The appropriateness of using a counter app in experimental studies assessing unwanted intrusive thoughts

Karina Wahl[1], Roselind Lieb[1], Martin Kollárik[1], & Christine Purdon[2]

[1] University of Basel, Department of Psychology, Clinical Psychology and Epidemiology, Missionsstr. 62a, 4055 Basel, Switzerland

[2] Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Corresponding author:

Karina Wahl, Clinical Psychology and Epidemiology, Department of Psychology, University of Basel, Missionsstr. 62a, 4055 Basel, Switzerland; email: karina.wahl@unibas.ch; phone: +41 61 207 0654

Acknowledgements

The authors wish to thank Martina Apostolo, Kim Buchmüller, Jeannine Häni, Pia Schülin, Andrea Schumacher, and Annalena Seitz for their help with the data collection.

Declaration of interest statement

All authors declare that they have no conflicts of interest.

JOCRD

Abstract

The reliable and valid assessment of unwanted intrusive thoughts (UITs) is crucial. The main aim of the current research was to investigate if individuals who used a counter app (a program on a mobile device that is used to count the frequency of an event by pressing the volume-up button) to assess UITs retrospectively overreported the number of UITs. The secondary aim was to establish preliminary psychometric qualities of the counter app method. A UIT was activated in $N = 87$ students. They were randomly allocated to one of three experimental conditions: counter app, thought monitoring, or free thinking. Retrospective descriptors of the UIT, including its frequency, were taken. The second study ($N = 118$) mainly aimed to replicate the results of the first study. In both studies, the retrospective frequency ratings of the UITs were 2 to 3 times higher in individuals who had used the counter app compared to those in the control conditions. Preliminary indicators of convergent validity and test–retest reliability were good; criterion, discriminant, and predictive validity were unsatisfactory. To conclude, using event marking such as a counter app can result in an overestimation of UITs. Alternative methods of assessment of UITs are discussed.

Keywords: counter app; experimental study; psychometric properties; unwanted intrusive thoughts

1. Introduction

The persistence of unwanted intrusive thoughts (UITs) is characteristic of many mental disorders and is particularly relevant in obsessive-compulsive disorder (OCD), where recurrent, persistent thoughts, images or impulses cause marked distress and typically result in compelling urges to undo the UIT or the potentially disastrous consequences (DSM 5; American Psychiatric Association, 2013). The occurrence of UITs is not restricted to OCD and mental disorders but seems to be a ubiquitous phenomenon; most people—regardless of ethnic or cultural background—experience UITs at least occasionally (Radomsky et al., 2014). UITs have been defined by Rachman (1981) as unwanted, repetitive thoughts, images or impulses that interrupt ongoing activity, are difficult to control and are attributed to an internal origin.

The reliable and valid assessment of UITs is crucial in the context of OCD-related research. Since UITs are internal, private cognitive phenomena and are often associated with feelings of shame and guilt, this can prove a difficult endeavour. Standardized questionnaires relying on retrospective verbal self-reports have been developed to assess the general tendency to experience UITs, and their psychometric qualities have been reported in previous work (e.g., Clark & Purdon, 1995). In experimental studies related to OCD, however, when the experimenter is interested in the frequency of UITs as they occur in real time, the assessment of UITs relies largely on methods whose psychometric properties are unknown (Clark & Purdon, 1995). Various methods have been used. In one common method, participants are instructed to indicate each occurrence of a target thought with a motor response such as pressing a button. The number of motor responses (button presses, bell rings, checkmarks, etc.) is assumed to reflect the number of UITs occurring during that period. Generally, this method is referred to as *event marking*, to distinguish it from other forms of thought assessment that do not require participants to actively report the occurrence of a target thought, such as thought sampling (e.g., Hirsch, Perman, Hayes, Eagleson, & Mathews, 2015)

or the think-aloud technique (Erskine, Georgiou, Joshi, Deans, & Colegate, 2017; Lavy & Vandenhout, 1990). During thought sampling, participants may think about whatever they want during a predefined period and are asked at several time points (e.g., 12 times in a 6-min period) to briefly indicate their thought content (i.e., whether they are currently thinking of the target thought). For the think-aloud technique, participants are asked to continuously say aloud whatever is going through their mind for a predefined period (e.g., 5 min). This stream-of-consciousness report is recorded and subsequently inspected by the experimenter, who evaluates the frequency of the target thoughts according to predetermined definitions.

Each method has advantages and disadvantages. Event marking has the advantage of easy implementation and analysis. Disadvantages are the potential for reactivity effects, which are discussed in detail below. Since participants are not required to monitor or actively report their thoughts during thought sampling or the think-aloud period, these methods might show fewer reactivity effects compared to event marking. On the other hand, during thought sampling the occurrence of target thoughts might be easily missed since the requests for reporting thought content happen intermittently with potentially large periods of time that are not covered. As a result, the frequency of UITs is likely to be underestimated. The main disadvantage of the think-aloud technique is that a person instructed to speak thought content aloud might be very vulnerable to effects of social desirability. Surprisingly, despite event marking being commonly used in experimental studies assessing UITs, no study to date has systematically investigated its psychometric qualities. The current study aimed to establish whether event marking such as using a counter app is a reliable and valid means of assessing UITs.

The psychometric qualities of event marking have been questioned for different reasons. The major criticism refers to a spurious increase in the number of reported UITs as a result of event marking (Lavy & Vandenhout, 1990; Salkovskis & Campbell, 1994). Signalling the occurrence of a target thought might cue further target thoughts, thus increasing

the number of reported UITs compared to situations when event marking is not requested (Lavy & Vandenhout, 1990; Salkovskis & Campbell, 1994). In a similar vein, Purdon and Clark (2000) argued that monitoring the occurrence of a particular thought increases its salience in consciousness, and marked salience in turn might result in an increase in the reported frequency of that thought. The process of thought monitoring during event marking might mimic elevated awareness of the (re-)occurrence of obsessive thoughts in individuals diagnosed with OCD (Abramowitz, Tolin, & Street, 2001; Trinder & Salkovskis, 1994), a phenomenon that has been described as heightened cognitive self-consciousness by other authors (de Bruin, Muris, & Rassin, 2007; Janeck, Calamari, Riemann, & Heffelfinger, 2003; Purdon, Gifford, McCabe, & Antony, 2011). The potential increase in reported UIT frequency through event marking has never been tested directly, nor the possible *extent* to which UIT frequency might be overreported.

Additionally, no study has experimentally differentiated whether it is cognitive processes involved in thought monitoring alone or rather cognitive processes *in conjunction with* further sensorimotor processes involved in event marking that results in a potential overreporting of UITs. Cognitive processes include the awareness or constant monitoring of mind content, the decision about whether any content matches the defined characteristics of the target thought and the motivation to indicate that a target thought has occurred. With sensorimotor processes we mean all processing of sensory information and motoric responses that result from holding a marking device and pressing a button. We consider it plausible that sensorimotor processes might further increase the occurrence of UITs over and above the cognitive processes involved in thought monitoring alone. For example, the auditory or tactile information gained from pressing a button on a smartphone during event marking might increase the awareness of UITs and/or trigger new UITs.

The second criticism of event marking concerns the confounding of frequency and duration of a thought (Markowitz & Purdon, 2008; Purdon, 2001). This means that one

particular thought count, for example, ‘1’, might represent a short UIT or an excessively long chain of intrusive thoughts, or an unsuccessful struggle to replace a UIT. The latter two might last for the whole duration of a predefined interval yet also be marked as ‘1’ with event marking. The chain of thoughts or the struggle might arguably be much more distressing than one discrete thought occurrence. For this reason, some authors have suggested that because frequency and duration are distinct aspects of UITs, they should be assessed separately (Lambert, Hu, Magee, Beadel, & Teachman, 2014). Obsessional thoughts have been described as a ‘state of mind’ rather than discernible thoughts (Reed, 1985), so in the context of clinical samples, the distinction between thought frequency and duration seems particularly relevant. A related question is whether additional descriptors of UITs such as the percentage of time a person spends thinking about a UIT (hereafter for brevity, percentage of time), its intensity and its intrusiveness provide additional information over and above the information provided by frequency and duration. No study to date has differentiated these aspects of UITs and investigated their intercorrelations. It is plausible to assume that the frequency obtained with a counter app correlates strongly with the frequency obtained by retrospective ratings, even though a compromised reliability of retrospective frequency reports (Schwarz, 2007) might limit this association. This correlation would be an indication of a counter app’s convergent validity. In comparison, associations between the frequency obtained with a counter app and the retrospectively assessed duration, percentage of time, intensity and intrusiveness of the UIT should be lower, and this would be an indication of the counter app’s discriminant validity.

Finally, social desirability might conceivably have a large impact on event marking. Event marking externalizes the occurrence of a private, sometimes unpleasant or unacceptable thought, and this observable externalization might be influenced by the presence of an experimenter to a larger degree than other, more subtle methods of thought assessment, such as retrospective ratings on the computer (Purdon & Clark, 2000). Social desirability could

influence the results in two ways: Participants might want to give the impression that they are very adaptive individuals and do not experience these kinds of thoughts, resulting in an *under*reporting of the number of thoughts. Alternatively, they might want to meet the ostensible expectations of the experimenter and fully comply with instructions, thus *over*reporting the number of UITs. To our knowledge, no study to date has examined whether event marking is associated with high levels of social desirability, which would be necessary to provide a thorough evaluation of various aspects of thought assessment as an appropriate means of assessing UITs.

Currently, there is a sharp increase in the development and use of treatment apps for mental disorders (Roncero, Belloch, & Doron, 2018; Van Ameringen, Turna, Khalesi, Pullia, & Patterson, 2017; Van den Bergh & Lehnen, 2019) and one might ask whether a counter app might be a useful feature for the assessment and/or treatment of OCD. For example, it might be used to estimate the frequency of obsessions in a specific threatening situation, or changes in the frequency of obsessions before and after exposure exercises during cognitive behavioural therapy. Thus the question of whether using a counter app method is an appropriate way to assess UITs might have direct clinical implications for new technical tools used in the treatment of OCD. In the current study, we addressed the prevailing critique of event marking by focussing on a counter app as a contemporary event marker of UITs in laboratory-based experimental studies (Wahl, van den Hout, & Lieb, 2019). The main purpose was to estimate if, and to what degree, the use of a counter app increases the number of retrospectively reported UITs occurring during a short interval in the laboratory, relative to the retrospectively reported UITs in two control conditions, that is, free thinking and thought monitoring. Free thinking was chosen since it controls for natural thinking processes and possesses high ecological validity (Lavy & Van den Hout, 1990). Monitoring of thoughts was chosen as an additional control condition to differentiate between the impact of heightened cognitive awareness, as occurs in thought monitoring, and the impact of heightened cognitive

awareness *in conjunction with* sensorimotor processes, as occurs in event marking with a counter app.

If it is the combination of cognitive awareness and the sensorimotor processes involved in event marking that accounts for a potential increase in UITs, then event marking should result in a higher frequency of reported UITs compared to thought monitoring and free thinking. If it is mainly the cognitive awareness (without sensorimotor processes) that accounts for the potential increase, then thought monitoring alone should result in a higher frequency of UITs compared to free thinking.

