Journal of Evidence-Based Psychotherapies,
Vol. 23, Special Issue, March 2023, 67-94.

10.24193/jebp.2023.1.4

EFFICACY AND ACCEPTABILITY OF VIRTUAL REALITY IMAGERY RESCRIPTING FOR PTSD DUE TO CHILDHOOD SEXUAL ABUSE: A MULTIPLE BASELINE STUDY

Sven van Kuik[1], Arnold A.P. van Emmerik[2], Willem-Paul Brinkman[3], Elizabeth Uduwa-Vidanalage[2], Cliff Schouten[2], Arnoud Arntz[2]*

[1] PsyQ, Purmerend, The Netherlands
[2] University of Amsterdam, Amsterdam, The Netherlands
[3] Delft University of Technology, The Netherlands

Abstract

This study piloted the efficacy and acceptability of Virtual Reality Imagery Rescripting (VR-ImRs) compared to conventional Imagery Rescripting (ImRs) for PTSD due to childhood sexual abuse (CSA). Eight adult patients with clinician-rated PTSD due to CSA as their primary diagnosis participated, of whom six completed the full treatment. A non-concurrent multiple baseline design with cross-over elements was used, with randomly assigned baseline lengths and treatment conditions. After baseline and a 5-session 'education and exploration' phase, six sessions of either ImRs or VR-ImRs were given, followed by another six sessions of the opposite treatment condition and a 5-week follow-up without treatment. The primary outcome was PTSD symptoms (PCL-5), and secondary outcomes were negative and positive emotions (added PCL-5 items), anxiety and depressive symptoms (HADS) and trauma-related cognitions (PTCI). Data were analyzed with mixed regression. Results showed a significant linear reduction of trauma symptoms and negative emotions only during ImRs. No significant treatment effects on positive emotion, anxiety and depressive symptoms were found for both treatment conditions. Both treatment conditions showed significant positive effects on trauma-related cognitions. This study does not support the efficacy of VR-ImRs in reducing PTSD symptoms. Possibly VR-ImRs keeps people from reprocessing their memories, making it less effective.

Keywords: Posttraumatic stress disorder; Childhood sexual abuse; Virtual reality; Imagery rescripting; Cognitive behavioral therapy; Clinical trial.

[*] Correspondence concerning this article should be addressed to Arnold A. P. van Emmerik, Nieuwe Achtergracht 129-B, 1018 WS Amsterdam, The Netherlands.
E-mail: A.A.P.vanEmmerik@uva.nl

Posttraumatic stress disorder (PTSD) is a debilitating condition causing severe distress to individuals (American Psychiatric Association, 2013). Being exposed to actual or threatened death, serious injury, or sexual violence may lead to PTSD. Patients suffering from PTSD continue re-experiencing the traumatic event(s) in various ways, such as unwanted upsetting memories, nightmares or flashbacks, and in addition display symptoms of avoidance, negative changes in cognitions or mood, and increased arousal or reactivity. Fortunately, there are various evidence-based treatments available such as exposure-based therapies, Eye Movement Desensitization and Reprocessing (EMDR), and conventional imagery rescripting (ImRs). ImRs is a treatment technique that has been used in PTSD since the late twentieth century (see for example Smucker et al., 1995). Recently, the interest in this treatment has increased rapidly due to the growing body of evidence suggesting that ImRs is an effective psychological intervention for PTSD (for a review, see Arntz, 2012; Morina et al., 2017). With this growing evidence, the possibilities for its application also broadened. For example, Jung and Steil (2013) showed that ImRs reduced feelings of being contaminated in patients with PTSD due to childhood sexual abuse, and Kunze et al. (2017) showed that ImRs helped in reducing nightmares. A recent study showed that ImRs is as effective as EMDR in reducing PTSD symptoms arising from childhood trauma, with approximately 60 percent of these childhood traumas being sexual abuse (Boterhoven de Haan et al., 2020). In brief, in ImRs the therapist instructs patients to vividly imagine the traumatic memory, followed by an alternative script which is either made up by patient or therapist, depending on the phase of the treatment. The alternative script usually starts at the most difficult part of the memory, the so-called the hotspot, and ends in a scenario that better meets the patient's emotional needs. For an extensive explanation of the background, rationale, and treatment protocol see Arntz (2012, 2015) and Arntz and Weertman (1999).

However, as with all therapeutic interventions, not all patients benefit (sufficiently) from ImRs (Boterhoven de Haan et al., 2020). First, some patients find it difficult to generate an alternative trauma scenario (Arntz & van Genderen, 2009). For example, avoidance, dysfunctional beliefs, or certain personality characteristics may hinder the formation of such alternative trauma scenarios and thus the effectiveness of ImRs. Other hindrances may include patients not being able to identify and activate a relevant memory, the intervention being perceived as unrealistic, patients not being able or willing to take the child perspective (Arntz & Weertman, 1999; Arntz & van Genderen, 2009), or patients not wanting to close their eyes during the process (Arntz, 2015). Second, as in cognitive therapy, in some patients functional and realistic cognitions and helpful alternative trauma scenarios are clearly there, but lack conviction ("I know it, but it doesn't feel that way"), thus making it difficult for them to engage emotionally (Stott, 2007; Cresswell & Waite, 2009).

For patients experiencing these difficulties, virtual reality imagery rescripting (VR-ImRs) may be an experientially powerful alternative technique to rescript traumatic memories and the meaning these memories have for them. Virtual reality (VR) could help these patients to visualize the traumatic event and engage with it emotionally (Botella et al., 2015; Maples-Keller et al., 2017). Furthermore, VR may allow more control over the experience since the therapist controls depicting the images (Maples-Keller et al., 2017). There is growing research suggesting that VR-based therapies can be useful in treatments for a large variety of mental disorders (Low et al., 2020; Maples-Keller et al., 2017; Morina et al., 2015; Opriş et al., 2012; Gerardi et al., 2010; Beck et al., 2007). There is also growing consensus that VR-based therapies can be helpful in treating PTSD, although the evidence is largely limited to war veterans (Van Meggelen, 2022; for a review, see Botella et al., 2015; but see Reger et al., 2016 for findings that imaginal exposure was superior to VR-exposure at 12- and 26-week follow-up).

Given the increasing evidence for ImRs (i.e., non-VR), identifying methods that could further strengthen its effects for specific patient groups is an important next step. Therefore, this exploratory proof-of-concept study was conducted as a first test of VR-ImRs in eight patients with PTSD due to childhood sexual abuse. We used a non-concurrent multiple baseline cross-over design testing effectiveness of active treatment against time and nonspecific attention, with a comparison of VR-ImRs and ImRs, with the aim of obtaining preliminary quantitative and qualitative data on the efficacy and acceptability of VR-ImRs. In addition, we also evaluated the feasibility of VR-ImRs in routine care.

Method

Participants

Participants were eligible if they met the criteria of PTSD due to childhood sexual abuse (as index trauma) that took place before the age of 16. Additional inclusion criteria included: (1) sufficient fluency in Dutch to complete research procedures; (2) able to attend sessions twice a week during the treatment period; (3) agree to abstain from medication changes or other psychological therapies for the duration of the study (i.e., until five weeks post-treatment). Exclusion criteria included: (1) PTSD due to trauma that occurred within the past six months; (2) acute suicide risk; (3) comorbid psychotic disorder; (4) comorbid bipolar disorder type 1; (5) comorbid alcohol or drug dependence; (6) IQ < 80; (7) neurological problems (e.g., dementia); (8) medication changes or PTSD-focused therapy within the past three months; (9) benzodiazepine medication (unless abstinent for two weeks). Potential participants were fully informed, both verbally and in writing, about the study before agreeing to participate and providing written informed consent. Of the 15 patients that were screened for eligibility, eight participants were included and

randomly allocated to the various baseline lengths and intervention orders (see Figure 1 for the complete patient flow). Table 1 describes the characteristics of the participants. The study protocol was approved by the ethical committee of the Faculty of Social and Behavioural Sciences and pre-registered in the Netherlands Trial Register (ID: NTR7029).

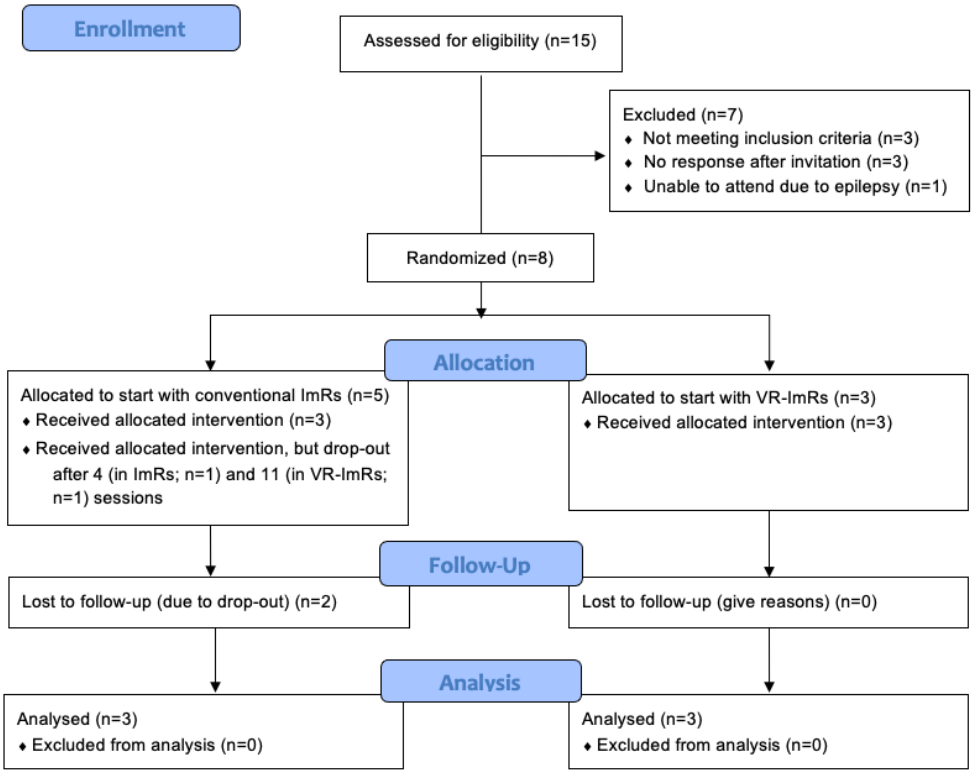


Figure 1.

