative overview. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 46–76). Guilford Press.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*(1), 147–158. https://doi.org/10.1037/0033-2909.101.1.147

Buunk, B. P. (1990). Relational interaction satisfaction scale (RISS). In J. Touliatos, B. F. Perlmutter, & M. A. Straus (Eds.), *Handbook of family measurement techniques* (p. 106). Saga Publications.

+*Carlson, R. G., Rappleyea, D. L., Daire, A. P., Harris, S. M., & Liu, X. (2017). The effectiveness of couple and individual relationship education: Distress as a moderator. *Family Process*, *56*(1), 91–104. https://doi.org/10.1111/famp.12172

+Casey, P., Cowan, P. A., Cowan, C. P., Draper, L., Mwamba, N., & Hewison, D. (2017). Parents as partners: A U.K. trial of a U.S. couples-based parenting intervention for at-risk low-income families. *Family Process*, *56*, 589–606. https://doi.org/10.1111/famp.12289

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). Academic Press.

Conradi, H. J., De Jonge, P., Neeleman, A., Simons, P., & Sytema, S. (2011). Partner attachment as a predictor of long-term response to treatment with couples therapy. *Journal of Sex & Marital Therapy*, *37*(4), 286–297. https://doi.org/10.1080/0092623X.2011.582435

Conradi, H. J., Dingemanse, P., Noordhof, A., Finkenauer, C., & Kamphuis, J. H. (2018). Effectiveness of the 'hold me tight' relationship enhancement program in a self-referred and a clinician-referred sample: An emotionally focused couples therapy-based approach. *Family Process*, *57*(3), 613–628. https://doi.org/10.1111/famp.12305

Conradi, H. J., Gerlsma, C., van Duijn, M., & de Jonge, P. (2006). Internal and external validity of the experiences in close relationships questionnaire in an American and two Dutch samples. *European Journal of Psychiatry*, *20*, 258–269.

Conradi, H. J., Kamphuis, J. H., & de Jonge, P. (2018). Adult attachment predicts the seven-year course of recurrent depression in primary care. *Journal of Affective Disorders*, *225*, 160–166. https://doi.org/10.1016/j.jad.2017.08.009

Conradi, H. J., Noordhof, A., & Kamphuis, J. H. (2021). Satisfying and stable couple relationships: Attachment similarity across partners can partially buffer the negative effects of attachment insecurity. *Journal of Marital and Family Therapy*, *47*, 682–697. https://doi.org/10.1111/jmft.12477

Cook, W. L., & Kenny, D. A. (2005). The actor-partner interdependence model: A model of bidirectional effects in developmental studies. *International Journal of Behavioral Development*, *29*(2), 101–109. https://doi.org/10.1080/01650250444000405

*Cowan, P. A., Cowan, C. P., Pruett, M. K., Pruett, K., & Gillette, P. (2014). Evaluating a couples group to enhance father involvement in low-income families using a benchmark comparison. *Family Relations*, *63*(3), 356–370. https://doi.org/10.1111/fare.12072

Dalgleish, T. L., Johnson, S. M., Burgess Moser, M., Wiebe, S. A., & Tasca, G. A. (2015). Predicting key change events in emotionally focused couple therapy. *Journal of Marital and Family Therapy*, *41*(3), 260–275. https://doi.org/10.1111/jmft.12101

+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. *Journal of Consulting and Clinical Psychology*, *88*, 283–294. https://doi.org/10.1037/ccp0000479

*Dupree, D. G., Whiting, T. B., & Harris, S. M. (2016). A person-oriented analysis of couple and relationship education. *Family Relations*, *65*, 635–646. https://doi.org/10.1111/fare.12222

Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, *14*(1), 43–53. https://doi.org/10.1037/a0014699

+*Gordon, K. C., Cordova, J. V., Roberson, P. N. E., Miller, M., Gray, T., Lenger, K. A., Hawrilenko, M., & Martin, K. (2019). An implementation study of relationship checkups as home visitations for low-income at-risk couples. *Family Process*, *58*(1), 247–265. https://doi.org/10.1111/famp.12396

*Gubits, D., Lowenstein, A. E., Harris, J., & Hsueh, J. (2014). Do the effects of a relationship education program vary for different types of couples? Exploratory subgroup analysis in the supporting healthy marriage evaluation (OPRE report 2014–22). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Halford, W. K., & Bodenmann, G. (2013). Effects of relationship education on maintenance of couple relationship satisfaction. *Clinical Psychology Review*, *33*(4), 512–525. https://doi.org/10.1016/j.cpr.2013.02.001

+Halford, W. K., Pepping, C. A., Hilpert, P., Bodenmann, G., Wilson, K. L., Busby, D., Larson, J., & Holman, T. (2015). Immediate effect of couple relationship education on low-satisfaction couples: A randomized clinical trial plus an uncontrolled trial replication. *Behavior Therapy*, *46*(3), 409–421. https://doi.org/10.1016/j.beth.2015.02.001

*Halford, W. K., Sanders, M. R., & Behrens, B. C. (2001). Can skills training prevent relationship problems in at-risk couples? Four-year effects of a behavioral relationship education program. *Journal of Family Psychology*, 15(4), 750–768. https://doi.org/10.1037/0893-3200.15.4.750

*Halford, W. K., & Wilson, K. L. (2009). Predictors of relationship satisfaction four years after completing flexible delivery couple relationship education. *Journal of Couple & Relationship Therapy*, 8(2), 143–161. https://doi.org/10.1080/15332690902813828

+*Hsueh, J., Alderson, D. P., Lundquist, E., Michalopoulos, C., Gubits, D., Fein, D., & Knox, V. (2012). The supporting healthy marriage evaluation: Early impacts on low-income families (OPRE report 2012–11). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

+*Job, A. K., Baucom, D. H., & Hahlweg, K. (2017). Who benefits from couple relationship education? Findings from the largest German CRE study. *Journal of Couple & Relationship Therapy*, 16(2), 79–101. https://doi.org/10.1080/15332691.2016.1178616

Johnson, S. (2008). *Hold me tight: Seven conversations for a lifetime of love*. Little & Brown.