In line with previous authors, we expected that using a counter app would result in a higher frequency of retrospectively estimated UITs than thought monitoring and free thinking (Hypothesis 1). Since Salkovsis and Campbell (1994) and Purdon and Clark (2000) have argued that *monitoring alone* increases the salience of UITs, we additionally predicted that thought monitoring would result in a greater frequency of reported UITs than free thinking (Hypothesis 2). Given the argument that event marking increases the *salience* of the UIT (Purdon & Clark, 2000), it can conceivably be argued that not only frequency, but also the retrospectively rated duration, percentage of time, intensity and intrusiveness of the UIT would be increased in participants using the counter app compared to the two control groups. On the other hand, frequency and duration are not necessarily influenced in the same way by methods of thought assessment (Lambert et al., 2014). Given these conflicting presumptions, we decided not to specify any hypotheses for the additional UIT descriptions (duration, percentage of time, intensity and intrusiveness) but to analyse potential group differences in an exploratory way. Additionally, in line with arguments discussed above, we expected that participants using a counter app would experience a greater need to behave in a socially desirable way compared to participants using thought monitoring or free thinking (Hypothesis 3).

The hypotheses were tested in two independent studies with almost identical designs. In the first study, retrospectively rated frequency of the UIT was compared between individuals who were instructed to use a counter app, to monitor their thoughts or to think about whatever they wanted. The second study aimed to replicate the results of the first study, under conditions in which the thinking during the relevant interval was canalised by an additional focussed breathing task and was thus conceivably more standardized across conditions than in the first study. As a secondary aim of the study, we analysed the data of both studies to examine preliminary indicators of psychometric qualities of the counter app method.

2. Study 1

2.1. Methods

2.1.1. Participants

Undergraduate psychology students (N = 87) were recruited to participate in the present study (M_{age} = 24.83 years, SD = 7.89; 71.3% female) in exchange for course credit. Inclusion criterion was age ≥ 18 years. We excluded participants who had experienced a close family member or friend being involved in a car accident as well as participants who met the cut-off score of the Beck Depression Inventory, second edition (BDI-II; Schneider, Härter, & Schorr, 2017) to minimise the risk of causing great distress with the activation of the UIT.

2.1.2. Procedure

The study was approved by the Ethics Committee of the Faculty of Psychology, (blinded for review purposes). Figure 1 shows a detailed depiction of the study's procedure, which included two consecutive periods to allow us to assess the test–retest reliability of the counter app. Upon arrival, participants completed informed consent and self-report measures—(Beck Anxiety Inventory (BAI); Beck, Brown, Epstein, & Steer, 1988; Beck Depression Inventory, second edition (BDI-II); Beck, Steer, & Brown, 1996; Obsessive-Compulsive Inventory, revised (OCI-R); Foa et al., 2002; Positive and Negative Affect

Schedule (PANAS); Watson, Clark, & Tellegen, 1988; White Bear Suppression Inventory (WBSI); Wegner & Zanakos, 1994)—and rated their distress on a visual analogue scale (VAS), using the online tool Limesurvey (Questback GmbH, 2013). This was followed by the activation of the UIT in Phase 1. Immediately after the activation of the UIT, participants completed VAS ratings of distress again. Subsequently, participants were provided with one of the three instructions of the first thought phase (counter app, thought monitoring, or free thinking), depending on random allocation to groups. This first thought phase was followed by the retrospective ratings of the thought descriptors (frequency and duration of the UITs, the percentage of time, intensity and intrusiveness). Next, a short movie about travelling around the world (3 min), which was shown to reduce carry-over effects for the subsequent measurements, concluded Phase 1 of the study. Phase 2 was identical to Phase 1, with the exception that the UIT was *re*activated, using an identical procedure. At the end, participants completed measures of the frequency of mental and behavioural neutralizing, the Revised Obsessive Intrusions Inventory (ROII; Purdon & Clark, 1994) and two questions assessing social desirability.

2.1.3. Tasks

2.1.3.1. Activation and reactivation of the UIT

The UIT was activated with a paradigm that has been used in various previous studies (e.g., Rachman, Shafran, Mitchell, Trant, & Teachman, 1996; van den Hout, Kindt, Weiland, & Peters, 2002; van den Hout, van Pol, & Peters, 2001). After being provided with a pen and paper, participants were asked to think of a loved one and to get a clear picture of this person in their mind. Once they had a clear picture in mind, they were asked to write down and read aloud the following sentence: 'I wish that person x would die in a horrible car accident this evening', replacing 'person x' with the beloved person's name. Participants were subsequently told that any thoughts, images or impulses related to the beloved person dying in a car accident were going to be referred to as the target UIT in the following parts of the

study. The reactivation of the UIT in Phase 2 of the study was identical to the activation in Phase 1, except that the experimenter asked participants to think *again* of the beloved person that they were thinking of previously, and to get *again* a clear picture of that person in their mind.

2.1.3.2. Thought phase (experimental manipulation)

Participants were randomly allocated to one of three experimental groups (counter app, thought monitoring or free thinking). In the counter app group, participants were provided with a smartphone and the following instructions on a computer screen:

> During the next 5 min, you may think about anything you like. You might think of your target unwanted intrusive thought, but you do not have to. However, if at any time you think of your target unwanted intrusive thought, please record the occurrence of each thought by pressing the "+" key on the smartphone once for each occurrence. It is important that you continue in the same way for the full 5 min. (adapted from Marcks & Woods, 2005)

In the thought-monitoring group, participants were provided with the following instructions:

> During the next 5 min, you may think about anything you like. You might think of your target unwanted intrusive thought, but you do not have to. During these 5 min, please monitor your thoughts very carefully. In particular, note in your mind whether you experience any target unwanted intrusive thought. It is important that you continue in the same way for the full 5 min.

In the free-thinking group, participants were provided with the following instructions: 'During the next 5 min, you may think about anything you like'.

2.1.3.3. Counter app frequency

We used a counter app (FunCoolApps, 2016) on a smartphone to assess the number of UIT occurrences. Participants were asked to press the volume-up button whenever the target

thought occurred. The display with the number of counts was covered to prevent visual feedback on the number of counts. No other applications were visible or usable for the duration of the study.

2.1.4. Standardized measures, assessment of UITs, thought descriptives, social desirability and neutralizing

2.1.4.1. Standardized questionnaires

The (BDI-II; Beck et al., 1996; German version: Hautzinger, Keller, & Kühner, 2006) assesses depressive symptoms with 21 items. The reliability and validity of the scale are well established (Wang & Gorenstein, 2013).

The (BAI; Beck et al., 1988; German version: Magraf & Ehlers, 2007) is a 21-item self-report measure of anxiety symptoms. The psychometric properties of the scale are excellent (Beck et al., 1988; Magraf & Ehlers, 2007).

The (OCI-R; Foa et al., 2002; German version: Goenner, Leonhart, & Ecker, 2007) is a well-established 18-item measure of OC symptom severity, focussing on six symptom domains: washing, obsessions, hoarding, ordering, neutralizing and checking. The total score has good psychometric properties (Goenner et al., 2007; Goenner, Leonhart, & Ecker, 2008).

The WBSI (Wegner & Zanakos, 1994) was originally designed to assess the tendency to suppress UITs using 15 items. The scale has demonstrated good reliability (Höping & de Jong-Meyer, 2003; Muris, Merckelbach, & Horselenberg, 1996) but poor validity, since recent studies demonstrated that it also measures the tendency to experience intrusive thoughts (Blumberg, 2000; Höping & de Jong-Meyer, 2003).

The (PANAS; Krohne, Egloff, Kohlmann, & Tausch, 1996; Watson et al., 1988) assesses current positive (10 items) and negative (10 items) affect. Reliability and validity are well established (Crawford & Henry, 2004; Krohne et al., 1996).

The (ROII; Purdon & Clark, 1994) was designed to assess the occurrence, frequency and impact of UITs. The ROII Part 1 contains 52 items to assess the frequency of UITs

related to aggression, sex, dirt and contamination. Eight of the original 10 items of the ROII Part 2 were used to measure the appraisals of the activated UIT (unpleasantness, guilt, worry the UIT will come true, unacceptability, likelihood the UIT will come true, importance to control, harm/danger and responsibility).

2.1.4.2. Additional self-rated measures

Participants rated their distress on a VAS of 0 (*extremely low*) to 9 (*extremely high*). A retrospective rating of thought descriptors was used to assess the frequency, duration, percentage of time, intensity and intrusiveness of UITs immediately after the thought phase:

1. 'Please estimate (as accurately as possible) how many times thoughts about the car accident came to your mind during the last 5 minutes'. Participants had to provide the estimated number of target thoughts as an answer.
2. 'Please indicate how long the target thoughts lasted on average'. Answers were provided on a VAS of 0 (*only very briefly*) to 100 (*for an extremely long time*).
3. 'Please indicate what percentage of the time you were thinking of the target thoughts during the last 5 minutes'. Answers were provided as percentages, from 0% (*no time at all*) to 100% (*100% of the time*).
4. 'Please indicate the degree of intensity of your target thoughts during the last 5 minutes'. Answers were given on a VAS of 0 (*not at all*) to 100 (*extreme*).
5. 'Please indicate the degree of intrusiveness of your target thoughts during the last 5 minutes. Intrusiveness means that the target thoughts intruded into your mind with great force, persisted for a long time, occurred frequently and possibly interrupted what you were thinking at that time'. Answers were given on a VAS of 0 (*not at all*) to 100 (*extreme*).

The following two items were used to assess two aspects of social desirability:

1. 'To what extent were your frequency ratings of the target thought influenced by the experimenter being present, for example, because you felt you were being

observed or evaluated?' Answers were provided on a scale of 0 (*not at all influenced by presence of experimenter*) to 4 (*extremely influenced by presence of experimenter*).

2. 'Given the presence of the experimenter, did you feel that you wanted to make a certain impression?' Answers were provided on a scale of 0 (*not at all*) to 4 (*very much so*).

At the end of the study, 11 items were provided to assess different kinds of covert neutralizing and 14 items to assess different kinds of overt neutralizing during the course of the experiment. Items included modifications from previous work on neutralizing (Freeston, Ladouceur, Thibodeau, & Gagnon, 1991; Goodman et al., 1989; Purdon & Clark, 1993, 1994; Rachman et al., 1996) and were extended with additional items reflecting neutralizing strategies found in previous studies by our group. A detailed description of the assessment is provided by Kollárik, et al. (2020). Participants had to indicate for each covert and overt strategy whether they had used it to neutralize the activated UIT at any point during the study (e.g., 'Did you reassure yourself?' 'Did you say a prayer?' 'Did you knock on wood?'). In addition, they had to indicate whether they anticipated performing a neutralizing behaviour once the study was finished (e.g., 'I will talk to the person named in the sentence to reassure myself that nothing terrible has happened'). Finally, participants were provided with the sheet of paper on which they had previously written the car accident sentence and were given the opportunity to perform any overt neutralizing that they felt they needed to do (e.g., to rip the paper or to change the name in the sentence). The experimenter rated whether participants performed any neutralizing behaviour as either 1 (present) or 0 (absent).

For this study, covert neutralizing was operationalized as the sum of all different covert strategies used to neutralize the UIT during the course of the study. Overt neutralizing was operationalized as the sum of all overt strategies performed at some point during the study and any anticipated overt neutralizing behaviour.