Table 1. Demographic and Clinical Characteristics of Participants (N = 8).

Variable		Mean (SD)/Number (%)
Age (in years)		38.0 (12.1)
Gender	Female	5 (63%)
	Male	3 (38%)
Relationship status	With partner	7 (88%)
	No partner	1 (13%)
Living situation	Alone	3 (38%)
	With partner	4 (50%)
	With parents or other family members	1 (13%)
Living with children		5 (63%)
Education (highest)	Elementary school/primary education	3 (38%)
	Lower secondary education	1 (13%)

Variable		Mean (SD)/Number (%)
	Higher secondary education	1 (13%)
	Post-secondary vocational education	1 (13%)
	University for applied sciences	2 (25%)
PTSD duration (months)		191.6 (165.2)
Current use of medication	Anti-depressant	3 (38%)
Previous treatment	Individual therapy	8 (100%)
	Group therapy	3 (38%)
	Partner relationship/family therapy	1 (13%)
	Day treatment	1 (13%)
Comorbid disorder	Mood disorders	4 (50%)
	Anxiety disorders	4 (50%)
	Bulimia Nervosa	1 (13%)
	ADHD	1 [a] (13%)
	Elder-child relationship problem	1 [a] (13%)
	Circadian rhythm-sleep wake disorder; all subtypes	1 [a] (13%)
Number of comorbid disorders		2.0 (1.6)
Alcohol or substance use		5 (62.5%)

Note. [a] Disorders diagnosed by trained clinicians.

Study design

A non-concurrent multiple baseline case series design with cross-over elements was used, with a baseline varying in length between five, six, or seven weeks to differentiate between time effects and experimental effects (Arntz et al., 2013). The baseline phase started with a pretest and then involved twice-weekly process measurements of patients' PTSD symptoms that continued throughout the study until the end of the follow-up phase. The baseline phase was followed by a 5-session education and exploration phase consisting of twice-weekly sessions that served to (1) build a therapeutic relation, (2) provide education about the upcoming treatment, and (3) obtain the information about the original traumatic scenarios, and the information that was needed to devise an alternative, helpful rescripting scenario, since the VR-ImRs worlds needed to be prepared in advance. The education and exploration phase also served (4) as a control phase for non-specific treatment factors such as attention, understanding, and focus on trauma, and (5) to familiarize patients with the VR-equipment in a non-emotional way. After the education and exploration phase, participants were randomly allocated to one of two treatment orders: (1) six treatment sessions of VR-ImRs followed by six treatment sessions of ImRs, both given twice a week, or (2) six treatment sessions of ImRs followed by six treatment sessions of VR-ImRs, both given twice a week. Finally, both 6-week treatment phases were followed by a 5-week no intervention follow-up phase to examine the treatment's long-term effects, after which regular treatment was offered if needed (see Figure 2 for a schematic presentation of the study design, including assessment moments).

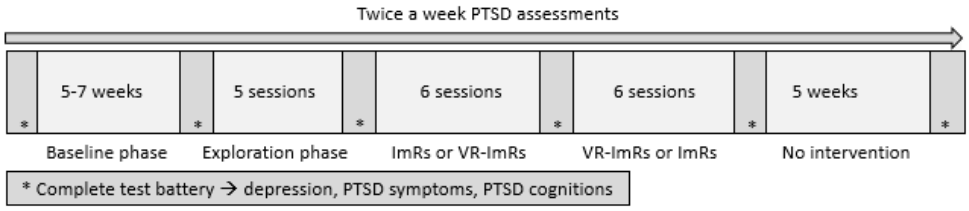


Figure 2.

Primary outcome

Self-reported PTSD symptoms were assessed with the PTSD Checklist for DSM-5 (PCL-5; Weathers et al., 2013b). This is one of the most commonly used measures of PTSD in both clinical and research settings and has excellent psychometric properties (e.g., internal consistency ($\alpha = .94$), test-retest reliability ($r = .82$), and convergent ($r = .74$ to $.85$) and discriminant ($r = .31$ to $.60$) validity; Blevins et al., 2015). The time frame in the instruction of the PCL-5 was changed from "in the past month" to "since the previous assessment", and within the treatment phase to "since the last session", since participants were asked to complete the PCL-5 twice a week, making the total number of PCL-5 assessments 37, 39 or 41 depending on baseline phase length. PCL-5 items are rated 0 ('Not at all') to 4 ('Extremely') resulting in a total score that ranges between 0–80, with higher scores reflecting greater symptom severity.

Secondary outcomes

Seven items were added to the PCL-5 to assess shame, anger, guilt, disgust, sadness, anxiety, and happiness (Boterhoven de Haan et al., 2020). These items were not included in the PCL-5 total score, but were analyzed separately as two factors (negative emotions and positive emotions).

Changes in severity of PTSD symptoms were also assessed by trained independent research assistants, that were blind for treatment condition, with the Clinician-Administered PTSD Scale for DSM-5 (CAPS-5), a structured interview that assesses PTSD symptoms during the last month (Weathers et al., 2013a). The CAPS-5 yields a dimensional total severity score ranging from 0 to 80, with higher scores reflecting greater PTSD symptom severity, a dimensional score per symptom cluster (intrusion symptoms, avoidance symptoms, negative alterations in cognitions and mood, and alterations in arousal and reactivity), and an assessment of diagnostic status. The CAPS-5 has excellent psychometric properties in various populations (e.g., internal consistency ($\alpha = .80$ to $.90$) and interrater reliability ($\kappa = .90$ or higher); Weathers et al., 2001).

The Hospital Anxiety Depression Scale (HADS) was used to measure self-reported depression (HADS-D) and anxiety (HADS-A) (Zigmond & Snaith, 1983). It consists of 14 items, of which seven measure depression (e.g., "I still enjoy the things I used to enjoy") and seven measure anxiety ("Worrying thoughts go through

my mind"), with scores ranging from 0 (not at all) to 3 (most of the time). Total scores can range from 0 to 42. It showed excellent internal consistency (HADS-A: α = .68 to .93 (mean α = .83); HADS-D: α = .67 to .90 (mean α = .82); Bjelland et al., 2002).

The Post Traumatic Cognitions Inventory (PTCI) is a 33-item self-report questionnaire used to assess cognitions associated with posttraumatic psychopathology. It consists of three subscales that measure negative cognitions about self (21 items; e.g., "Nothing good can happen to me anymore"), negative cognitions about the world (seven items; "You can never know who will harm you") and self-blame (five items; "The event happened because of the way I acted"), with all items rated from 1 ('totally disagree') to 7 ('totally agree'). The PTCI includes a total score (range: 33 to 231) with excellent internal consistency (α = .97) (Foa et al., 1999).

Finally, in-depths interviews were held after the follow-up phase (see Appendix 1: Qualitative assessment in VR-ImRs for the complete interview). The interviews focused on the perspectives of patients on the treatments, with a special focus on the acceptability and feasibility of VR-ImRs.

Procedure

Participants were recruited within a mental health care institute in the Netherlands from May 2018 to June 2020. PTSD was diagnosed with the CAPS-5 (Weathers et al., 2013a) and comorbid disorders were diagnosed with the use of the Mini International Neuropsychiatric Interview (M.I.N.I.-plus; Sheehan et al., 1998). After inclusion in the study, and again at the pre-education and exploration, pre-treatment (both conditions), post-treatment and follow-up (5-weeks post-treatment) assessments, the complete set of outcome measures were completed through Qualtrics. In addition, the primary outcome measure (PCL-5) was completed twice a week throughout the complete trial.

Computerized randomization of participants occurred after the inclusion of the first participant and was executed by an independent person. Participants were randomized to baseline length and treatment order. If a participant dropped out, the allocation of that participant was put back in the pool and the remaining allocations were randomized again. Traumata treated per treatment condition were divided based on theme (e.g., sexual abuse, violence, emotional abuse) and severity so that both treatment conditions were approximately comparable in terms of trauma exposure of participants, but with at least one session per condition targeting the index trauma (always childhood sexual abuse). The participants and therapist (first author) were not informed about the outcome of the randomization until the last session of the education and exploration phase.

Treatment and therapists

Treatment consisted of 12 90-minute treatment sessions, preceded by five sessions of preparation (education and exploration phase). All treatment sessions were video recorded. Two licensed psychologists (one of them the first author) were involved, both with experience in ImRs, and in PTSD treatment in general. The therapists received training in the study protocol and treatment supervision throughout the study from an ImRs expert (fifth author), using video recordings of sessions. The ImRs protocol developed by Arntz and Weertman (1999) formed the basis for both treatment conditions. Deviations from this protocol resulted from the fact that in VR-ImRs, it was necessary to select the upcoming trauma target prior to the next session since the VR-ImRs worlds including the rescripted scenarios needed to be prepared in advance. For the same reason, each session addressed only one trauma memory. Furthermore, in each condition patients received three sessions in which the therapist intervened in the rescripting part followed by three sessions where the patients themselves rescripted. Finally, in both the ImRs and VR-ImRs conditions, trauma reprocessing immediately began in the first session since the usual "practice rescripting" was already done in the preceding education and exploration phase. A research assistant attended the VR-ImRs sessions to assist with controlling the VR-equipment.