Kamphuis, J. H., Noordhof, A., & Hopwood, C. J. (2021). When and how assessment matters: An update on the treatment utility of clinical assessment (TUCA). *Psychological Assessment*, 33(2), 122–132. https://doi.org/10.1037/pas0000966

+*Kennedy, N. (2017). Conversations for connection: An outcome assessment of the hold me tight relationship education program for couples (Doctoral dissertation). University of Ottawa.

Koeter, M. W. J., & Ormel, J. (1991). *In Dutch: General health questionnaire. Dutch adaptation, manual*. Swets & Zeitlinger.

Levenson, R. W., Haase, C. M., Bloch, L., Holley, S. R., & Seider, B. J. (2015). Emotion regulation in couples. In J. J. Gross (Ed.), *Handbook of emotion regulation* (2nd ed.). Guilford Press.

+*Lundquist, E., Hsueh, J., Lowenstein, A. E., Faucetta, K., Gubits, D., Michalopoulos, Ch., & Knox, V. (2014). A family-strengthening program for low-income families: Final impacts from the supporting healthy marriage evaluation (OPRE report 2014-09A). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 1(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

Markman, H. J., & Rhoades, G. K. (2012). Relationship education research: Current status and future directions. *Journal of Marital and Family Therapy*, 38, 169–200. https://doi.org/10.1111/j.1752-0606.2011.00247.x

Miller, L. C., Berg, J. H., & Archer, R. L. (1983). Openers: Individuals who elicit intimate self-disclosure. *Journal of Personality and Social Psychology*, 44(6), 1234–1244. https://doi.org/10.1037/0022-3514.44.6.1234

*Mitchell, A. M., Owen, J., Adelson, J. L., France, T., Inch, L. J., Bergen, C., & Lindel, A. (2015). The influence of dyadic coping in relationship education for low- income racial and ethnic minority couples. *Journal of Family Therapy*, 37(4), 492–508. https://doi.org/10.1111/1467-6427.12057

*Petch, J., Halford, W. K., Creedy, D. K., & Gamble, J. (2012). Couple relationship education at the transition to parenthood: A window of opportunity to reach high-risk couples. *Family Process*, 51(4), 498–511. https://doi.org/10.1111/j.1545-5300.2012.01420.x

*Roberson, P. N. E., Lenger, K. A., Gray, T., Cordova, J., & Gordon, K. C. (2020). Dyadic latent profile analyses and multilevel modeling to examine differential response to couple relationship education. *Journal of Family Psychology*, 34(7), 879–885. https://doi.org/10.1037/fam0000667

Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and other dyads. *Journal of Marriage and the Family*, 38(1), 15–28. https://doi.org/10.2307/350547

Stanley, S. M., & Markman, H. J. (1992). Assessing commitment in personal relationships. *Journal of Marriage and the Family*, 54(3), 595–608. https://doi.org/10.2307/353245

*Stanley, S. M., Rhoades, G. K., Loew, B. A., Allen, E. S., Carter, S., Osborne, L. J., Prentice, D., & Markman, H. J. (2014). A randomized controlled trial of relationship education in the U. S, Army: 2-year outcomes. *Family Relations*, 63, 482–495. https://doi.org/10.1111/fare.12083

Stanley, S. M., Rhoades, G. K., & Markman, H. J. (2006). Sliding versus deciding: Inertia and the premarital cohabitation effect. *Family Relations*, 55(4), 499–509. https://doi.org/10.1111/j.1741-3729.2006.00418.x

Stanley, S. M., Rhoades, G. K., & Whitton, S. W. (2010). Commitment: Functions, formation, and the securing of romantic attachment. *Journal of Family Theory & Review*, 2(4), 243–257. https://doi.org/10.1111/j.1756-2589.2010.00060.x

Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., Oltmanns, T. F., & Shrout, P. E. (2017). It's time to broaden the replicability conversation: Thoughts for and from clinical psychological science. *Perspectives on Psychological Science*, 12(5), 742–756. https://doi.org/10.1177/1745691617690042

*Van Widenfelt, B., Hosman, C., Schaap, C., & van der Staak, C. (1996). The prevention of relationship distress for couples at risk: A controlled evaluation with nine-month and two-year follow-ups. *Family Relations*, 45(2), 156–165. https://doi.org/10.2307/585286

+*Williamson, H. C., Rogge, R. D., Cobb, R. J., Johnson, M. D., Lawrence, E., & Bradbury, T. N. (2015). Risk moderates the outcome of relationship education: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 83, 617–629. https://doi.org/10.1037/a0038621

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Conradi, H. J, Noordhof, A., Dingemanse, P., & Kamphuis, J. H. (2023). Do high-risk couples profit more or less from couple relationship education programs than low-risk couples? Room for improvement and vulnerability effects. *Family Process*, *62*, 591–608. https://doi.org/10.1111/famp.12801