2.1.5. Statistical analysis

All raw data were visually checked for outliers, using boxplots, prior to analysis. Two outliers in the counter app frequencies in Phase 1 were adjusted such that the outlier was assigned a value one unit above the next highest score (Winsorizing). For better readability, the statistical analysis and expected effects are reported in each Results section rather than here. Alpha was set at <.05 for each hypothesis. For the exploratory analysis of multiple variables (the retrospective descriptor ratings with the exception of frequency), we controlled alpha inflation by using multivariate general linear model (GLM) analysis. As effect sizes we report η_p^2 for GLM analyses, *d* for planned contrasts and *t* tests and *r* for Pearson correlations.

2.2. Results of Study 1

Manipulation checks including differences between experimental groups in participant characteristics and successful UIT activation are presented in Supplement A.

2.2.1. Group differences in frequencies and other retrospective descriptors of the UIT (Hypotheses 1 and 2)

For the retrospective frequency ratings, the main hypothesis was that using a counter app would result in higher retrospective frequency ratings than the two control techniques. Since we tested this hypothesis in two consecutive phases, we first used a GLM analysis in a 2 × 3 mixed-model design with time (Phase 1 vs. Phase 2) as a within-subject factor and experimental group (counter app, thought monitoring, free thinking) as a between-subjects factor to account for potential phase effects, such as habituation, practise or carry-over effects. These would be reflected in a significant main effect of phase or a significant interaction of phase and group. This analysis was followed by planned contrasts comparing the retrospective ratings of the counter app condition with those of the two control conditions, separately for each phase, to specifically test Hypothesis 1. Means and standard deviations of the retrospective frequency ratings are shown in Table 1; Figure 2 depicts the results graphically. The GLM resulted in a significant effect of phase, $F(1, 84) = 27.47$, $p < .001$, η_p^2

= .25, reflecting an overall decrease in frequency of the UIT from Phase 1 to Phase 2. The main effect of group was not significant, $F(2, 84) = 2.10$, $p = .129$, $\eta_p^2 = .05$. However, a significant interaction of phase and group indicated that group differences varied with the phases, $F(2, 84) = 3.57$, $p = .033$, $\eta_p^2 = .08$.[1]

Planned contrasts show that in Phase 1, retrospectively rated frequency of UITs was higher in the counter app condition compared to both control conditions, with a medium effect size, $t(84) = 2.08$, $p = .04$, $d = 0.45$. In Phase 2, the differences in frequencies of UITs between the counter app condition and the two control conditions, with a small effect size, did not reach significance, $t(84) = 1.25$, $p = .21$, $d = 0.27$.

To test Hypothesis 2, independent-samples *t* tests were used to compare thought monitoring with free thinking, separately for each phase. Thought monitoring alone did not result in higher frequency of UITs than free thinking, neither in Phase 1, $t(59) = -0.91$, $p = .366$, $d = 0.24$, nor in Phase 2, $t(59) = -0.96$, $p = .342$, $d = 0.25$ (Hypothesis 2). Effect sizes for differences between thought monitoring and free thinking were small.

Means and standard deviations of the remaining retrospective ratings (duration, percentage of time, intensity and intrusiveness) are shown in Table 1. Since we were interested in the question of whether there were group differences on *any* of these retrospective ratings at this early exploratory stage, all variables were analysed using a multivariate mixed-model design with phase (Phase 1 vs. Phase 2) as a within-subject factor and experimental group (counter app, monitoring, free thinking) as a between-subjects factor. The multivariate results demonstrate a significant effect of phase, Pillai's trace, $V = .377$, $F(4, 82) = 12.24$, $p < .001$, $\eta_p^2 = .38$. Neither group, Pillai's trace, $V = .09$, $F(8, 164) = .99$, $p = .47$, $\eta_p^2 = .05$, nor the interaction of phase and group, Pillai's trace, $V = .13$, $F(8, 164) = 12.24$, $p = .17$, $\eta_p^2 = .06$, were significant.

[1] Since experimental groups differed on OC symptoms (OCI-R) at baseline, we reran the GLM analysis with OCI-R as a covariate, which resulted in an identical pattern of results.

2.2.2. Group differences in social desirability (Hypothesis 3)

Means and standard deviations for social desirability are shown in Table 1. A planned contrast comparing social desirability in the counter app condition with the two control conditions was used to test Hypothesis 3. Neither the tendency to be influenced by the experimenter, $t(84) = -0.01$, $p = .996$, $d < 0.001$, nor the intention to make a certain impression, $t(84) = -0.84$, $p = .404$, $d = -0.20$, was greater in the counter app condition than in the two control conditions.

2.2.3. Preliminary indicators of validity and reliability of counter app assessment

Means and standard deviations of the counter app frequencies during the 5-min period are shown in Table 1. Associations between the counter app frequency and the retrospective thought descriptors (frequency, duration, percentage of time, intensity and intrusiveness) were used to determine convergent and divergent validity. High positive correlations between two different assessments of the same construct (frequency assessed with the counter app and the retrospective ratings) would indicate convergent validity, while low positive correlations between two different constructs (counter app frequency and retrospective duration, percentage, intensity and intrusiveness, respectively) would indicate discriminant validity.

UITs are typically associated with concurrent distress and subsequent neutralizing, that is, deliberate overt or covert actions aiming to undo the UIT or the distress that is associated with it (Bocci & Gordon, 2007). The correlation between counter app frequency and distress was used to determine criterion validity. The functional relationship between UITs and neutralizing suggests that the degree of neutralizing depends on characteristics of the UIT, such as its frequency and the distress associated with it, and previous studies support this idea (Bocci & Gordon, 2007; Rupp et al., 2019). Predictive validity of the counter app would thus be reflected in positive associations with subsequent neutralizing behaviour. A high correlation between the counter app frequencies in Phase 1 and Phase 2 would indicate

test–retest reliability. Since indicators of validity and reliability were defined a priori, we did not adjust for multiple testing. However, results should be considered as first indications.

Intercorrelations of retrospective thought descriptors and correlations of thought descriptors with distress and neutralizing are shown in Table 2. The product-moment correlation between frequency assessed with the counter app and frequency assessed retrospectively (convergent validity) was $r = .64$ in Phase 1 and $r = .41$ in Phase 2. The correlations between the frequency assessed with the counter app and descriptors of the UIT assessed retrospectively (discriminant validity) ranged from $r = .25$ between counter app frequency and duration of UIT to $r = .53$ between counter app frequency and intrusiveness in Phase 1. In Phase 2, correlations between counter app frequency and the retrospective ratings were generally lower, ranging from $r = .04$ for correlation between counter app frequency and intrusiveness and $r = .22$ between counter app frequency and duration of UIT. Counter app frequency was not related to distress (criterion validity). Correlations between counter app frequency and subsequent covert and overt forms of neutralizing (predictive validity) were zero to small and nonsignificant. The correlation between counter app frequency in Phase 1 and counter app frequency in Phase 2 (test–retest reliability) was $r = .64$. The remaining retrospective variables were correlated to a medium to large degree in both phases, based on Cohen's (1992) definitions (small ($r = .10$), medium ($r = .30$), large ($r = .50$)).

2.3. Discussion of Study 1

Participants who used a counter app reported a higher frequency of UITs than those who were instructed to monitor the occurrence of the UIT and those who were instructed to think freely in the first phase of the study, but not in the second. The predicted group differences were observable in both phases and reached significance only when the effect was of a medium size in Phase 1. Thus Hypothesis 1 was partially supported. The decrease in the overestimation of the UIT frequencies in Phase 2 can be explained by practise effects. In Phase 1, participants who used the counter app estimated the number of UITs to be 2 to 3

times higher than the number estimated by participants who thought about whatever they wanted. The overestimation is unlikely to be accounted for by cognitive processes alone, since thought monitoring did not result in a higher frequency of UITs than free thinking (Hypothesis 2). The impact of event marking seems to be specific to frequency, rather than affecting other aspects of UITs, such as duration, percentage of time and intensity. Additionally, we observed an overall decrease from Phase 1 to Phase 2 in UIT frequency.

There is no indication that aspects of social desirability played a larger role when UITs were assessed with a counter app compared to thought monitoring and free thinking. Thus, Hypothesis 3 was not supported.

Regarding the psychometric properties of the counter app, the pattern of correlations provides a first indication of satisfactory convergent validity and acceptable test–retest reliability of the counter app as a measurement of UIT frequency. Discriminant validity was low in Phase 1 but seems to increase with practise. Criterion and predictive validity were unsatisfactory.

The study provides a stringent test of differences between the experimental groups. However, strictly speaking, it remains unclear whether the group differences can be accounted for by overestimation of the true occurrence of UITs, as was suggested by Salkovskis and Campbell (1994) and Purdon (2020), or by underestimation of thoughts in the control groups. There is some indication that the activation of a UIT in healthy participants results in a spontaneous effort to suppress the UIT and also to take other measures to deal with the associated emotional distress (Purdon & Clark, 2001; Salkovskis & Reynolds, 1994). It can be argued that participants in the free-thinking and possibly also in the thought-monitoring condition had more mental capacity to suppress or neutralize the UIT compared to those in the counter app condition, resulting in a lower occurrence and report of UITs.

3. Study 2

In the second study, we therefore aimed to improve the internal validity of the research design by reducing the number of alternative interpretations of the group differences. All participants were asked to perform a simultaneous focussed-breathing task, thereby limiting the possibility of spontaneous mental countermeasures in all groups to the same degree. As a second and minor modification, we asked participants to evaluate distress directly before and after the thought phases in which the UIT was assessed. This allowed us to test the effects of using a counter app on distress. Because a higher number of UITs would conceivably result in higher distress, we expected that participants using a counter app would experience more distress than the control groups. We tested all previous hypotheses and the following additional hypothesis in Study 2: Using a counter app would result in higher distress than thought monitoring and free thinking combined (Hypothesis 4).

3.1. Methods

3.1.1. Participants

Undergraduate psychology students at the University of (blinded for review purposes) (N = 118; M_{age} = 26.14 years, SD = 9.98; 77.1% female) participated in the study for course credit. Inclusion criterion was age ≥ 18 years. Exclusion criteria were a BDI-II score of ≥ 13 (Schneider et al., 2017) and the experience of a close family member or friend being involved in a car accident. The study was approved by the Ethics Committee of (blinded for review purposes).

3.1.2. Procedure

All participants completed informed consent prior to the beginning of the study. Figure 3 shows the details of the study's procedure, which was identical to that of Study 1 with two exceptions: A focussed breathing task was practised at the beginning of the experimental session and integrated into the instructions of the thought phase (see Sections 3.1.3.1 and 3.1.3.2). The VAS distress rating was assessed immediately before and after the thought phases in order to directly test the impact of the experimental group on distress.

3.1.3. Tasks

Activation and reactivation of the UIT were identical to in Study 1.

3.1.3.1. Focussed breathing

A focussed breathing task was introduced in Study 2 to limit the possibility of spontaneous mental countermeasures in all groups to the same degree. At the beginning of the experimental session, participants were asked to practise focussing their attention on their breathing. The instructions were a slightly modified version of the instructions used by Hirsch et al. (2015):

> To start with, we are going to practice focussing your attention on your breathing for a short time. Try to focus your attention on your breathing. It is completely normal that your thoughts wander, but if this happens, try to refocus your attention back on your breathing again.