Treatment integrity was assessed by two blind, independent raters who viewed randomly selected video tapes of 18 sessions out of a total of 96 available tapes, evenly divided over participants and phases (i.e., six tapes per treatment phase, and six for the education and exploration phase, which served as control for the two treatment phases). Each tape was rated using the ImRs Adherence and Competency Scale, where each item is rated 0–4 (Boterhoven de Haan et al., 2020). Items addressed the presence and order of specific ImRs techniques and the competence with which they were conducted. The mean rating for ImRs was 3.44 and the mean rating for VR-ImRs was 3.27, which indicates satisfactory adherence for both treatment conditions. The mean rating for tapes from the education and exploration phase was .12, which was expected because no active treatment was supposed to be given in this phase. Treatment integrity was thus supported by a significant difference between the education and exploration phase versus the treatment phases as determined by a one-way ANOVA ($F(2,15) = 470.971$, $p < .001$), with Tukey post hoc tests showing higher treatment integrity during the treatment phases compared to the education and exploration phase (p's $< .001$) and no significant difference between both treatment phases ($p = .372$).

Equipment and software

The VR equipment included an Alienware laptop, an Oculus Rift Virtual Reality headset with integrated headphones, and an Xbox One controller. The Virtual Reality headset was capable to detect and translate head movement into movement

within the virtual world. The research assistant navigated in the environment using the Xbox controller. Online environments and scenarios were constructed using pre-built objects (i.e., walls, floors, closets, beds), buildings (i.e., courthouse, prison) and characters (man, woman, boy, girl) (see Brinkman et al., 2022 for a visual presentation of actual footage used in the present study).

When patients entered the virtual world, they were presented with a scene that presents their memory of the trauma. These personalized 3D worlds were predesigned, using a program called 3D WorldBuilder (Tielman, 2017), based on information given by the patient. Patients took a first person perspective, and the avatar used for the patients resembled the patient as much as possible, although there were only limited choices. When the child perspective was taken, the child representation of the given character was chosen. Patients were unable to move around the world by themselves, instead their avatars were directed through the virtual world by the therapist or research assistant scene by scene.

Statistical analysis

Since this is a proof-of-concept pilot study it is exploratory in nature. Based on an a priori power analysis, due to the within-Ss repeated measures design, a total of 6 participants was required to detect a difference with a large effect size of f=0.3 between the linear trends of VR-ImRs and ImRs over six weeks (with a power [1-β] of .80, at $p < .05$, 12 repeated assessments, assuming a correlation of 0.7 between repeated measures). If this difference is detected and is in favor of ImRs, this suggests superiority of ImRs over VR-ImRs and little support to further investigate VR-ImRs, at least in its present form. In contrast, a difference in favor of VR-ImRs suggests superiority of VR-ImRs over ImRs and identifies VR-ImRs as a promising intervention that is worthy of further study. Differences with an effect size that is smaller than f=0.3 may require more nuanced conclusions regarding the usefulness of further studies of VR-ImRs.

Statistical analysis was done with SPSS, version 27 for Windows. Based on similar prior research, within and between effects of the treatments were assessed using mixed regression analysis (Arntz et al., 2013; Vlaeyen et al., 2010). Specifically, differences in means and slopes in the development over time of PCL-5 scores between the education and exploration phase, treatment conditions, and follow-up phase were examined in comparison with the baseline phase. The fixed model part consisted of (1) a general linear time effect starting with time = 0 when the first assessment of a participant was conducted ("Date within phase"), (2) a phase variable containing each phase ("Phase"), but with baseline as reference phase (thus contrasting each phase to baseline) and (3) five centered time-within-condition covariates (e.g. "Time within ImRs"), one for every phase to assess time by phase interactions, that is, linear time effects per phase (cf. Vlaeyen et al., 2001). Due to

the cross-over design, (4) the order of the treatment conditions ("Order") and all interactions between order and the above-mentioned fixed effects were added to the fixed part to control for effects of treatment order.

The random model part consisted of a random intercept for participant to capture between-subject outcome variation, plus ARMA11 for the within-subject covariance structure. Random slopes to allow interindividual variation in time and condition effects lead to reduced fit of the model or convergence problems and were therefore not included. The analytic strategy was to first assess a general time effect, next to assess the full model with all predictors entered, and then to delete in backward fashion the effects involving order, and time within phase effects that were non-significant. If the main time effect was non-significant, it was deleted at the last step. If time effects (i.e., linear trend) within VR-ImRs and ImRs were significant, they were tested against each other. If however one (or both) trend would be nonsignificant, a direct comparison was considered obsolete. Cohen's *d* effect sizes were calculated for the change of a phase with respect to baseline: *d* = the mean outcome difference between baseline and current phase, derived from the fixed part of the mixed regression, divided by the *SD* of the baseline assessments.

Since scores on the PCL-5 differed substantially between participants a bias in the analysis was observed, resulting in estimated means differing from the actual observed means. Due to this bias we decided to calculate person-centered variables for the PCL-5 scores by subtracting the mean score per participant from each measurement point. This solved the problem. We report estimated means and figures of back-transformed estimated scores.

For the secondary measures of the PCL-5, the seven added items, an exploratory factor analysis was carried out, in line with previous research (Watson & Tellegen, 1985). This exploratory factor analysis indicated that the six added PCL-5 items that measured negative emotions (shame, anger, guilt, disgust, sadness, and anxiety) loaded strongly on one factor, while the item that measured happiness did not load on this factor. A new variable was therefore computed as the mean score of these six items, while the score on the happiness item was retained, making it possible to analyze negative and positive emotions separately. The same analytic strategy was used as for the primary outcome.

The secondary measures HADS and PTCI were also analyzed with mixed regression (AR1 had the best fit), without modelling of slopes within phases as only single assessments per phase were available. CAPS-5 scores were analyzed with Wilcoxon Signed Ranks Test for difference in total scores and McNemar Test for remission from PTSD. Additionally, a thematic content analysis was carried out on the qualitative assessments of acceptability and feasibility.

Results

PCL-5

Compared to baseline, visual inspection of the individual PCL-5 total scores of the six participants suggests a decrease of these scores during ImRs and lower scores during follow-up. During VR-ImRs, PCL-5 total scores appeared to improve in only three participants (1, 5 and 6) and to increase in the other three (2, 4 and 8) (Figure 3; see Supplementary Figure 2 for the negative and positive emotion items that were added to the PCL-5).

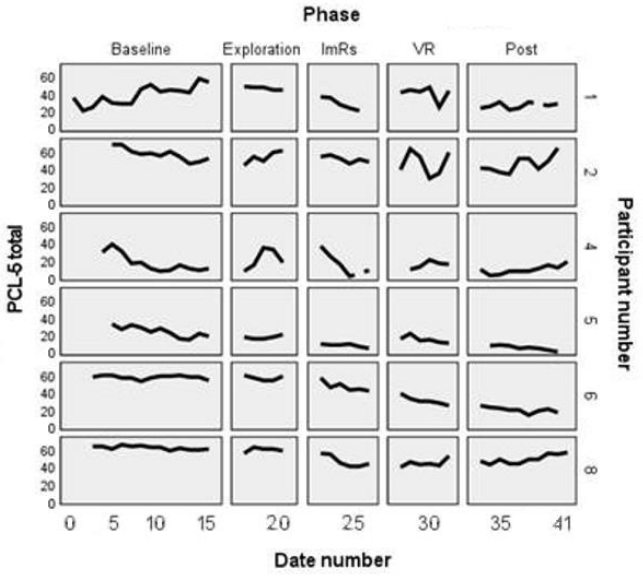


Figure 3.

Note. Participants 1, 4 and 5 were randomly assigned to order VR-ImRs – ImRs, thus receiving VR-ImRs before ImRs. Participants 2, 6 and 8 were assigned to order ImRs – VR-ImRs. Date within phase: each data point represents roughly half a week as measures were assessed twice a week. PCL-5 = PTSD Checklist for DSM-5.]

Table 2 presents the results of the mixed regression analysis. The education and exploration phase had no effect on the PCL-5 scores, but the main effects of both treatment conditions (i.e., the change halfway each treatment condition compared to baseline) were significant, as was the main effect of follow-up (as compared to baseline). This shows that overall, the treatments were effective in reducing PTSD symptoms and this was maintained at follow-up. The time effect within ImRs was significant, whereas the time effect within VR-ImRs failed to reach significance and was therefore deleted from the model. Effect sizes of ImRs vs baseline, and follow-

up vs baseline were large, while VR-ImRs vs baseline showed a medium to large effect size (Table 2). After the effect sizes were calculated we also calculated the additive Cohen's *d* to see to what extent each phase contributed to the total effect. Additive Cohen's *d*'s showed that ImRs was responsible for 73 percent of the total effect, while VR-ImRs was only responsible for 17 percent.

Supplementary Figure 1 depicts the predicted means from the analysis. When all predictors were entered the linear effect of time, order, order by phase, all centered time-within-condition effects, and all order by centered time-within-condition effects were non-significant (p's > .18) except for the highly significant effect of time-within-ImRs (p = .001). After stepwise deleting, the effect of time-within-ImRs remained the only significant time effect (p < .001) showing a steep decrease of PTSD symptoms within this treatment condition. Order and order by phase effects were also non-significant, but were kept in the model to control for order effects, as described in the method section. These results showed that ImRs had a far greater effect on the decrease of self-reported PTSD symptoms than VR-ImRs did, and that the order in which the treatments were given did not affect these outcomes.

