Personality and Individual Differences 91 (2016) 89–97

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Psychometric qualities of the Thought Suppression Inventory-Revised in different age groups

Kevin van Schie [a,b,*,1], Sabine Wanmaker [b], Iris Yocarini [b], Samantha Bouwmeester [b]

[a] Clinical Psychology, Utrecht University, PO Box 80140, NL-3508 TC Utrecht, The Netherlands
[b] Institute of Psychology, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received 24 April 2015
Received in revised form 1 November 2015
Accepted 28 November 2015
Available online 4 December 2015

Keywords:
Thought suppression
Intrusions
Dimensionality
Mokken scale analysis

ABSTRACT

Intrusive thoughts about negative events are core symptoms of several psychiatric disorders. Because current instruments for the assessment of thought suppression are unsatisfactory, we developed and evaluated the dimensionality and validity of a questionnaire that distinguishes between three major facets of thought suppression – intrusions, suppression attempts, and effective suppression – that affect psychopathology distinctly. Participants ($N = 784$) divided over three age groups, 25 years and younger ($n = 351$), between 26 and 50 years ($n = 202$), and 51 years or older ($n = 231$), completed the Thought Suppression Inventory-Revised. The data were analyzed with sophisticated nonparametric item response theory. Exploratory Mokken scale analysis revealed a three-factor structure, which was affirmed with confirmatory analyses. The Suppression Attempts scale appeared to be a weak scale, specifically in the two older age groups. Since suppression most likely depends on inhibitory ability – which declines with age – suppression attempts probably have increasingly variable outcomes (i.e., failure or success), which complicates measuring this factor. Overall, our findings suggest that three facets of thought suppression can be measured especially in younger individuals, but that for individual measurements particularly in the older age groups the Suppression Attempts scale has to be used with caution.

1. Introduction

Unwanted, intrusive thoughts about negative events are central to a number of psychological disorders, such as obsessive–compulsive disorder (OCD; Julien, O'Connor, & Aardema, 2007), post-traumatic stress disorder (PTSD; Ehlers & Clark, 2000; Shipherd & Salters-Pedneault, 2008), and depression (Wenzlaff, 2005; Wenzlaff & Wegner, 2000). Such thoughts are also common in the daily lives of healthy individuals (Brewin, Christodoulides, & Hutchinson, 1996; Clark & Rhyno, 2005; Rachman & de Silva, 1978) and it is self-evident that people occasionally attempt to suppress these thoughts. The extent to which individuals are successful in doing this varies and likely differs with age; therefore studying thought suppression in non-clinical individuals of different ages could serve as a useful model for pathological thought suppression. However, current instruments for thorough assessment of thought suppression are unsatisfactory, because questionnaire items do not have simple structure or are unscalable. Thus, the current study set out to develop an instrument that properly distinguishes between three major facets that make up the process of thought suppression – intrusions, suppression attempts, and effective suppression – in several age groups.

The development of an instrument that encompasses all aspects of thought suppression is especially relevant because in the last decade, the idea that thought suppression is always ineffective and counterproductive has been challenged by numerous studies. These studies either show that rebound effects – an increase of intrusive thoughts after suppression – are inconsistent (Magee, Harden, & Teachman, 2012; Purdon, 1999), or that suppressed memories can actually become less accessible (e.g., Anderson & Green, 2001; Depue, Curran, & Banich, 2007; van Schie, Geraerts, & Anderson, 2013; for reviews see Anderson & Hanslmayr, 2014; Anderson & Huddleston, 2011). Interestingly, engaging in thought suppression can also reduce the number of unwanted memories that intrude into awareness (Benoit, Hulbert, Huddleston, & Anderson, 2014; Levy & Anderson, 2012). Hence, the blanket term 'thought suppression' may long have been equated erroneously with ineffectiveness (e.g., in models of psychopathology; Ehlers & Clark, 2000; Rachman, 1998), while evidently it can be effective in reducing intrusive memories. Thus, if certain people are able to regulate negative affect by effectively suppressing intrusive memories that evoke feelings of fear, anxiety or anger, this could reduce psychopathological symptoms. Research on individual differences in thought suppression may shed light on why some people are better in regulating unwanted intrusive thoughts than others.