Participants then practiced focussed breathing for 20 s.

3.1.3.2. Thought phases (experimental manipulation)

Participants were randomly allocated to one of three experimental groups (counter app, thought monitoring or free thinking). Instructions in the respective groups were identical to those in Study 1, except that each instruction was preceded by asking participants to focus on their breathing, in the same way that they had practised earlier in the experimental session. In particular, participants were told: 'As before, I would like you to focus on your breathing'. For example, in the counter app group, the complete instructions were as follows:

> As before, I would like you to focus on your breathing. During the next 5 min, you may think about anything you like. You might think of your target unwanted intrusive thought, but you do not have to. However, if at any time you think of your target unwanted intrusive thought, please record the occurrence of each thought by pressing the "+" key on the smartphone once for each occurrence. It is important that you continue in the same way for the full 5 min. (adapted from Marcks & Woods, 2005)

3.1.4. Measures and assessment of UITs and neutralizing

The standardized questionnaires (BDI-II, BAI, OCI-R, WBSI), VAS ratings of distress, the retrospective questionnaire, ratings of social desirability, assessment of covert and overt neutralizing and assessment of the counter app frequency were identical to those in Study 1. Additionally, the Thought-Action Fusion Scale (German version: Jana Hansmeier, Cornelia Exner, Winfried Rief, & Julia Glombiewski, 2014; TAFS; Shafran, Thordarson, & Rachman, 1996) was used to assess thought–action fusion bias. The questionnaire consists of two subscales measuring thought–action fusion with regard to morality (TAF-Moral, 12 items) and probability (TAF-Likelihood, 7 items). Reliability and validity are acceptable (Hansmeier et al., 2014; Meyer & Brown, 2013; Rassin, Merckelbach, Muris, & Schmidt, 2001).

3.1.5. Statistical analysis

No outliers were detected on any of the variables using visual inspection of boxplots. All statistical analyses and procedures were identical to those in Study 1. Study 1 resulted in a medium-sized effect ($d = 0.45$) for the main research question (i.e., a planned contrast for the difference between the retrospective frequency ratings of the UIT in the counter app group compared to both control groups). An a priori power analysis for Study 2 with alpha = .05, beta = .80, an anticipated medium-sized effect of $d = 0.45$ and an n_2/n_1 ratio of 2/1 resulted in a total sample size of $N = 114$.

3.2. Results of Study 2

Manipulation checks including differences between experimental groups in participant characteristics and successful UIT activation are presented in Supplement B.

3.2.1. Group differences in frequencies and other retrospective descriptors of the UIT (Hypotheses 1 and 2)

Means and standard deviations of the retrospective frequency ratings are shown in Table 3. Figure 4 shows the results graphically. A significant effect of phase, $F(1, 115) =$

17.26, $p < .001$, $\eta_p^2 = .13$, reflects a decrease in the retrospective frequencies of the UITs from Phase 1 to Phase 2. The main effect of group was significant, $F(2, 115) = 3.32$, $p = .040$, $\eta_p^2 = .06$, and the interaction of group and phase was not significant, $F(2, 115) = 0.07$, $p = .936$, $\eta_p^2 = .00$, indicating that group differences in frequency ratings did not differ across the two phases.[2] Subsequent planned contrasts show that in both phases, the difference between the counter app condition and the two control conditions was significant: in Phase 1, $t(115) = 2.06$, $p = .042$, $d = 0.41$; in Phase 2, $t(115) = 2.86$, $p = .005$, $d = 0.56$, with medium effect sizes.

Thought monitoring alone did not result in higher retrospective frequency of UITs compared to free thinking, neither in Phase 1, $t(78) = 0.20$, $p = .840$, $d = 0.05$, nor in Phase 2, $t(78) = 0.79$, $p = .432$, $d = 0.18$. Effect sizes were very small.

Means and standard deviations of the remaining retrospective ratings (duration, percentage of time, intensity and intrusiveness) are shown in Table 3. The multivariate results demonstrate a significant effect of phase, Pillai's trace, $V = .27$, $F(4, 112) = 10.46$, $p < .001$, $\eta_p^2 = .27$, and a significant interaction of phase and group, Pillai's trace, $V = .16$, $F(8, 226) = 2.40$, $p = .017$, $\eta_p^2 = .08$. The main effect of group was not significant, Pillai's trace, $V = .03$, $F(8, 226) = .41$, $p = .913$, $\eta_p^2 = .01$.

3.2.2. Group differences in social desirability (Hypothesis 3)

Table 3 shows means and standard deviations for social desirability. Neither the tendency to be influenced by the experimenter, $t(115) = -0.429$, $p = .669$, $d = -0.08$, nor the intention to make a certain impression, $t(115) = -0.933$, $p = .353$, $d = -0.18$, was greater in the counter app group than in the two control groups.

3.2.3. Group differences in distress (Hypothesis 4)

[2] Since the experimental groups differed on negative affect at baseline, we reran the GLM analysis with negative affect as a covariate. The main effect phase was not significant, $F(1, 114) = 0.47$, $p = .493$, nor the interaction of phase and group, $F(2, 114) = 0.15$, $p = .864$. The main effect of group was significant, $F(2, 114) = 4.00$, $p = .021$. The covariate negative affect did not interact significantly with phase, $F(1, 114) = 3.70$, $p = .057$.

Means and standard deviations of distress pre- and postexperimental manipulation are displayed in Table 3, separately for each phase. To examine potential phase effects, distress was first analysed using GLM analysis with a 2 × 2 × 3 mixed-model design, with time (pre vs. post experimental manipulation) and phase (Phase 1 vs. Phase 2) as within-subject factors and experimental group (counter app vs. thought monitoring vs. free thinking) as a between-subjects factor. GLM analysis revealed significant main effects of phase, $F(1, 115) = 62.02$, $p < .001$, $\eta_p^2 = .35$, and time, $F(1, 115) = 106.93$, $p < .001$, $\eta_p^2 = .48$, which were qualified by a significant interaction of phase and time, $F(1, 115) = 17.56$, $p < .001$, $\eta_p^2 = .13$. The interaction reflects a larger reduction from pre- to postexperimental manipulation in Phase 1 compared to Phase 2. Neither the main effect of group, $F(2, 115) = 1.51$, $p = .224$, $\eta_p^2 = .03$, nor any interactions with group were significant: Group × Phase, $F(2, 115) = 1.51$, $p = .262$, $\eta_p^2 = .02$; Group × Time, $F(2, 115) = 2.67$, $p = .07$, $\eta_p^2 = .04$; Group × Time × Phase, $F(2, 115) = 1.90$, $p = .153$, $\eta_p^2 = .03$. Since the main effect of group and its interaction with time were not significant, we did not perform any follow-up analyses such as contrasts comparing the counter app group with the two control conditions.

3.2.4. Secondary aim: Preliminary indicators of validity and reliability of counter app assessment

Table 4 shows the intercorrelations of retrospective thought descriptors and correlations of retrospective thought descriptors with neutralizing. For details of how the different indicators of the psychometric properties were determined, please refer to Section 2.2.3. The product-moment correlation between frequency assessed with the counter app and frequency assessed retrospectively (convergent validity) was excellent in both phases, $r = .86$ in Phase 1 and $r = .90$ in Phase 2. Correlations of frequency assessed with the counter app with duration of UIT, percentage of time, intensity and intrusiveness (discriminant validity) ranged from $r = .45$ (intrusiveness) to $r = .76$ (percentage of time) in Phase 1 and from $r = .37$ (intensity) to $r = .61$ (percentage of time) in Phase 2. The correlation of frequency with the

counter app and distress after thought activation (criterion validity) was small and reached significance only in Phase 1. Frequency of the counter app was correlated with subsequent overt neutralizing to a small to medium degree; the small correlation with subsequent covert neutralizing did not reach significance (predictive validity). The remaining retrospective variables were intercorrelated and also correlated with distress to a medium to large degree in both phases.

3.2.5. Additional analysis: The persistence of UITs

All indicators of the persistence of the UIT at the end of the study, that is, distress, the retrospective thought descriptors and neutralizing at the end of the study, showed considerable variance. Out of interest, we examined whether the persistence of the UIT was associated with the trait measures in an additional exploratory regression analysis. As the outcome variable, we computed a persistence score. We argue that persistence would comprise the following aspects: (a) The UIT would still be present, which would be reflected in high ratings of any of the retrospective thought descriptors. (b) The UIT would be experienced as distressing, which (c) would result in neutralizing behaviour. The decision on which of the thought descriptors would be included was based on an examination of the correlations with covert and overt neutralizing, as an indication of criterion validity. UIT intensity showed the highest correlations. Thus the new variable persistence was computed as the sum of retrospectively rated intensity in Phase 2, distress in Phase 2, and the sum of all covert and overt neutralizing strategies at the end of the study. Internal consistency (Cronbach's alpha) of this persistence score was low (r = .58). Since item statistics suggested that internal consistency would improve if overt and covert neutralizing were deleted, we recomputed persistence as the sum of distress and retrospectively rated intensity, with good reliability (Cronbach's alpha = .78).[3] As predictors, we examined the intercorrelations of the

[3] We performed regression analyses with both variants of persistence, with almost identical results. R^2 was slightly higher when persistence consisted of four variables instead of two.

trait measures and decided to enter depressive symptoms (BDI-II), OC symptoms (OCI-R), tendency to suppress thoughts (WBSI) and thought–action fusion (TAFS), as they were moderately correlated (r = .19 to r = .51). We excluded the BAI from the list of predictors since it showed a high conceptual overlap with the BDI (r = .71). Multicollinearity was not a problem; all tolerance indices were >.25 and the variance inflation factors were far below 5, meeting the recommendations of Urban and Mayerl (2006). All predictors were entered simultaneously. Table 5 shows the regression coefficients (*B*s), the standardized coefficients (*Betas*), standard errors and p values. Only thought–action fusion predicted the persistence of the UIT, with an adjusted R^2 = .12.

3.3. Discussion of Study 2

Participants who had used the counter app to record their UITs retrospectively reported experiencing almost twice as many UITs compared to individuals who had not used a counter app. Thus, Hypothesis 1 was supported in both phases of Study 2. Thought monitoring did not result in more UITs compared to free thinking, neither in Phase 1 nor in Phase 2. Hypothesis 2 was thus not supported. The second study strongly suggests that event marking, but not thought monitoring alone, leads to an overestimation of UITs.

Although the overall number of UITs seemed lower for the counter app group in Study 2 compared to Study 1, adding a focussed breathing task to the thought instructions did not alter the overall pattern of results. This means that the difference in the frequencies of UITs between the counter app condition and the control conditions is unlikely to be accounted for by an underestimation of UITs in participants who did not use a counter app but rather is likely an overestimation of UITs by those who used the counter app. A significant interaction between phase and group in the multivariate analysis of the remaining thought descriptors (duration, percentage of time, intensity and overall intrusiveness) suggests that using a counter app might also have affected the retrospective ratings of other characteristics of the UIT. Future studies should address this research question directly, using a study design that

focusses on various UIT descriptors. Using a counter app did not result in a higher degree of socially desirable behaviour compared to the control groups. Thus Hypothesis 3 was not supported.Distress decreased from Phase 1 to Phase 2, and also during the respective thought phases, to the same degree in each experimental group. Hypothesis 4, stating that using a counter app might result in a higher degree of distress, was not supported. Since the counter app group reported retrospectively more UITs than controls, this finding could mean that distress is not a function of the frequency of UITs. This view is supported by a small correlation between the counter app frequency and distress after thought activation. Alternatively, the finding could mean that the increase in retrospectively reported UITs caused by the counter app is indeed artificial and reflects a report bias, rather than a real increase in the occurrence of UITs.