For the negative emotion items that were added to the PCL-5, the same results were found as for the PCL-5 total scores: education and exploration had no effect on the negative emotions scores, but the main effects of both treatment conditions were significant, as was the main effect of follow-up (Table 2). When all predictors were entered, again all of the above-mentioned predictors were non-significant (p's > .11) except the highly significant effect of time-within-ImRs, which remained significant after stepwise deleting (p < .001). After stepwise deleting, the effect of VR-ImRs by order also became significant, suggesting that it mattered which condition was offered first (see Supplementary Figure 1). Effect sizes of ImRs compared to baseline and of follow-up compared to baseline were large, while VR-ImRs showed a medium to large effect size. Additive Cohen's *d*'s showed that ImRs was responsible for 69 percent of the total effect and VR-ImRs for 19 percent.

For the happiness item that was added to the PCL-5, the only significant main effect was found for the ImRs treatment condition. This effect did not last however, because at follow-up the main effect was not significant anymore. Of all the above-mentioned predictors entered none was significant (p's > .10), except for the effect of VR-ImRs by order which became significant after stepwise deleting the other non-significant effects (Table 2). Effect sizes were small to medium for all phases, even though the additive Cohen's *d*'s showed that only the treatment conditions contributed to the total effect, while the negative additive Cohen's *d* for follow-up showed that the effect of the treatment phases loses some of its effect over time.

Table 2. Results of mixed regression analyses of PCL-5 total, negative, and positive emotions scores.

	Parameter	ß	SE	Df	*t*	*p*	Effect size Cohen's *d*[a]	Additive Cohen's *d*
PCL-5 Total	Intercept	8.59	2.26	10.05	3.78	.003		
	Order	-5.09	4.48	10.77	-1.14	.28		
	EE	-1.29	2.21	124.16	-.58	.56	.07	.07
	ImRs	-9.99	2.89	41.12	-3.46	.001	.90	.70
	VR-ImRs	-12.40	2.89	40.33	-4.30	<.001	.67	.16
	Follow-up	-17.87	3.26	15.88	-5.48	<.001	.96	.03
	Time within ImRs	-2.69	0.76	111.66	-3.56	<.001		
	EE * order	-.62	4.39	127.76	-.14	.89		
	ImRs * order	.88	5.72	47.07	.15	.88		
	VR-ImRs * order	11.98	6.04	38.45	1.98	.05		
	Follow-up * order	8.74	6.23	21.01	1.40	.18		
Negative emotions	Intercept	.58	.14	13.58	4.05	.001		
	Order	-.40	.289	14.52	-1.39	.19		
	EE	-.01	.15	131.50	-.07	.94	.009	.009
	ImRs	-.66	.19	45.99	-3.50	.001	.92	.76
	VR-ImRs	-.81	.19	45.17	-4.31	<.001	.68	.21
	Follow-up	-1.34	.21	18.86	-6.36	<.001	1.11	.13
	Time within ImRs	-.18	.05	125.73	-3.49	<.001		
	EE * order	.18	.29	134.58	.60	.55		
	ImRs * order	.18	.38	51.27	.47	.64		
	VR-ImRs * order	1.07	.40	41.26	2.70	.01		
	Follow-up * order	.48	.40	24.59	1.19	.25		
Positive emotions	Intercept	-.26	.15	15.33	-1.78	.10		
	Order	.39	.30	15.33	1.32	.21		
	EE	-.11	.20	98.02	-.57	.57	-1.00	-1.00
	ImRs	.47	.22	37.82	2.11	.04	.42	.36
	VR-ImRs	.44	.22	36.99	1.97	.06	.39	.37
	Follow-up	.45	.22	19.41	2.04	.06	.40	-.23
	EE * order	-.63	.39	98.02	-1.61	.11		
	ImRs * order	-.001	.45	37.82	-.003	1.00		
	VR-ImRs * order	-1.03	.45	36.99	-2.29	.03		
	Follow-up * order	-.53	.44	19.41	-1.20	.25		

Note. EE = education and exploration; PCL-5 = PTSD Checklist for DSM-5. Predictors were coded as follows: Phase coding for Baseline (4), EE (3); ImRs (2), VR-ImRs (1), and Follow-up (0) so that Baseline was the reference category; Time-within-Condition: 0 for measurements outside the condition, centered time (with half-week as unit) for measurements within condition (e.g., -2.5, -1.5, -.5, .5, 1.5, 2.5 for a 3- week condition); Order coding for ImRs-VR (-.5), and VR-ImRs (.5) to correct for the order of the treatment conditions. Beta's for main effects of phase represent the time point halfway the phase, not the end.

[a] Effect size Cohen's *d* = effect/SD, with SD derived from the baseline assessments as explained in the Method section, and effect equal to the beta's of the Mixed Regression model. Positive values denote improvement. Effect sizes represent effect at the end of the intervention, and were corrected for order.

HADS

For both the anxiety and depression scores on the HADS, mixed regression showed that education and exploration had no main effect, and that the main effects

of both treatment conditions were non-significant (Table 3). The only significant main effect was the effect of follow-up (as compared to baseline) for the HADS anxiety scores. Since HADS scores were only assessed once after each phase we could not account for effects within phases (slopes), so the only predictors entered were order, and order by phase. Of these predictors, order was significant for the HADS depression scores, showing that these scores decreased more when offered ImRs as the first treatment condition (see Supplementary Figure 4). For both treatment conditions medium to large effect sizes were found for anxiety scores, and for depression scores a small to medium effect size for conventional ImRs and a medium to large effect size for VR-ImRs. Additive Cohen's *d*'s showed that VR-ImRs had a slightly larger contribution to the effect on anxiety and depression scores than conventional ImRs. Individual scores and predicted means for secondary outcomes are depicted in Supplementary Figure 3 and 4.

Table 3. Results of mixed regression analyzes of HADS Anxiety and Depression scores.

	Parameter	ß	SE	Df	*t*	*p*	Effect size Cohen's *d*[a]	Additive Cohen's *d*
Anxiety	Intercept	14.50	1.63	11.97	8.87	<.001		
	Order	-7.00	3.27	11.97	-2.14	.05		
	Post baseline	-1.33	1.43	18.06	-.93	.36	.25	.25
	Post EE	-1.33	1.82	23.07	-.73	.47	.25	<.001
	Post ImRs	-3.33	2.08	22.83	-1.60	.12	.62	.31
	Post VR-ImRs	-4.17	2.08	22.83	-2.00	.06	.78	.40
	Post follow-up	-4.83	2.21	18.87	-2.19	.04	.90	-.06
	Post baseline * order	-.67	2.86	18.06	-.23	.82		
	Post EE * order	.67	3.64	23.07	.18	.86		
	Post ImRs * order	-.67	4.16	22.83	-.16	.87		
	Post VR-ImRs * order	5.00	4.16	22.83	1.20	.24		
	Post follow-up * order	-.33	4.41	18.87	-.08	.94		
Depression	Intercept	12.50	1.87	8.44	6.69	<.001		
	Order	-9.67	3.74	8.44	-2.59	.03		
	Post baseline	-.83	1.23	19.22	-.68	.51	.14	.14
	Post EE	-.33	1.65	22.35	-.20	.84	.06	-.09
	Post ImRs	-2.17	2.00	23.97	-1.08	.29	.37	.23
	Post VR-ImRs	-3.83	2.00	23.97	-1.92	.07	.65	.40
	Post follow-up	-2.83	2.22	22.16	-1.28	.22	.48	-.20
	Post baseline * order	.33	2.47	19.22	.14	.89		
	Post EE * order	.67	3.29	22.35	.20	.84		
	Post ImRs * order	1.67	4.00	23.97	.42	.68		
	Post VR-ImRs * order	5.67	4.00	23.97	1.41	.17		
	Post follow-up * order	3.00	4.45	22.16	.68	.51		

Note. EE = education and exploration; ImRs = conventional ImRs; HADS = Hospital Anxiety Depression Scale. Predictors were coded as follows: Phase coding for Start baseline (5), Post baseline (4), Post EE (3); Post ImRs (2), Post VR-ImRs (1), and Post follow-up (0) so that Baseline was the reference category; Order coding for ImRs-VR (-.5), and VR-ImRs (.5) to correct for the order of the treatment conditions.

[a] Effect size Cohen's *d* = effect/SD, with SD derived from the baseline assessments as explained in Method section, and effects equal to the beta's of the Mixed Regression model. Positive values denote improvement. Effect size represents effect at the end of the intervention, and were corrected for order.

PTCI

Mixed regression of PTCI total scores showed that after the education and exploration phase, participants already had a significantly lower score in comparison to the assessment at the start of the baseline (Table 4). Education and exploration had no main effect on the PTCI scores, but the main effects of both treatment conditions were significant, as was the main effect of follow-up (as compared to start of baseline). Since PTCI scores were also only assessed once after each phase we could not account for effects within phases (slopes), so again the only predictors entered were order, and order by phase, which all were non-significant. The direct comparison between VR-ImRs and ImRs failed to reach significance. For both treatment conditions medium effect sizes were found, although additive Cohen's *d*'s showed that VR-ImRs contributed two and a half times as much to the effect than conventional ImRs did.

CAPS-5 PTSD Severity

The median score on the CAPS-5 decreased from pre-treatment (*Md* = 37) to post-treatment (*Md* = 5), although the effect was non-significant, likely due to the fact that two participants actually experienced a higher severity score at the end of treatment (Table 5). A Wilcoxon Signed Rank Test showed a statistically non-significant reduction in severity of PTSD symptoms according to the CAPS-5 following the complete treatment (both conditions), $z = -1.57$, $p = .116$, with a medium effect size ($r = .45$).

Table 4. Results of mixed regression analyzes of PTCI total scores.