The counter app method had excellent convergent validity in Study 2. Indications of discriminant validity were mixed, showing considerable overlap with percentage of time and both duration and intrusiveness of the UIT. Criterion and predictive validity were low. The test–retest reliability of the counter app was high. Adding a task that canalises the stream of consciousness seems to have resulted in higher convergent validity compared to Study 1.

The additional exploratory analysis regarding trait predictors of the persistence of the UIT shows that only a tendency to misinterpret the occurrence of UITs according to a thought–action fusion bias, and not a tendency to suppress UITs, depressive symptoms or OC symptoms, predicted whether the UIT would last during the experimental study. This means that individuals who believed that experiencing the car accident thought rendered the event more likely to happen or that experiencing the car accident thought was morally equivalent to *wanting* the event to happen were more likely to experience a distressing and intense UIT at the end of the study. Thus the belief that intrusive thoughts have moral and actual consequences might not only be associated with symptoms of OCD (e.g., Bailey, Wu,

Valentiner, & McGrath, 2014) but might also be involved in the short-term persistence of UITs.

4. General discussion

Our findings indicate that event marking increases the number of reported UITs by a factor of 2 to 3, and this method cannot be recommended for experimental studies in which the frequency of naturally occurring UITs is of primary interest. The overestimation seems to be accounted for by thought monitoring processes *in conjunction with* additional sensorimotor processes involved in event marking. Therefore our findings do not support the argument that mainly the cognitive processes of event marking increase the number of reported UIT occurrences (Purdon & Clark, 2002; Salkovskis & Campbell, 1994). However, rejecting this view might be premature, since it could be a question of degree. Differences between thought monitoring and free thinking were in the predicted direction and of potential practical relevance for experimental studies in which the absolute number of UITs is of interest. Replication studies with sufficient power to account for possibly small effect sizes should follow.

Whether these findings have implications for the assumption that heightened self-consciousness in OCD is a mechanism that eventually results in an increase in obsessive thoughts cannot be answered without a replication study in individuals diagnosed with OCD. The naturally occurring thought-monitoring processes in OCD might differ from the thought-monitoring instructions of the study (e.g., the consequences of not identifying thought content correctly might be perceived as more severe).

Study 2 mainly served to replicate the results of Study 1 in a more rigorous study design that reduced alternative explanations for the group differences. Introducing a focussed breathing task, thereby limiting the possibility of spontaneous mental countermeasures in all groups to the same degree, reduced the overall number of UITs, which is consistent with previous studies (Salkovskis & Reynolds, 1994), but did not change the overall pattern of

results. This consistency of findings in conditions of more standardized thinking further fosters our confidence in the argument that using a counter app results in an *over*reporting of UITs, rather than the control groups *under*reporting UITs.

Whether using a counter app exclusively affected retrospective frequency ratings or also other retrospective thought descriptors (duration, percentage of time, intensity or frequency) cannot be answered unambiguously, since the two studies resulted in contrary findings. Therefore various thought descriptors should be addressed more directly in future studies. All thought descriptors were moderately to highly intercorrelated. Some descriptors shared more conceptual overlap than others. For example, UIT intensity and intrusiveness were highly intercorrelated, whereas frequency and duration were only moderately correlated. Thus UIT frequency and duration should be seen as different but conceptually overlapping dimensions, which is only partially consistent with the view of Lambert et al. (Lambert et al., 2014). Different degrees of associations between the thought descriptors and distress or subsequent neutralizing suggest that the target descriptor of UITs in future experimental studies should be carefully considered. For example, UIT intensity and intrusiveness demonstrated the highest associations with distress and neutralizing and might thus be particularly relevant if emotional or behavioural aspects of the UITs are investigated in a clinical context. Since our findings suggest that using a counter app might result in strong reactivity effects, using a counter app in the context of new technologies in the treatment of OCD cannot be recommended. Instead, an app feature that asks individuals to retrospectively estimate the intrusiveness of obsessions might be less reactive and also clinically more valid in terms of the associated distress and the subsequent neutralizing behaviour.

As an interesting side effect, the study design allowed us to examine which traits predicted the persistence of UITs across both experimental phases. Only the belief that UITs have moral and actual consequences, that is, a thought–action fusion bias, and not depressive or OC symptomatology or the tendency to suppress UITs was associated with distressing and

intense UITs at the end of the study, a finding that is consistent with current cognitive-behavioural models of OCD (Clark & Purdon, 2016; Salkovskis & Millar, 2016).

There is no indication that aspects of social desirability played a larger role when UITs were assessed with a counter app compared to thought monitoring or free thinking, in either study. Thus Hypothesis 3 was not supported.

The following limitations are important for the interpretation of the results. The retrospective ratings of thought descriptors have high face validity, but the method is not validated in the strict sense of rigorous validation studies. Thus for a more exhaustive investigation of the validity of event marking, a variety of other UIT assessment methods such as thought sampling or the think-aloud technique should be considered in future studies. We used a contemporary form of event marking (i.e., a counter app), and means and standard deviations were comparable to those found in previous studies using a similar UIT activation and event marking (Grisham & Williams, 2009; Lin & Wicker, 2007).Whether results can be generalized to other forms of event marking such as key pressing or checkmarking is unknown. Using a counter app is a relatively covert form of event marking and it could be that other more overt forms of event marking, for example, bell ringing or verbal indications, result in an even greater overestimation of UITs.

The UIT in our study was standardized and negatively valenced, and it would be interesting to test in future studies whether results generalize to other forms of intrusive thoughts, such as clinically relevant obsessions. The associated distress and the appraisal ratings indicate that the UITs activated in the study can be considered approximations of obsessional thoughts. However, it can be assumed that despite being comparable on some dimensions, the natural frequency of obsessional thoughts is higher than that of UITs (Purdon, Rowa, & Antony, 2005), which might affect its assessment with a counter app. For clinically relevant obsessions, an alternative or supplementary assessment method that directly

measures intrusiveness rather than frequency might be useful, such as thought replacement (Edwards & Dickerson, 1987; Olafsson et al., 2014; Purdon et al., 2011).

The study design did not allow us to distinguish between an increased occurrence of UITs and an increased awareness of UITs, reflected in a report bias, caused by using a counter app. Pressing a button could potentially remind the person of the UIT and thus trigger the next UIT, or pressing a button might merely increase the awareness of potential UITs, which might subsequently be more easily detected than if the button had not been pressed. Future studies might consider a study design that makes it possible to disentangle these two plausible mechanisms.

5. Conclusions

Can the counter app method be considered an appropriate method of assessing UITs in experimental studies? Advantages are its high convergent validity and test–retest reliability, that its use does not increase distress, and that it is not influenced by social desirability. However, the overestimation of UITs is considerable, and discriminant, criterion and predictive validity tend to be unsatisfactory. Adding a task that canalises the stream of consciousness seems to increase convergent validity considerably and predictive validity slightly. Retrospective frequency ratings, which are provided immediately after the time interval of interest, are a practical alternative to using a counter app, although convergent and discriminant validity appear to be slightly lower than for the counter app. Alternatively, investigators who are interested in an economic measure that is more sensitive to the emotional aspects of the UIT might choose to assess the retrospective intensity or overall intrusiveness of the UIT, which have high criterion and predictive validity.

References

Abramowitz, J. S., Tolin, D. F., & Street, G. P. (2001). Paradoxical effects of thought suppression: A meta-analysis of controlled studies. *Clinical Psychology Review, 21*(5), 683–703. doi:10.1016/s0272-7358(00)00057-x

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Bailey, B. E., Wu, K. D., Valentiner, D. P., & McGrath, P. B. (2014). Thought-action fusion: Structure and specificity to OCD. *Journal of Obsessive-Compulsive and Related Disorders, 3*(1), 39–45. doi:10.1016/j.jocrd.2013.12.003

Beck, A. T., Brown, G., Epstein, N., & Steer, R. A. (1988). An inventory for measuring clinical anxiety - Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*(6), 893–897. doi:10.1037/0022-006x.56.6.893

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the beck depression inventory-II*. San Antonio, TX: Psychological Corporation.

Blumberg, S. J. (2000). The white bear suppression inventory: revisiting its factor structure. *Personality and Individual Differences, 29*(5), 943–950. doi:10.1016/S0191-8869(99)00245-7

Bocci, L., & Gordon, P. K. (2007). Does magical thinking produce neutralising behaviour? An experimental investigation. *Behaviour Research and Therapy, 45*(8), 1823–1833. doi:10.1016/j.brat.2007.02.003

Clark, D. A., & Purdon, C. (2016). Still cognitive after all these years? Perspectives for a cognitive Behavioural theory of obsessions and where we are 30 years later: A commentary. *Australian Psychologist, 51*(1), 14–17. doi:10.1111/ap.12200

Clark, D. A., & Purdon, C. L. (1995). The assessment of unwanted intrusive thoughts - A review and critique of the literature. *Behaviour Research and Therapy, 33*(8), 967–976. doi:10.1016/0005-7967(95)00030-2

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. doi:10.1037/0033-2909.112.1.155

Crawford, J. R., & Henry, J. D. (2004). The positive and negative affect schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43*, 245–265. doi:10.1348/0144665031752934

de Bruin, G. O., Muris, P., & Rassin, E. (2007). Are there specific meta-cognitions associated with vulnerability to symptoms of worry and obsessional thoughts? *Personality and Individual Differences, 42*(4), 689–699. doi:10.1016/j.paid.2006.08.015

Edwards, S., & Dickerson, M. (1987). Intrusive unwanted thoughts - A 2-Stage model of control. *British Journal of Medical Psychology, 60*, 317–328. doi:10.1111/j.2044-8341.1987.tb02750.x

Erskine, J. A. K., Georgiou, G. J., Joshi, M., Deans, A., & Colegate, C. (2017). Ageing and thought suppression performance: Its relationship with working memory capacity, habitual thought suppression and mindfulness. *Consciousness and Cognition, 53*, 211–221. doi:10.1016/j.concog.2017.06.019

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The obsessive-compulsive inventory: Development and validation of a short version. *Psychological Assessment, 14*(4), 485–496. doi:10.1037//1040-3590.14.4.485

Freeston, M. H., Ladouceur, R., Thibodeau, N., & Gagnon, F. (1991). Cognitive intrusions in a non-clinical population. I. Response style, subjective experience, and appraisal. *Behaviour Research and Therapy, 29*(6), 585–597.

FunCoolApps. (2016). Click Counter (Version 1.4) [Mobile application software]. Retrieved from https://play.google.com/store/apps/details?id=com.useless.counter

Goenner, S., Leonhart, R., & Ecker, W. (2007). Das Zwangsinventar OCI-R—die deutsche Version des Obsessive–Compulsive Inventory-Revised: Ein kurzes Selbstbeurteilungsinstrument zur mehrdimensionalen Messung von Zwangssymptomen [The German version of the Obssessive–Compulsive Inventory-Revised: a brief self-report measure for the multidimensional assessment of obsessive–compulsive symptoms]. *Psychotherapie Psychosomatik Medizinische Psychologie, 57*, 395–404.