	Parameter	ß	SE	Df	*t*	*p*	Effect size Cohen's *d*[a]	Additive Cohen's *d*
Total	Intercept	109.00	17.85	4.80	6.11	.002		
	Order	-85.33	35.70	4.80	-2.39	.06		
	Post baseline	-12.83	5.99	19.63	-2.14	.05	.23	.23
	Post EE	-11.33	8.36	20.69	-1.36	.19	.20	-.03
	Post ImRs	-27.83	10.81	22.13	-2.57	.02	.50	.14
	Post VR-ImRs	-33.50	10.81	22.13	-3.10	.005	.60	.34
	Post follow-up	-29.83	12.70	23.13	-2.36	.03	.53	-.14
	Post baseline * order	2.33	12.00	19.63	.20	.85		
	Post EE * order	4.00	16.71	20.69	.24	.81		
	Post ImRs * order	-15.67	21.63	22.13	-.72	.48		
	Post VR-ImRs * order	13.00	21.63	22.13	.60	.55		
	Post follow-up * order	8.33	25.34	23.13	.33	.75		

Note. EE = education and exploration; ImRs = conventional ImRs; PTCI = Post Traumatic Cognitions Inventory. Predictors were coded as follows: Phase coding for Start baseline (5), Post baseline (4), Post EE (3); Post ImRs (2), Post VR-ImRs (1), and Post follow-up (0) so that Baseline was the reference category; Order coding for ImRs-VR (-.5), and VR-ImRs (.5) to correct for the order of the treatment conditions.

[a] Effect size Cohen's *d* = effect/S.D., with S.D. derived from the baseline assessments as explained in Method section, and effects equal to the beta's of the Mixed Regression model. Positive values denote improvement. Effect size represents effect at the end of the intervention, and were corrected for order.

Table 5. Clinician Administered PTSD Scale scores pre- and post-treatment per participant

Participant	Pre-treatment	Post-treatment
1	37 – 13*	3 – 1
2	39 – 16*	48 – 15
4	26 – 9*	7 – 2
5	25 – 11*	3 – 0
6	37 – 13*	2 – 0
8	45 – 16*	52 – 18*

Note. The first number indicates the severity of the PTSD symptoms (range 0 – 80). The second number indicates the number of PTSD symptoms present, i.e. with ratings > 1 (range 0 – 20). * Indicates the presence of a clinical diagnosis of PTSD.

Remission from PTSD

Out of the six participants who finished the complete treatment and were assessed with the CAPS-5 at follow-up, five did not meet the criteria of PTSD anymore. A McNemar Test showed a significant effect (p = .031, 1-tailed).

Additional findings

There were three main topics (expectations, satisfaction, complaints) that were assessed in qualitative interviews to evaluate the relative acceptability and feasibility of the treatments. Concerning expectations, all participants stated they did not have any specific expectations regarding conventional ImRs versus VR-ImRs. Some participants however indicated that they were particularly hopeful about VR-ImRs due to positive statements on the internet. Half of the interviewed participants said they had expected more advanced VR functionalities.

Concerning satisfaction about the extent to which both treatments helped to achieve their goals, two participants said that both treatments helped, and that a combination of both treatments had their preference. They would also recommend this to others even though one participant noted that for people with difficulties imagining VR-ImRs would be more helpful, otherwise conventional ImRs was recommended. One participant found it difficult to say which treatment helped the most because the VR images replaced the original trauma images, but this participant still preferred VR-ImRs over conventional ImRs and would also recommend VR-ImRs to others, because the images were tailored to a specific patient's story. However, the other three participants said conventional ImRs was the only intervention that helped, and they would prefer to receive conventional ImRs when given the choice and recommend other to do the same.

To assess for realisticness of the scenarios, participants were asked after each VR-ImRs session to what extent they found the presented scenarios realistic (range: 0 to 100). Almost all sessions were rated between 80 – 100, with only three sessions rated lower than 70 (M = 82.3; SD = 10.5). Bouchard et al. (2004) found

that this is a reliable way to measure presence in a virtual world in a minimally intrusive manner.

A complaint expressed by almost all participants was the faltering technique of VR-ImRs. One problem was that the software sometimes changed perspective while this was not intended (e.g., from victim to perpetrator, or victim to helicopter view). Furthermore, it was impossible to create a VR scenario that exactly replicated the participants' trauma memories. This made some participants use their own imagination to make the scenario more similar to their own memory. One participant found the soberness of the images pleasant because it made the scenario less aversive. All participants found VR-ImRs less aversive than conventional ImRs, except for one participant who admitted that she sometimes closed her eyes during the VR-ImRs sessions. Some participants experienced headaches during VR-ImRs, while others became nauseous, but it cannot be ruled out that this was because of anxiety. Also, the VR-goggles tended to fog up and the foam around the goggles became wet when participants were emotional, making it uncomfortable to wear.

A finding not reported by the participants but observed by the therapist during the VR-treatments is that VR-ImRs seemed to provoke more bodily enactments of the scenarios, such as moving the chair backwards to get away from an image, pushing a perpetrator away, and pushing against the table thinking this would help keeping a virtually presented door closed.

Regarding feasibility, the main finding was that building a VR-scenario took around two hours of preparation time per session. The VR-ImRs session itself took a little less time than ImRs, making the total time invested per participant around 18 hours for VR-ImRs and around nine hours for ImRs.

Discussion

This proof-of-concept pilot study constitutes the first exploratory comparison of ImRS and VR-ImRs as a treatment for PTSD due to childhood sexual abuse. We used a non-concurrent multiple baseline cross-over design testing effectiveness of both active treatments and comparing VR-ImRs to ImRs. The findings suggest an absence of significant changes during VR-ImRs, whereas changes were significant during ImRs. Superiority of ImRs over VR-ImRs in reducing PTSD symptoms as assessed with the total score of the PCL-5, can be concluded. We did not observe any improvement of depressive symptoms as assessed with the HADS. With regard to anxiety symptoms, we did not find any improvement during the treatment phases, although there was a main effect on anxiety scores at follow-up. However, due to the design of the study, these long-term effects can only be attributed to combined effect of the two treatments conditions. As far as reduction of trauma-related cognitions is concerned, both treatment conditions resulted in improved trauma-related cognitions while the conditions did not differ significantly from each other.

When looking at the primary outcome, mixed regression analyses showed that for the education and exploration, follow-up, and VR-ImRs phases, there was no evidence for significant time-within-condition effects. For ImRs however the time-within-condition effect was strong, indicating that only ImRs contributed to a decrease in PTSD symptoms during the treatment itself. Since there was no general time effect, mere passage of time can be ruled out as an alternative explanation, in concordance with earlier ImRs research (Arntz et al., 2013). The order of the treatment conditions did not contribute to the effect on the PCL-5 total score.

In sum, we found support for the efficacy of ImRs in reducing PTSD symptoms due to childhood sexual abuse, but not for VR-ImRs. Both conditions seemed ineffective in immediately reducing anxiety and depressive symptoms, but both ImRs and VR-ImRs were effective in reducing negative trauma-related cognitions.

The majority of our participants indicated that they preferred ImRs or a combination of both treatments to VR-ImRs alone. Research on underlying mechanisms suggests that activation of perceptual trauma memories is needed before change in the meaning of the traumatic experience can take place through ImRs (Arntz, 2012; Arntz, 2015). Probably the present VR environment was unable to sufficiently activate participants' fear structure because it was difficult to virtually recreate the exact 'hotspots' in the trauma scenarios accurately enough. Reger et al. (2016) already suggested this as an explanation for the larger effect of prolonged exposure at 12- and 26-week follow-up compared to VR exposure.

The lack of effect of VR-ImRs could also be due to the design of the present study, that required several modifications to the original ImRs protocol. The original protocol consists of six successive sessions in which the therapist intervenes in the traumatic scenario, followed by six sessions in which the patient intervenes him- or herself. We changed the protocol slightly to three sessions in which the therapist intervened followed by three sessions in which the patient intervened. Although this did not appear to hinder the efficacy of ImRs, it is unclear if this was also the case for VR-ImRs. Also, one could argue that the fact that VR-ImRs scenario's had to be prepared before each session and could not be changed during sessions, limited our flexibility during the actual rescripting if the scenario did not fit with the traumatic memories of the participant and/or the direction of the rescripting as it developed during the session. However, Ehlers et al. (2005) found that preparing a new scenario beforehand was also effective in itself, although in that study these preparations also included modifying negative appraisals through cognitive therapy techniques (e.g., Socratic questioning) before rescripting took place.

Even though the treatments resulted in large reductions of PTSD symptoms at follow-up, earlier studies showed a much larger effect size for ImRs (Arntz et al., 2013; Boterhoven de Haan et al., 2020; Morina et al., 2017). An explanation for these smaller effect sizes could be that in contrast with these earlier studies, participants in the present study only received six instead of the typical 12 sessions of ImRs. This could also explain the lack of effect of ImRs on the secondary outcomes, again in

contrast to earlier research (for depression, see Arntz et al., 2013; Boterhoven de Haan et al., 2020; for anxiety, see Morina et al., 2017).

A possible criticism could also be that the present study had a high drop-out rate (25%) in comparison to other studies of ImRs, that had drop-out rates of 8.1% and 0%, respectively (Boterhoven de Haan et al., 2020; Arntz, 2012). However, the present drop-out rate is in line with a meta-analysis of trauma-focused therapy for survivors of childhood abuse, which found an average dropout rate across studies of 22.3% (Ehring et al., 2014). The reasons given for drop-out in the present study were not directly linked to VR-ImRs; one participant had little belief that ImRs in general could help and one participant dropped-out due to logistic problems at the treatment facility (e.g., not warned when the therapist was ill, and not sent new directions when the treatment location was changed).

Another criticism might be that the participants who finished the full treatment were all treated by the same therapist, who is also the first author. We tried to overcome this by involving a second therapist in the study, but unfortunately both drop-outs were treated by this second therapist.

A strength of the present study is that we used an online questionnaire tool that kept the therapists blind to the outcome data and ensured that the ratings were not influenced by the presence of the therapist. Also, therapists and participants were blind to treatment condition up to the very last session of the education and exploration phase (of course, the use of VR-equipment made it impossible to blind participants and therapists to treatment condition).

In summary, the present study suggests that ImRs is superior to VR-ImRs in the tested form. The human imagery capacity might not be so simple to improve upon by artificial virtual reality. Despite the difficulties described above, there were some promising observations during VR-ImRs, such as the bodily re-enactments of the trauma scenarios that suggest the potential for intensive reliving of these scenarios in some patients. To give VR-ImRs a fair chance, studies of VR-ImRs using more advanced and user-friendly software are much needed. Possibly, adding multisensory experiences and actual pictures of the perpetrator to the avatar in the VR scenario would aid the activation of the fear structure. As Ma et al. (2021) point out, VR technologies are evolving rapidly so these and other changes are likely already possible with the newest equipment.

Maples-Keller et al. (2017) state that the use of VR treatment eliminates a potential barrier for people with difficulties imagining or visualizing. They also found that some people are willing to use VR treatment, but not to talk to a counselor in person. We found that within our small sample of patients with PTSD due to childhood sexual abuse however, none of them had difficulties imagining these aversive memories, which could have contributed to the effect of ImRs while by-passing the potential benefits of VR-ImRs. Using VR-ImRs in patients that have difficulty imagining or that are highly avoidant may be needed to unlock the full potential of VR-ImRs.

Authors' note

Declarations: Sven van Kuik is now at PsyQ, Amsterdam, The Netherlands. Arnoud Arntz has published on Imagery Rescripting and occasionally gives workshops on the Imagery Rescripting. The remuneration for these workshops goes to the University of Amsterdam to support research. No other interests declared.

Acknowledgements: We thank all patients and the second therapist, [blinded], involved in this study. We give special thanks to [blinded] for helping with adapting the software to our specific needs.

Funding: This study is part of the project Virtual e-Coaching and Storytelling Technology for Post-Traumatic Stress Disorder, which is financed by the Netherlands Organization for Scientific Research (pr. nr. 314-99-104).

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Arntz, A. (2012). Imagery rescripting as a therapeutic technique: Review of clinical trials, basic studies, and research agenda. *Journal of Experimental Psychopathology*, *3*(2), 189–208. https://doi.org/10.5127/jep.024211

Arntz, A. (2015). Imagery rescripting for posttraumatic stress disorder. In N. C. Thoma & D. McKay (Eds.), *Working with emotion in cognitive-behavioral therapy: Techniques for clinical practice* (pp. 203–215). The Guilford Press.

Arntz, A., & van Genderen, H. (2009). *Schema therapy for borderline personality disorder* (1st ed.). Wiley-Blackwell.

Arntz, A., Sofi, D., & van Breukelen, G. (2013). Imagery rescripting as treatment for complicated PTSD in refugees: A multiple baseline case series study. *Behaviour Research and Therapy*, *51*(6), 274–283. https://doi.org/10.1016/j.brat.2013.02.009

Arntz, A., & Weertman, A. (1999). Treatment of childhood memories: Theory and practice. *Behaviour Research and Therapy*, *37*(8), 715–740. https://doi.org/10.1016/s0005-7967(98)00173-9

Beck, J. G., Palyo, S. A., Winer, E. H., Schwagler, B. E., & Ang, E. J. (2007). Virtual reality exposure therapy for PTSD symptoms after a road accident: An uncontrolled case series. *Behavior Therapy*, *38*(1), 39–48. https://doi.org/10.1016/j.beth.2006.04.001

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the hospital anxiety and depression scale. *Journal of Psychosomatic Research*, *52*(2), 69–77. https://doi.org/10.1016/s0022-3999(01)00296-3

Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The posttraumatic stress disorder checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress*, *28*(6), 489–498. https://doi.org/10.1002/jts.22059

Botella, C., Serrano, B., Baños, R. M., & García-Palacios, A. (2015). Virtual reality exposure-based therapy for the treatment of post-traumatic stress disorder: A review of its efficacy, the adequacy of the treatment protocol, and its acceptability. *Neuropsychiatric Disease and Treatment*, 2533–2545. https://doi.org/10.2147/ndt.s89542

Boterhoven de Haan, K. L., Lee, C. W., Fassbinder, E., van Es, S. M., Menninga, S., Meewisse, M. L., Rijkeboer, M., Kousemaker, M., & Arntz, A. (2020). Imagery rescripting and eye movement desensitisation and reprocessing as treatment for adults with post-traumatic stress disorder from childhood trauma: randomised clinical trial. *The British Journal of Psychiatry*, *217*(5), 609–615. https://doi.org/10.1192/bjp.2020.158

Bouchard, S., Robillard, G., St-Jacques, J., Dumoulin, S., Patry, M., & Renaud, P. (2004). Reliability and validity of a single-item measure of presence in VR. *Proceedings. Second International Conference on Creating, Connecting and Collaborating through Computing*, 59–61. https://doi.org/10.1109/have.2004.1391882

Brinkman, W., van Kuik, S., van Emmerik, A. A. P., Uduwa-Vidanalage, E., Schouten, C., Arntz, A. (2022). Screenshots of Virtual Reality Imagery Rescripting tool. 4TU.ResearchData. Dataset. https://doi.org/10.4121/21170665.v1

Creswell, C., & Waite, P. (2009). The use of CBT with children and adolescents. In P. Waite & T. Williams (Eds.), *Obsessive compulsive disorder: Cognitive behaviour therapy with children and young people (CBT with children, adolescents and families)* (1st ed., pp. 19–32). Routledge.

Ehlers, A., Clark, D. M., Hackmann, A., McManus, F., & Fennell, M. (2005). Cognitive therapy for post-traumatic stress disorder: Development and evaluation. *Behaviour Research and Therapy*, *43*(4), 413–431. https://doi.org/10.1016/j.brat.2004.03.006

Ehring, T., Welboren, R., Morina, N., Wicherts, J. M., Freitag, J., & Emmelkamp, P. M. G. (2014). Meta-analysis of psychological treatments for posttraumatic stress disorder in adult survivors of childhood abuse. *Clinical Psychology Review*, *34*(8), 645–657. https://doi.org/10.1016/j.cpr.2014.10.004

Foa, E. B., Ehlers, A., Clark, D. M., Tolin, D. F., & Orsillo, S. M. (1999). The Posttraumatic Cognitions Inventory (PTCI): Development and validation. *Psychological Assessment*, *11*(3), 303–314. https://doi.org/10.1037/1040-3590.11.3.303

Gerardi, M., Cukor, J., Difede, J., Rizzo, A., & Rothbaum, B. O. (2010). Virtual reality exposure therapy for post-traumatic stress disorder and other anxiety disorders. *Current Psychiatry Reports*, *12*(4), 298–305. https://doi.org/10.1007/s11920-010-0128-4

Jung, K., & Steil, R. (2013). A randomized controlled trial on cognitive restructuring and imagery modification to reduce the feeling of being contaminated in adult survivors of childhood sexual abuse suffering from posttraumatic stress disorder. *Psychotherapy and Psychosomatics, 82*(4), 213–220. https://doi.org/10.1159/000348450

Kunze, A. E., Arntz, A., Morina, N., Kindt, M., & Lancee, J. (2017). Efficacy of imagery rescripting and imaginal exposure for nightmares: A randomized wait-list controlled trial. *Behaviour Research and Therapy*, *97*, 14–25. https://doi.org/10.1016/j.brat.2017.06.005

Low, T. L., Ho, R., Ho, C., & Tam, W. (2021). The efficacy of virtual reality in the treatment of binge-purging eating disorders: A meta-analysis. *European Eating Disorders Review*, *29*(1), 52–59. https://doi.org/10.1002/erv.2804

Ma, L., Mor, S., Anderson, P. L., Baños, R. M., Botella, C., Bouchard, S., Cárdenas-López, G., Donker, T., Fernández-Álvarez, J., Lindner, P., Mühlberger, A., Powers, M. B., Quero, S., Rothbaum, B., Wiederhold, B. K., & Carlbring, P. (2021). Integrating virtual realities and psychotherapy: SWOT analysis on VR and MR based treatments of anxiety and stress-related disorders. *Cognitive Behaviour Therapy, 50*(6), 509-526. https://doi.org/10.1080/16506073.2021.1939410

Maples-Keller, J. L., Bunnell, B. E., Kim, S. J., & Rothbaum, B. O. (2017). The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. *Harvard Review of Psychiatry*, *25*(3), 103–113. https://doi.org/10.1097/hrp.0000000000000138

Morina, N., Ijntema, H., Meyerbröker, K., & Emmelkamp, P. M. G. (2015). Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. *Behaviour Research and Therapy*, *74*, 18–24. https://doi.org/10.1016/j.brat.2015.08.010

Morina, N., Lancee, J., & Arntz, A. (2017). Imagery rescripting as a clinical intervention for aversive memories: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, *55*, 6–15. https://doi.org/10.1016/j.jbtep.2016.11.003

Opriş, D., Pintea, S., García-Palacios, A., Botella, C., Szamosközi, Ş., & David, D. (2012). Virtual reality exposure therapy in anxiety disorders: A quantitative meta-analysis. *Depression and Anxiety*, *29*(2), 85–93. https://doi.org/10.1002/da.20910

Reger, G. M., Koenen-Woods, P., Zetocha, K., Smolenski, D. J., Holloway, K. M., Rothbaum, B. O., Difede, J., Rizzo, A. A., Edwards-Stewart, A., Skopp, N. A., Mishkind, M., Reger, M. A., & Gahm, G. A. (2016). Randomized controlled trial of prolonged exposure using imaginal exposure vs. virtual reality exposure in active duty soldiers with deployment-related posttraumatic stress disorder (PTSD). *Journal of Consulting and Clinical Psychology*, *84*(11), 946–959. https://doi.org/10.1037/ccp0000134

Sheehan, D. V., Lecrubier, Y., Harnett Sheehan, K., Janavs, J., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The mini-

international neuropsychiatric interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, *59*(20), 22–33.