Goenner, S., Leonhart, R., & Ecker, W. (2008). The Obsessive-Compulsive Inventory-Revised (OCI-R): Validation of the German version in a sample of patients with OCD, anxiety disorders, and depressive disorders. *Journal of Anxiety Disorders, 22*(4), 734–749. doi:10.1016/j.janxdis.2007.07.007

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., . . . Charney, D. S. (1989). The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. *Arch Gen Psychiatry, 46*(11), 1006–1011.

Grisham, J. R., & Williams, A. D. (2009). Cognitive control of obsessional thoughts. *Behaviour Research and Therapy, 47*(5), 395–402. doi:10.1016/j.brat.2009.01.014

Hansmeier, J., Exner, C., Rief, W., & Glombiewski, J. (2014). Gedanken-Handlungs-Fusion und Zwangssymptome – eine Validierungsstudie der Thought Action Fusion-Skala (TAF-Skala) für den deutschen Sprachraum. *Zeitschrift für Klinische Psychologie und Psychotherapie, 43*(3), 204–214.

Hansmeier, J., Exner, C., Rief, W., & Glombiewski, J. (2014). Thought action fusion and obsessive-compulsive symptoms. A validation study of the Thought Action Fusion Scale (TAP Scale) for the German language area. *Zeitschrift Fur Klinische Psychologie Und Psychotherapie, 43*(3), 204–214. doi:10.1026/1616-3443/a000266

Hautzinger, M., Keller, F., & Kühner, C. (2006). *Das Beck Depressionsinventar II. Deutsche Bearbeitung und Handbuch zum BDI II*. Frankfurt am Main: Harcourt Test Services.

Hirsch, C. R., Perman, G., Hayes, S., Eagleson, C., & Mathews, A. (2015). Delineating the role of negative verbal thinking in promoting worry, perceived threat, and anxiety. *Clinical Psychological Science, 3*(4), 637–647. doi:10.1177/2167702615577349

Höping, W., & de Jong-Meyer, R. (2003). Differentiating unwanted intrusive thoughts from thought suppression: What does the White Bear Suppression Inventory measure? *Personality and Individual Differences, 34*, 1049–1055.

Janeck, A. S., Calamari, J. E., Riemann, B. C., & Heffelfinger, S. K. (2003). Too much thinking about thinking?: metacognitive differences in obsessive-compulsive disorder. *Journal of Anxiety Disorders, 17*(2), 181–195. doi:10.1016/s0887-6185(02)00198-6

Kollárik, M., van den Hout, M., Heinzel, C. V., Hofer, P. D., Lieb, R., & Wahl, K. (2020). Effects of rumination on unwanted intrusive thoughts: A replication and extension. *Journal of Experimental Psychopathology, 11*, 1–16. doi:10.1177/2043808720912583

Krohne, H. W., Egloff, B., Kohlmann, C. W., & Tausch, A. (1996). Investigations with a German version of the positive and negative affect schedule (PANAS). *Diagnostica, 42*(2), 139–156.

Lambert, A. E., Hu, Y., Magee, J. C., Beadel, J. R., & Teachman, B. A. (2014). Thought suppression across time: Change in frequency and duration of thought recurrence. *Journal of Obsessive-Compulsive and Related Disorders, 3*(1), 21–28. doi:10.1016/j.jocrd.2013.11.004

Lavy, E. H., & Van den Hout, M. A. (1990). Thought suppression induces intrusions. *Behavioural Psychotherapy, 18*(4), 251–258. doi:10.1017/s0141347300010351

Lin, Y. J., & Wicker, F. W. (2007). A comparison of the effects of thought suppression, distraction and concentration. *Behaviour Research and Therapy, 45*(12), 2924–2937. doi:10.1016/j.brat.2007.08.004

Magraf, J., & Ehlers, A. (2007). *Beck Angst-Inventar (BAI) - Manual*. Frankfurt am Main: Harcourt Test Services GmbH.

Marcks, B. A., & Woods, D. W. (2005). A comparison of thought suppression to an acceptance-based technique in the management of personal intrusive thoughts: a controlled evaluation. *Behaviour Research and Therapy, 43*(4), 433–445. doi:10.1016/j.brat.2004.03.005

Markowitz, L. J., & Purdon, C. (2008). Predictors and consequences of suppressing obsessional thoughts. *Behavioural and Cognitive Psychotherapy, 36*(2), 179–192. doi:10.1017/s1352465807003992

Meyer, J. F., & Brown, T. A. (2013). Psychometric evaluation of the thought-action fusion scale in a large clinical sample. *Assessment, 20*(6), 764–775. doi:10.1177/1073191112436670

Muris, P., Merckelbach, H., & Horselenberg, R. (1996). Individual differences in thought suppression. The White Bear Suppression Inventory: Factor structure, reliability, validity and correlates. *Behaviour Research and Therapy, 34*(5-6), 501–513. doi:10.1016/0005-7967(96)00005-8

Olafsson, R. P., Snorrason, I., Bjarnason, R. K., Emmelkamp, P. M. G., Olason, D. P., & Kristjansson, A. (2014). Replacing intrusive thoughts: Investigating thought control in relation to OCD symptoms. *Journal of Behavior Therapy and Experimental Psychiatry, 45*(4), 506–515. doi:10.1016/j.jbtep.2014.07.007

Purdon, C. (2001). Appraisal of obsessional thought recurrences: Impact on anxiety and mood state. *Behavior Therapy, 32*(1), 47–64. doi:10.1016/s0005-7894(01)80043-5

Purdon, C. (2020). Thought suppression. *Oxford Research Encyclopedia of Psychology*. Retrieved from https://oxfordre.com/psychology/view/10.1093/acrefore/9780190236557.001.0001/acrefore-9780190236557-e-84.

Purdon, C., & Clark, D. A. (1993). Obsessive intrusive thoughts in nonclinical subjects. Part I. Content and relation with depressive, anxious and obsessional symptoms. *Behaviour Research and Therapy, 31*(8), 713–720.

Purdon, C., & Clark, D. A. (1994). Obsessive intrusive thoughts in nonclinical subjects. Part II. Cognitive appraisal, emotional response and thought control strategies. *Behaviour Research and Therapy, 32*(4), 403–410.

Purdon, C., & Clark, D. A. (2000). White bears and other elusive intrusions - Assessing the relevance of thought suppression for obsessional phenomena. *Behavior Modification, 24*(3), 425–453. doi:10.1177/0145445500243008

Purdon, C., & Clark, D. A. (2001). Suppression of obsession-like thoughts in nonclinical individuals: impact on thought frequency, appraisal and mood state. *Behaviour Research and Therapy, 39*(10), 1163–1181. doi:10.1016/s0005-7967(00)00092-9

Purdon, C., & Clark, D. A. (2002). The need to control thoughts. *Cognitive Approaches to Obsessions and Compulsions: Theory, Assessment, and Treatment*, 29–43. doi:10.1016/b978-008043410-0/50004-0

Purdon, C., Gifford, S., McCabe, R., & Antony, M. M. (2011). Thought dismissability in obsessive-compulsive disorder versus panic disorder. *Behaviour Research and Therapy, 49*(10), 646–653. doi:10.1016/j.brat.2011.07.001

Purdon, C., Rowa, K., & Antony, M. M. (2005). Thought suppression and its effects on thought frequency, appraisal and mood state in individuals with obsessive-compulsive disorder. *Behaviour Research and Therapy, 43*(1), 93–108. doi:10.1016/j.brat.2003.11.007

Questback GmbH. (2013). Unipark [Survey software]. Retrieved from https://www.unipark.com/

Rachman, S. (1981). Part 1. Unwanted intrusive cognitions. *Advances in Behaviour Research and Therapy*, 89–99. doi:10.1016/0146-6402(81)90007-2

Rachman, S., Shafran, R., Mitchell, D., Trant, J., & Teachman, B. (1996). How to remain neutral: An experimental analysis of neutralization. *Behaviour Research and Therapy, 34*(11-12), 889–898. doi:10.1016/S0005-7967(96)00051-4

Radomsky, A. S., Alcolado, G. M., Abramowitz, J. S., Alonso, P., Belloch, A., Bouvard, M., . . . Wong, W. (2014). Part 1-You can run but you can't hide: Intrusive thoughts on six continents. *Journal of Obsessive-Compulsive and Related Disorders, 3*(3), 269–279. doi:10.1016/j.jocrd.2013.09.002

Rassin, E., Merckelbach, H., Muris, P., & Schmidt, H. (2001). The thought-action fusion scale: further evidence for its reliability and validity. *Behaviour Research and Therapy, 39*(5), 537–544.

Reed, G. F. (1985). *Obsessional experience and compulsive behaviour: A cognitive-structural approach*. New York: Academic Press.

Roncero, M., Belloch, A., & Doron, G. (2018). A novel approach to challenging OCD related beliefs using a mobile-app: An exploratory study. *Journal of Behavior Therapy and Experimental Psychiatry, 59*, 157–160. doi:10.1016/j.jbtep.2018.01.008

Rupp, C., Falke, C., Guehne, D., Doebler, P., Andor, F., & Buhlmann, U. (2019). A study on treatment sensitivity of ecological momentary assessment in obsessive-compulsive disorder. *Clinical Psychology & Psychotherapy*. doi:10.1002/cpp.2392

Salkovskis, P. M., & Campbell, P. (1994). Thought suppression induces intrusion in a naturally-occurring negative intrusive thoughts. *Behaviour Research and Therapy, 32*(1), 1–8. doi:10.1016/0005-7967(94)90077-9

Salkovskis, P. M., & Millar, J. F. A. (2016). Still cognitive after all these years? Perspectives for a cognitive behavioural theory of obsessions and where we are 30years later. *Australian Psychologist, 51*(1), 3–13. doi:10.1111/ap.12186

Salkovskis, P. M., & Reynolds, M. (1994). Thought suppression and smoking cessation. *Behaviour Research and Therapy, 32*(2), 193–201. doi:10.1016/0005-7967(94)90112-0

Schneider, F., Härter, M., & Schorr, S. (2017). *S3-Leitlinie/Nationale Versorgungs Leitlinie Unipolare Depression* (2nd ed.). Germany: Springer.

Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. In A. Stone, S. S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 11–26). New York, NY: Oxford University Press.

Shafran, R., Thordarson, D. S., & Rachman, S. (1996). Thought-action fusion in obsessive compulsive disorder. *Journal of Anxiety Disorders, 10*(5), 379–391. doi:10.1016/0887-6185(96)00018-7

Trinder, H., & Salkovskis, P. M. (1994). Personally relevant intrusions outside the laboratory - Long-term suppression increases intrusion. *Behaviour Research and Therapy, 32*(8), 833–842. doi:10.1016/0005-7967(94)90163-5

Urban, D., & Mayerl, J. (2006). *Regressionsanalyse: Theorie, Technik und Anwendung*. Wiesbaden: Springer VS.

Van Ameringen, M., Turna, J., Khalesi, Z., Pullia, K., & Patterson, B. (2017). There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. *Depression and Anxiety, 34*(6), 526–539. doi:10.1002/da.22657

Van den Bergh, O., & Lehnen, N. (2019). Embracing computational approaches can stimulate clinical psychology research. *Clinical Psychology in Europe*. doi:10.32872/cpe.v1i3.39237

van den Hout, M., Kindt, M., Weiland, T., & Peters, M. (2002). Instructed neutralization, spontaneous neutralization and prevented neutralization after an obsession-like

thought. *Journal of Behavior Therapy and Experimental Psychiatry, 33*(3-4), 177–189. doi:10.1016/s0005-7916(02)00048-4

van den Hout, M., van Pol, M., & Peters, M. (2001). On becoming neutral: effects of experimental neutralizing reconsidered. *Behaviour Research and Therapy, 39*(12), 1439–1448. doi:10.1016/s0005-7967(00)00109-1

Wahl, K., van den Hout, M., & Lieb, R. (2019). Rumination on unwanted intrusive thoughts affects the urge to neutralize in nonclinical individuals. *Journal of Obsessive-Compulsive and Related Disorders, 20*, 4–12. doi:10.1016/j.jocrd.2018.02.002

Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: A comprehensive review. *Revista Brasileira De Psiquiatria, 35*(4), 416–431. doi:10.1590/1516-4446-2012-1048

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect - The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. doi:10.1037/0022-3514.54.6.1063

Wegner, D. M., & Zanakos, S. (1994). Chronic thought suppression. *Journal of Personality, 62*(4), 615–640. doi:10.1111/j.1467-6494.1994.tb00311.x

Table 1

Means and Standard Deviations of Counter App Frequency, Retrospectively Rated Thought Descriptors and Social Desirability in Study 1

Variable	Experimental group					
	Phase 1			Phase 2		
	Counter app (n = 26)	Thought monitoring (n = 27)	Free thinking (n = 34)	Counter app (n = 26)	Thought monitoring (n = 27)	Free thinking (n = 34)
Counter app frequency	12.40 (8.57)			10.36 (15.65)		
Retrospective frequency	10.65 (8.18)	6.52 (18.88)	3.53 (2.89)	7.12 (7.04)	5.33 (19.00)	2.15 (3.59)
Retrospective duration	21.19 (18.41)	15.78 (16.57)	21.35 (20.02)	19.69 (20.48)	10.81 (20.31)	14.21 (18.23)
Retrospective percentage of time	29.15 (22.98)	17.63 (22.01)	23.83 (22.85)	20.38 (20.58)	8.26 (15.39)	13.24 (19.06)
Retrospective intensity	35.58 (23.18)	25.15 (27.70)	34.41 (27.56)	25.04 (23.14)	14.41 (25.53)	17.85 (22.07)
Retrospective intrusiveness	36.31 (24.96)	17.30 (23.87)	28.50 (28.75)	22.50 (22.00)	10.22 (20.30)	15.79 (21.48)
Social desirability						
Influenced by experimenter	.31 (.47)	.15 (.46)	.47 (.72)			
Make a certain impression	.08 (.52)	.11 (.58)	.25 (.62)			

Note. Standard deviations are presented in parentheses.

Table 2

Intercorrelations of Thought Descriptors and Correlations of Thought Descriptors With Neutralizing and Distress After Thought Activation (Reactivation), in Phase 1 (Below the Diagonal) and Phase 2 (Above the Diagonal) in Study 1

Measure	Counter app	Retrospective frequency	Retrospective duration	Retrospective percentage of time	Retrospective intensity	Retrospective intrusiveness	Covert neutralizing	Overt neutralizing	Distress
Counter app	(.64 **)	.41*	.22	.08	.19	.04	.17	.20	.02
Retrospective frequency	.64 **	(.95 **)	.66 **	.61 **	.61 **	.62 **	.29**	.31	.36 **
Retrospective duration	.25	.45 **	(.78 **)	.84 **	.84 **	.81 **	.47**	.22	.47**
Retrospective percentage of time	.47 **	.56 **	.72 **	(.72 **)	.86 **	.84 **	.52**	.45**	.56**
Retrospective intensity	.47 **	.43 **	.77 **	.76 **	(.72 **)	.86 **	.52**	.47**	.63**
Retrospective intrusiveness	.53 **	.56 **	.71 **	.82 **	.79 **	(.76 **)	.53**	.45**	.59**

Measure	Counter app	Retrospective frequency	Retrospective duration	Retrospective percentage of time	Retrospective intensity	Retrospective intrusiveness	Covert neutralizing	Overt neutralizing	Distress
Covert neutralizing	.02	.25*	.37**	.37**	.40**	.45**	-	.48	.27
Overt neutralizing	.21	.33**	.32**	.47**	.54**	.50**	.48**	-	.48 **
Distress	.15	.28 **	.32 **	.44 **	.38 **	.48 **	.21	.39	(.71 **)

Note. Correlations between scores on each measure in Phases 1 and 2 appear within parentheses on the diagonal. Neutralizing was assessed only once at the end of the study. All correlations with counter app were based on $n = 25$, all other correlations on $N = 87$.

*$p < .05$. **$p < .001$.

Table 3

Means and Standard Deviations of Counter App Frequency, and Retrospectively Rated Thought Descriptors, Distress and Social Desirability in Study 2

Variable	Phase 1			Phase 2		
	Experimental group			Experimental group		
	Counter app (n = 38)	Thought monitoring (n = 39)	Free thinking (n = 41)	Counter app (n = 38)	Thought monitoring (n = 39)	Free thinking (n = 41)
Counter app frequency	6.61 (5.27)			5.18 (5.34)		
Retrospective frequency	6.18 (5.76)	3.74 (3.59)	4.00 (7.12)	4.53 (5.28)	2.18 (1.96)	2.66 (3.27)
Retrospective duration	16.11 (18.94)	18.23 (14.59)	12.27 (14.70)	10.87 (13.79)	13.36 (16.96)	16.95 (24.10)
Retrospective percentage of time	17.39 (18.94)	18.77 (20.69)	13.66 (19.18)	11.45 (16.10)	12.46 (16.97)	17.44 (23.67)
Retrospective intensity	25.76 (26.95)	28.00 (23.59)	25.12 (26.71)	13.18 (17.51)	14.31 (18.46)	19.34 (25.74)
Retrospective intrusiveness	18.95 (21.20)	22.79 (24.24)	19.76 (23.78)	11.55 (13.50)	18.82 (24.48)	19.34 (26.55)
Distress						
Preexperimental manipulation	51.08 (29.67)	50.97 (30.99)	59.15 (28.34)	30.39 (25.43)	39.05 (29.56)	40.10 (30.73)
Postexperimental manipulation	27.11 (22.16)	36.92 (26.32)	30.37 (23.54)	16.87 (16.04)	29.62 (25.60)	27.85 (29.18)
Social desirability						

Variable	Phase 1			Phase 2		
	Experimental group			Experimental group		
	Counter app (n = 38)	Thought monitoring (n = 39)	Free thinking (n = 41)	Counter app (n = 38)	Thought monitoring (n = 39)	Free thinking (n = 41)
Influenced by experimenter	0.39 (0.79)	0.49 (0.56)	0.41 (0.63)			
Make a certain impression	0.08 (0.27)	0.13 (0.41)	0.17 (0.44)			

Note. Standard deviations are presented in parentheses.

Table 4

Intercorrelations of Thought Descriptors and Correlations of Thought Descriptors With Neutralizing and Distress After Thought Activation and Reactivation, in Phase 1 (Below the Diagonal) and Phase 2 (Above the Diagonal) in Study 2

Measure	Counter app	Retrospective frequency	Retrospective duration	Retrospective percentage of time	Retrospective intensity	Retrospective intrusiveness	Covert neutralizing	Overt neutralizing	Distress
Counter app	(.71 **)	.90 **	.57 **	.61 **	.37 *	.39 *	.21	.16	.23
Retrospective frequency	.86 **	(. 73 **)	.38 **	.54 **	.36 **	.28**	.13	.26 *	.26 **
Retrospective duration	.46 **	.42 **	(.48 **)	.78 **	.85 **	.70 **	.38 **	.41 **	.56 **
Retrospective percentage of time	.76 **	.64 **	.71 **	(.76 **)	.66 **	.66 **	.26 **	.38 **	.44 **
Retrospective intensity	.47 **	.51 **	.60 **	.63 **	(.71 **)	.69 **	.40 **	.43 **	.66 **
Retrospective intrusiveness	.45 **	.54 **	.60 **	.69 **	.76 **	(.69 **)	.38 **	.39 **	.54 **
Covert neutralizing	.27	.21 *	.27 *	.33 **	.40 **	.38 **	_	.51 **	.56 **

Measure	Counter app	Retrospective frequency	Retrospective duration	Retrospective percentage of time	Retrospective intensity	Retrospective intrusiveness	Covert neutralizing	Overt neutralizing	Distress
Overt neutralizing	.36 *	.26 **	.47 **	.45 **	.58 **	.49 **	.50 **	_	.50 **
Distress	.34 *	.22 *	.37 **	.32 **	.46 **	.51 **	.41 **	.49	(.63 **)

Note. Correlations between scores on each measure at Phase 1 and Phase 2 appear within parentheses on the diagonal. Neutralizing was assessed only once at the end of the study. All correlations with counter app were based on $n = 39$, all other correlations on $N = 118$.

*$p <$. 05. **$p < .001$.

Table 5

Depression, Obsessive-Compulsive Symptoms, Tendency to Suppress Thoughts and Thought–Action Fusion Predicting Persistence of the Unwanted Intrusive Thought in Study 2

Measure	*B*	*SE*	*Beta*	*p*
BDI-II	.32	1.10	.03	.763
OCI-R	.79	.55	.15	.157
WBSI	-.36	.36	-.11	.322
TAFS	1.10	.32	.33	.001

Note. BDI-II = Beck Depression Inventory, second edition; OCI-R = Obsessive-Compulsive Inventory; TAFS = Thought-Action Fusion Scale; WBSI = White Bear Suppression Inventory.