Smucker, M. R., Dancu, C., Foa, E. B., & Niederee, J. L. (1995). Imagery rescripting: A new treatment for survivors of childhood sexual abuse suffering from posttraumatic stress. *Journal of Cognitive Psychotherapy*, *9*(1), 3–17. https://doi.org/10.1891/0889-8391.9.1.3

Stott, R. (2007). When head and heart do not agree: A theoretical and clinical analysis of rational-emotional dissociation (RED) in cognitive therapy. *Journal of Cognitive Psychotherapy*, *21*(1), 37–50. https://doi.org/10.1891/088983907780493313

Tielman, M. L., Neerincx, M. A., Bidarra, R., Kybartas, B., & Brinkman, W. (2017). A therapy system for post-traumatic stress disorder using a virtual agent and virtual storytelling to reconstruct traumatic memories. *Journal of Medical Systems, 41*(125), 1-10. https://doi.org/10.1007/s10916-017-0771-y

Van Meggelen, M., Morina, N., van der Heiden, C., Brinkman, W., Yocarini, I. E., Tielman, M. L., Rodenburg, J., van Ee, E., van Schie, K., Broekman, M. E., & Franken, I. (In press). A randomized controlled trial to pilot the efficacy of a computer-based intervention for the treatment of post-traumatic stress disorder. *Frontiers in Digital Health.*

Vlaeyen, J. W., de Jong, J., Geilen, M., Heuts, P. H., & van Breukelen, G. (2001). Graded exposure in vivo in the treatment of pain-related fear: A replicated single-case experimental design in four patients with chronic low back pain. *Behaviour Research and Therapy*, *39*(2), 151–166. https://doi.org/10.1016/s0005-7967(99)00174-6

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*(2), 219–235. https://doi.org/10.1037/0033-2909.98.2.219

Weathers, F. W., Keane, T. M., & Davidson, J. R. (2001). Clinician-administered PTSD scale: a review of the first ten years of research. *Depression and Anxiety*, *13*(3), 132–156. https://doi.org/10.1002/da.1029

Webb, A. K., Vincent, A. L., Jin, A. B., & Pollack, M. H. (2015). Physiological reactivity to nonideographic virtual reality stimuli in veterans with and without PTSD. *Brain and Behavior*, *5*(2), 1–9. https://doi.org/10.1002/brb3.304

Weathers, F.W., Blake, D.D., Schnurr, P.P., Kaloupek, D.G., Marx, B.P., & Keane, T.M. (2013a). The clinician-administered PTSD scale for DSM-5 (CAPS-5). [Assessment] Available from www.ptsd.va.gov

Weathers, F.W., Litz, B.T., Keane, T.M., Palmieri, P.A., Marx, B.P., & Schnurr, P.P. (2013b). The PTSD Checklist for *DSM-5* (PCL-5). Scale available from the National Center for PTSD at www.ptsd.va.gov

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x

Appendix 1: Qualitative assessment in VR-ImRs

The overall goal of the present interview is to evaluate how the patient experienced VR-ImRs compared to conventional ImRs.

Theme: Expectation
Initial question:

What did you expect from the VR-ImRs therapy? What did you expect from the ImRs therapy?

Follow up question:

How logical did VR-ImRs seem to you? How logical did ImRs seem?

Theme: Satisfaction
Initial question:

To what extent has VR-ImRs rescripting been useful (or not) in achieving your objectives / in reducing PTSD symptoms? To what extent has ImRs been useful?

Follow up questions:

1. How would you describe your overall experience with VR-ImRs compared to ImRs?
2. Did you encounter any difficulties with VR-ImRs? If so, what difficulties? / How about with conventional ImRs? If so, what difficulties?
3. How did you experience that within the VR treatment it was not always possible to recreate every detail of your story? Was this lack of flexibility a problem? Did you add any imagery to the VR-ImRs? Was it beneficial to add imagery to VR-ImRs?
4. Did it matter that the VR worlds did not represent 100% your real memory?
5. Was imagining a problem in ImRs?
6. To what extent did you find VR-ImRs aversive? To what extent did you find ImRs aversive?

Theme: Preference
Initial question:

After experiencing both treatments, if you had to choose your own treatment plan what would you select from the following? And why?

a.) VR-ImRs
b.) conventional ImRs
c.) combination of VR-ImRs and conventional ImRs (in what order?)
d.) a totally different treatment

Follow up questions:

(*Based on what the patient answered above choose one of the following*)

1. In what aspects did you find VR-ImRs better than ImRs/Why did you prefer VR-ImRs over conventional ImRs?
2. In what aspects did you find ImRs better than VR-ImRs/Why did you prefer conventional ImRs over VR-ImRs?

3. Why did you prefer the combination?
4. What made you choose a totally different treatment?

Theme: Recommendation
Initial question:

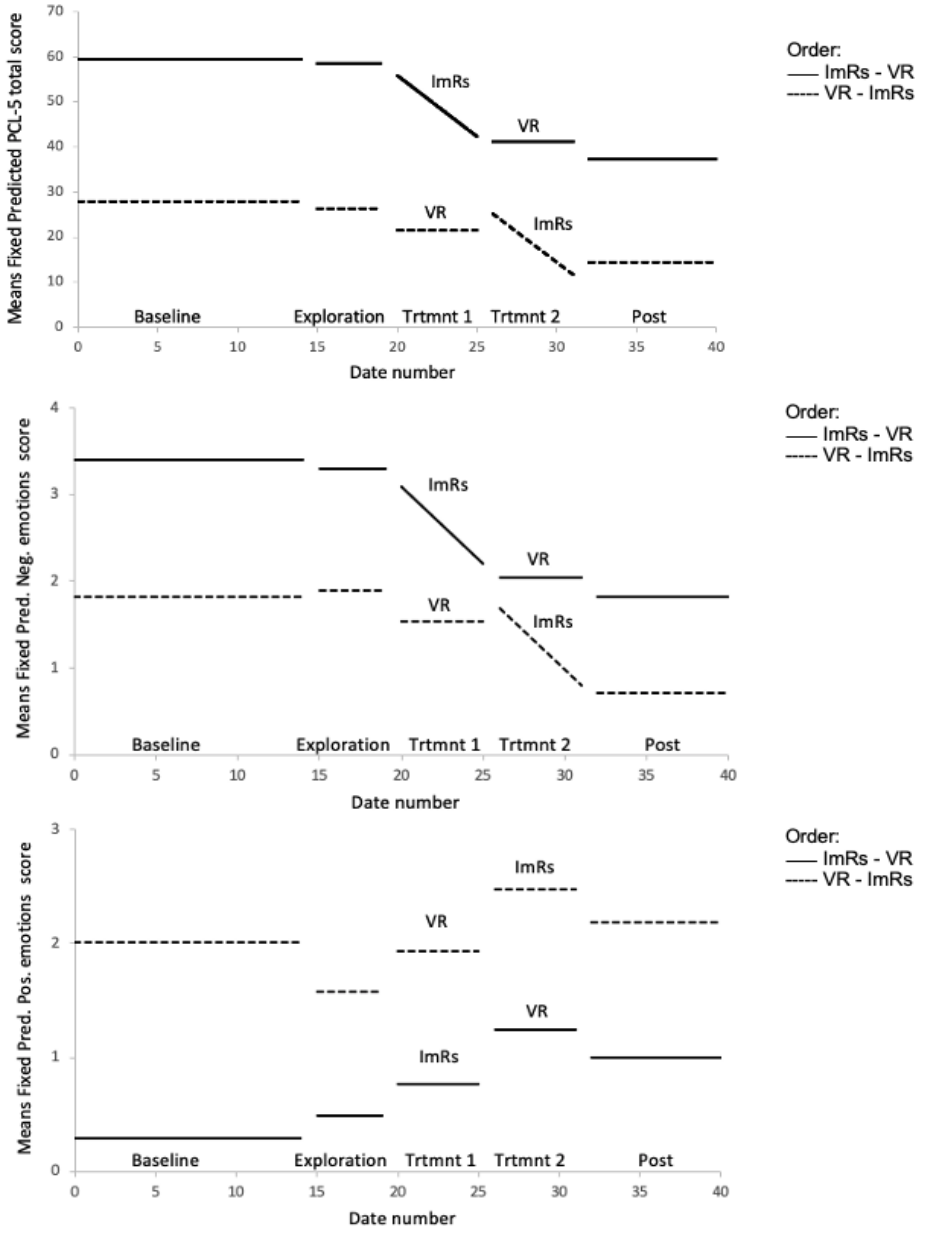
How confident would you feel in recommending VR-ImRs to a friend who experiences similar problems? How about ImRs?

Follow up questions:

1. Can you indicate anything related to the treatment received that could be useful for future improvements?/ Do you have any ideas as to what worked or did not work with VR-ImRs and ImRs therapy?
2. Is there anything else you would like to add?