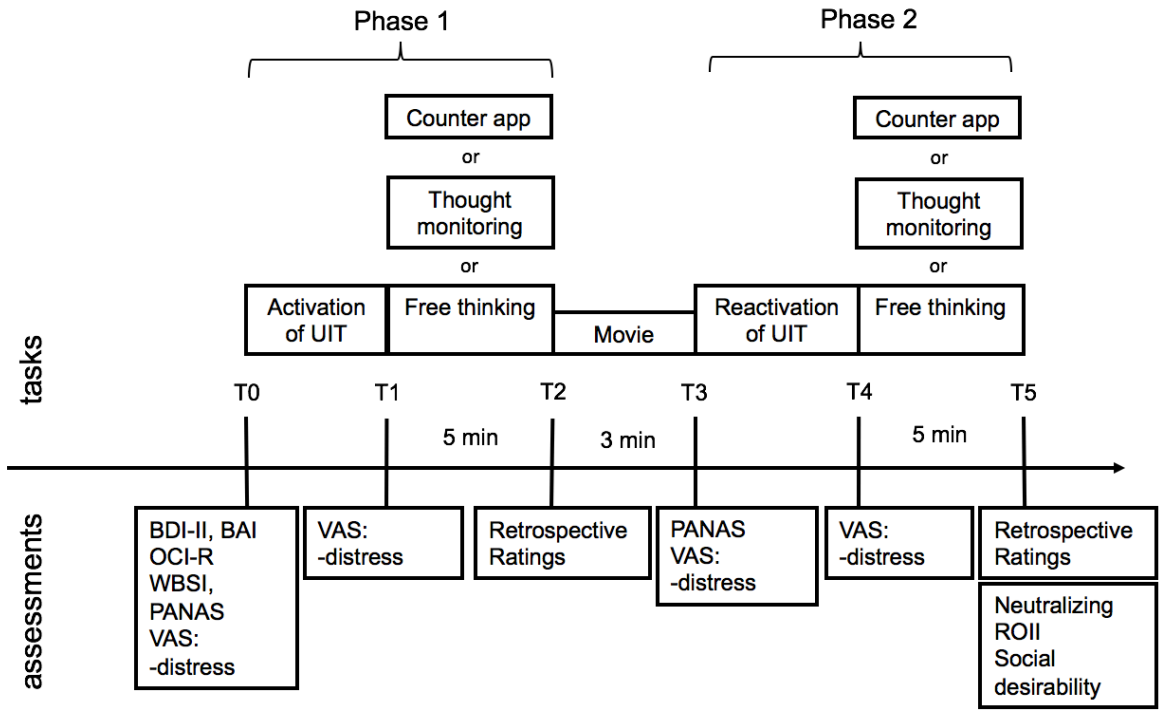

Figure 1. Experimental procedure of Study 1. BAI = Beck Anxiety Inventory; BDI-II = Beck Depression Inventory, second edition; OCI-R = Obsessive-Compulsive Inventory, revised; PANAS = Positive and Negative Affect Schedule; ROII = Revised Obsessive Intrusions Inventory; UIT = unwanted intrusive thought; VAS = visual analogue scale; WBSI = White Bear Suppression Inventory.

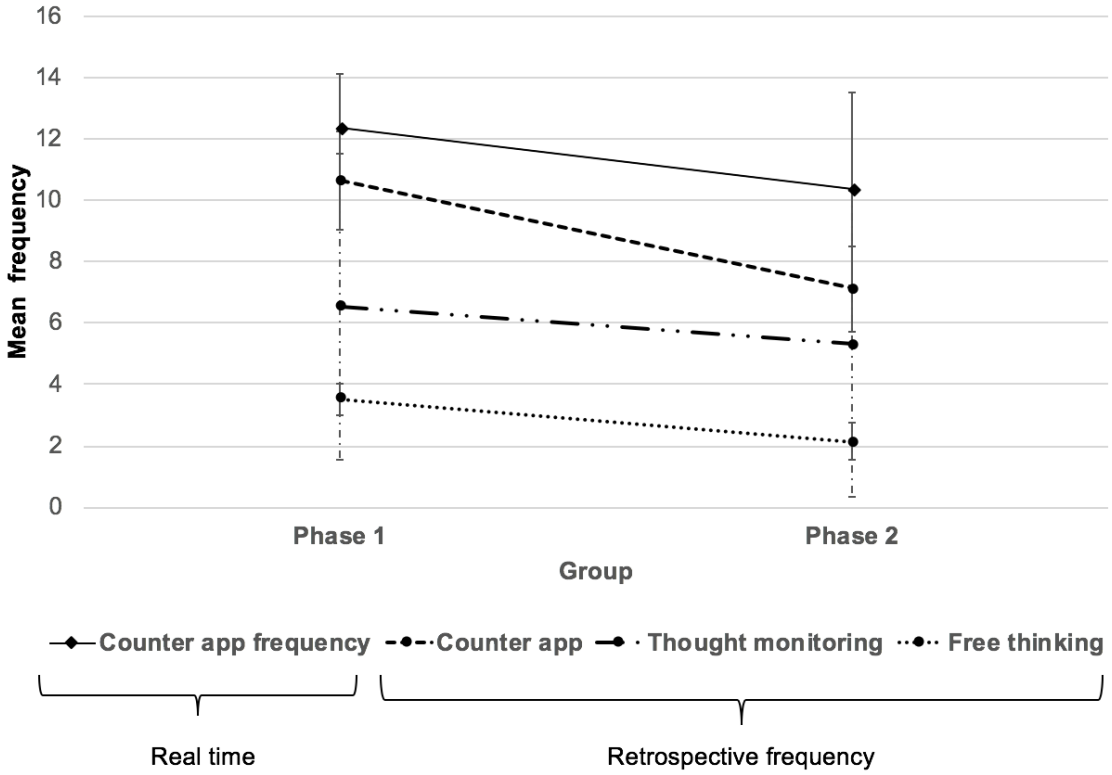

Figure 2. Mean thought frequencies and standard errors per phase and group in Study 1.

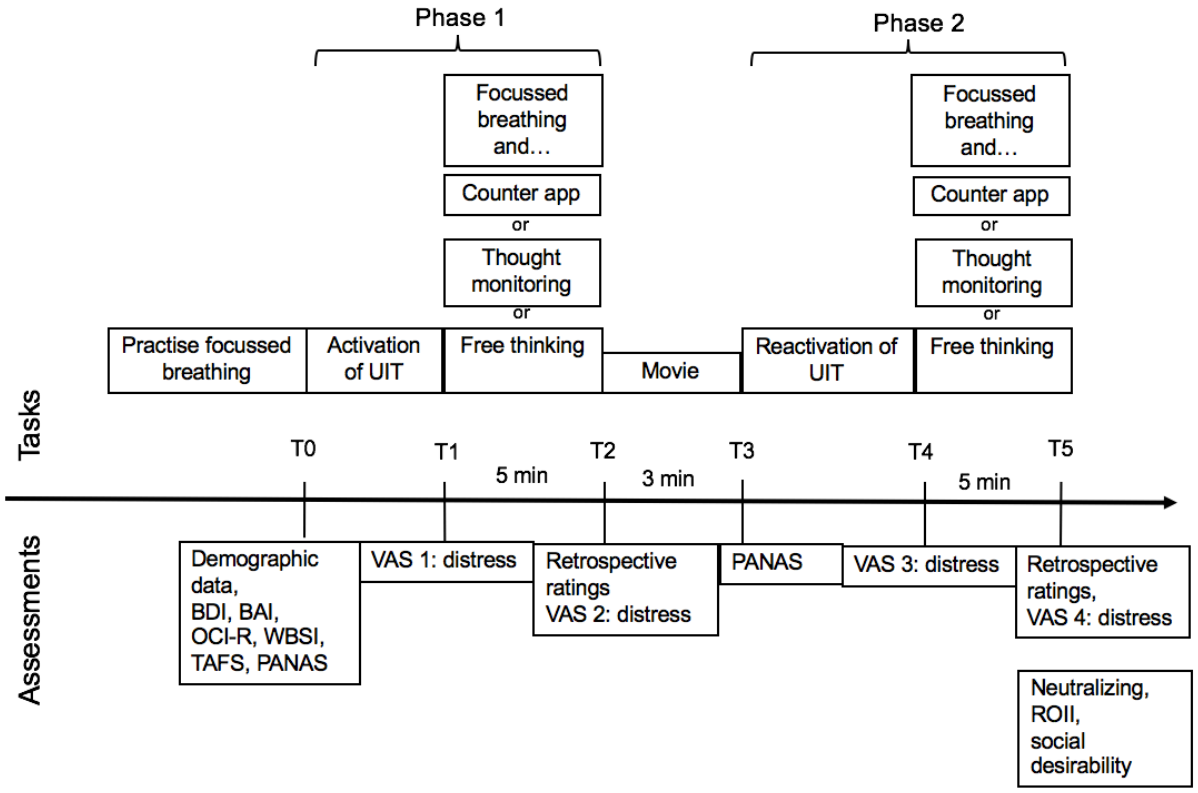

Figure 3. Experimental procedure of Study 2. BAI = Beck Anxiety Inventory; BDI-II = Beck Depression Inventory, second edition; OCI-R = Obsessive-Compulsive Inventory, revised; PANAS = Positive and Negative Affect Schedule; ROII = Revised Obsessive Intrusions Inventory; TAFS = Thought-Action Fusion Scale; UIT = unwanted intrusive thought; VAS = visual analogue scale; WBSI = White Bear Suppression Inventory.

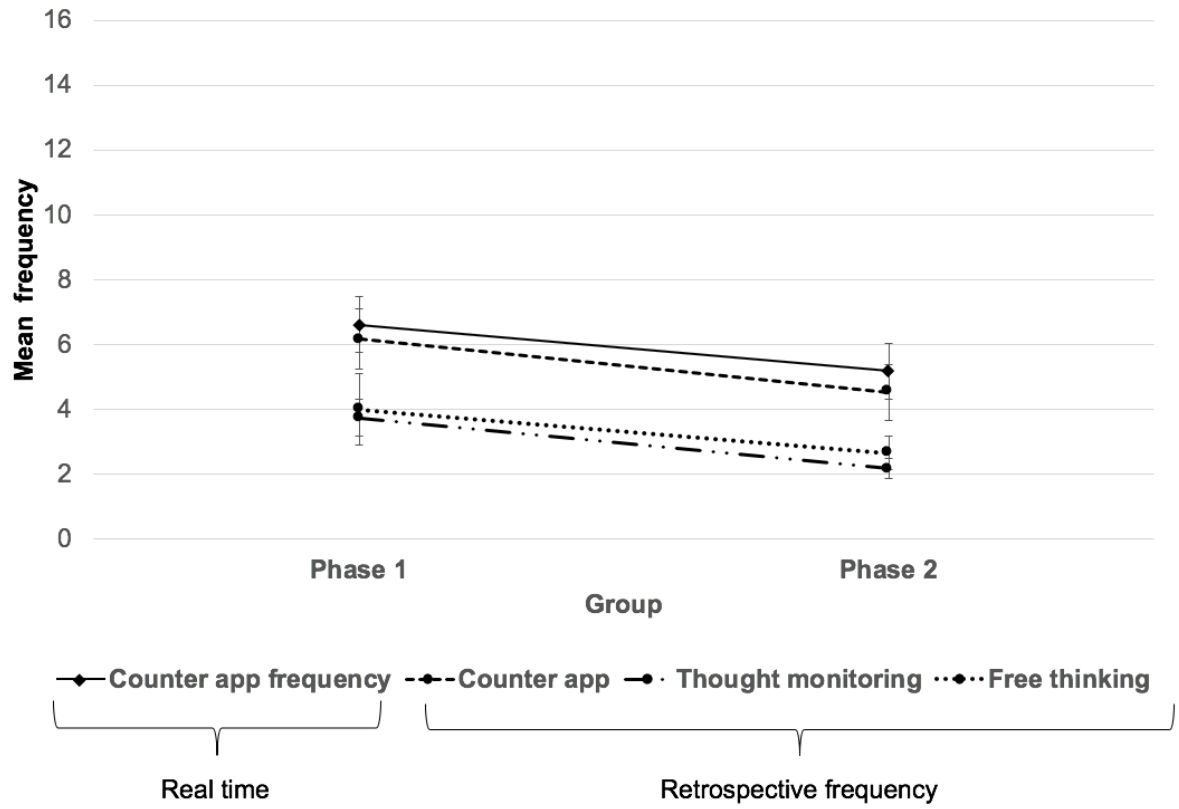

Figure 4. Mean thought frequencies and standard errors per phase and group in Study 2.