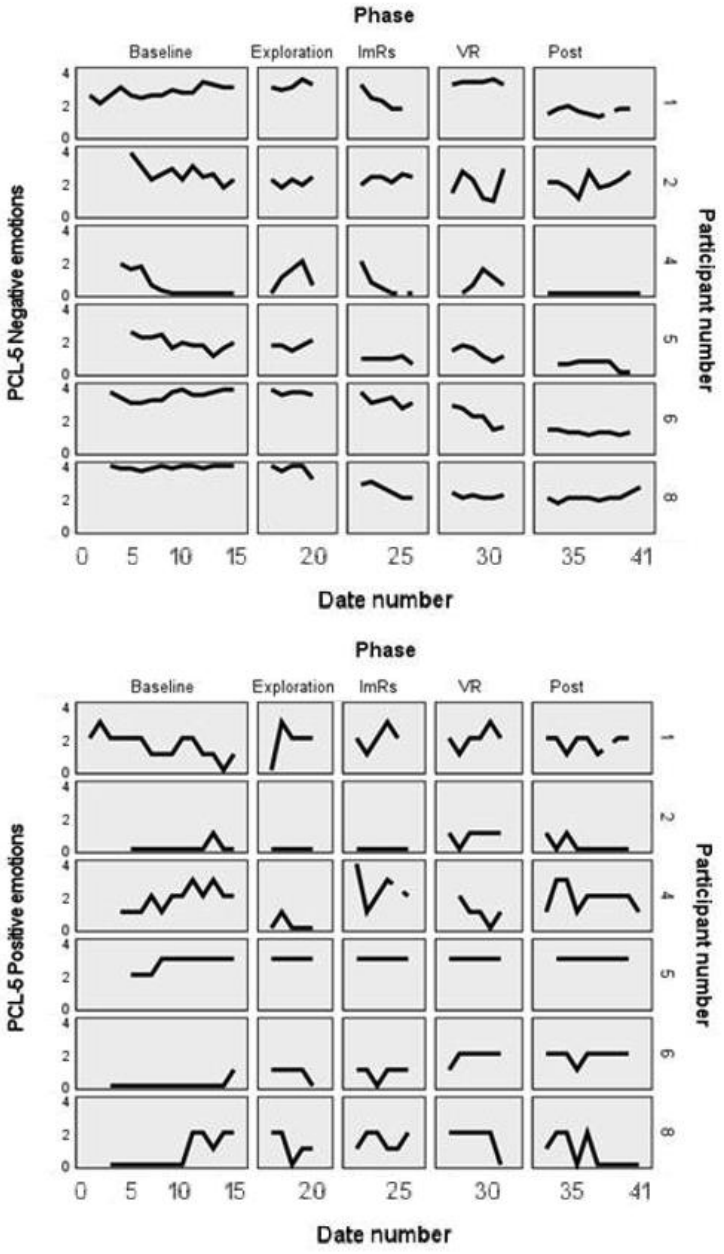
Appendix 2: Supplementary Figures

Supplementary Figure 1. Predicted estimated means of PCL-5 by the mixed regression model



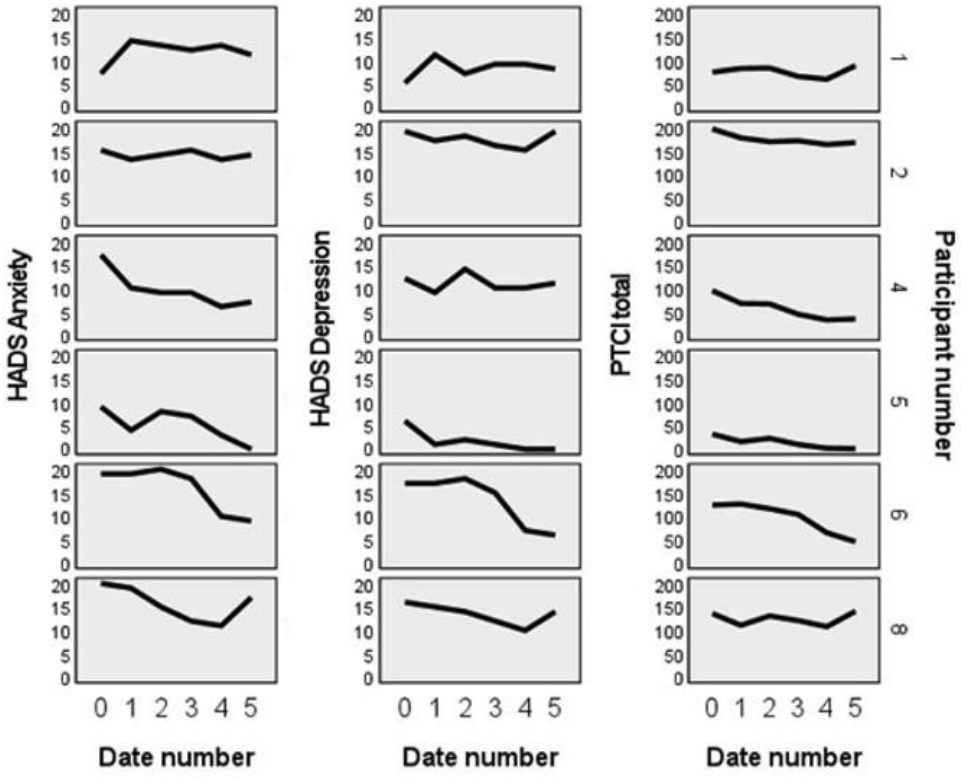
Note. PCL-5 total (upper panel), PCL-5 Negative emotions (middle panel) and PCL-5 Positive emotions (bottom panel).

Supplementary Figure 2. Individual PCL-5 negative and positive emotions scores



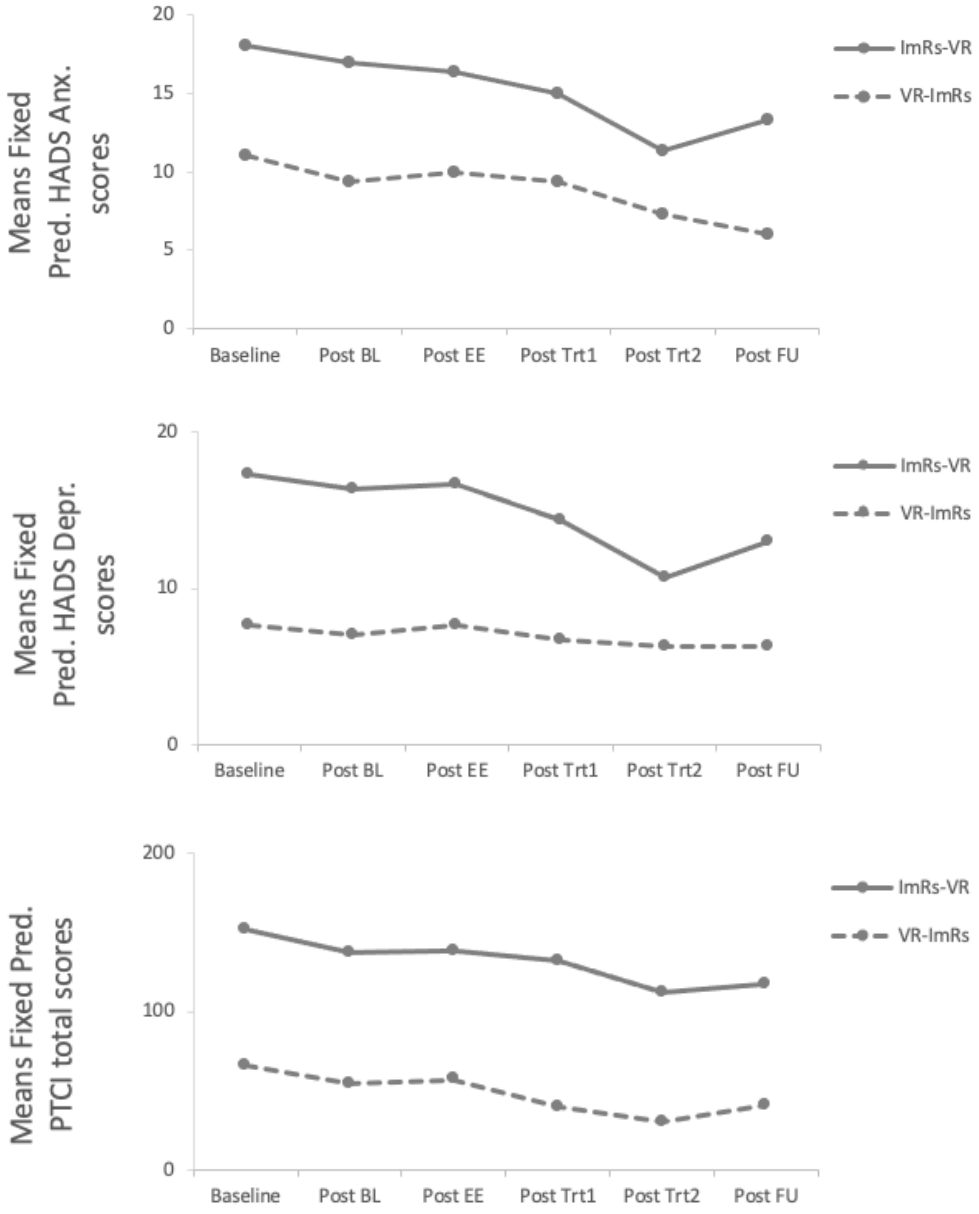
Note. Upper panel: Individual PCL-5 negative scores; bottom panel: Individual positive emotions scores. Participants 1, 4 and 5 were randomly assigned to order VR – ImRs, thus receiving VR-ImRs before getting ImRs. Participants 2, 6 and 8 were assigned to order ImRs – VR. Date within phase: each data point represents roughly half a week as measures were assessed twice a week.

Supplementary Figure 3. Individual HADS Anxiety, HADS depression and PTCI total scores



Note. Date numbers were coded as follows: Start baseline (0), Post baseline (1), Post education and exploration (2), Post ImRs (3), Post VR-ImRs (4), and Post follow-up (5). Participants 1, 4 and 5 were randomly assigned to order VR – ImRs, thus receiving VR-ImRs before getting ImRs. Participants 2, 6 and 8 were assigned to order ImRs – VR.

Supplementary Figure 4. Predicted estimated means of HADS and PTCI by the mixed regression model



Note. HADS Anxiety (upper panel), HADS Depression (middle panel) and PTCI total (bottom panel).