* Corresponding author at: Clinical Psychology, Utrecht University, PO Box 80140, NL-3508 TC Utrecht, The Netherlands.
E-mail addresses: k.vanschie@uu.nl (K. van Schie), wanmaker@fsw.eur.nl (S. Wanmaker), yocarini@fsw.eur.nl (I. Yocarini), bouwmeester@fsw.eur.nl (S. Bouwmeester).
[1] The work was performed at both universities.

http://dx.doi.org/10.1016/j.paid.2015.11.060
0191-8869/

Moreover, it is likely that thought suppression may change over an individual's lifetime, as with age – at least some – inhibitory abilities decline (Hasher & Zacks, 1988). Indeed, on a behavioral level older adults experience more pro-active interference in memory than younger adults, which is exactly what one would expect if older adults are less able to engage in inhibitory control (Biss, Campbell, & Hasher, 2013; Lustig, Hasher, & Tonev, 2001). Furthermore, decreased inhibitory control hinders older adults from intentionally forgetting unwanted episodic memories (Anderson, Reinholz, Kuhl, & Mayr, 2011). Hence, age-related decreases in inhibitory control may therefore lead to increases of intrusive, recurrent memories, which are core symptoms of psychiatric disorders such as PTSD and OCD. To determine the exact relationship between thought suppression and psychopathology, it is essential to adequately measure all aspects of thought suppression.

The White Bear Suppression Inventory (WBSI; Wegner & Zanakos, 1994) was the first instrument enabling the assessment of individual thought suppression and its relation to psychopathology. However, the WBSI seemed to lack a consistent factor dimensionality and therefore highlights that thought suppression is not a unidimensional construct; a substantial number of studies found that at least some of the items capture another construct, namely intrusions (Blumberg, 2000; Höping & De Jong-Meyer, 2003; Luciano et al., 2006; Muris, Merckelbach, & Horselenberg, 1996; Pica, Pierro, & Giannini, 2014; Rassin, 2003; Rodríguez, Delgado, Rovella, & Cubas León, 2008; Schmidt et al., 2009; Spinhoven & Van der Does, 1999). This makes interpretation difficult because it is unclear whether a low score implies successful thought suppression, the absence of experienced intrusions, or both (Blumberg, 2000; Rassin, 2003). This continuing debate on the construct validity and the imperfections of the WBSI necessitated the development of an instrument that overcame its shortcomings by using separate scores for different constructs.

Contrary to the WBSI, the three-factor Thought Suppression Inventory (TSI, Rassin, 2003) seemed to successfully differentiate thought suppression (attempts) from intrusions, and from successful thought suppression. In a student population, Rassin (2003) showed that intrusions were strongly related to general psychopathology and obsessive–compulsive symptoms, while suppression attempts were not.[2] Additionally, successful suppression on the TSI correlated negatively with WBSI intrusions and suppression components, showing that the WBSI – which was intended to measure suppressing thoughts – essentially measures *failed suppression*. Thus, Rassin (2003) emphasized the inherent bias of the WBSI and introduced the TSI as an alternative questionnaire for integral thought suppression.

Though the TSI was a first instrument to assess the complete construct of thought suppression, it is not without flaw. Using sophisticated tools from modern item response theory, Wismeijer (2012) revealed that in an elderly population 8 out of 15 TSI items had unsatisfactory psychometric properties. These items did not have a simple structure or were unscalable (e.g., item 12 loaded equally on all scales 'I am able to suppress unpleasant experiences to the point that I hardly remember them'). Consequently, Wismeijer proposed rephrasing or replacing of these items.

Since the TSI proved to be psychometrically unsound and its generalizability unclear, the goal of our study was two-fold. First, we critically examined the TSI's items, developed new items, and rephrased old items. This was done in order to create a valid revised questionnaire that adequately distinguishes between thought suppression attempts, successful thought suppression and unwanted intrusive thoughts. Second, because thought suppression likely varies with age and since previous studies predominantly examined the psychometric properties of the TSI with age-restricted samples (viz., students (mean age = 20.4, SD = 2.5) and the elderly (mean age = 65.0, SD = 9.58); Rassin, 2003; Wismeijer, 2012), our study used a broad stratified quota sample with individuals in the age range of 16 to 83 years. This allowed us to examine if the psychometric properties of the TSI-Revised (TSI-R) are widely supported by all subgroups, and not only by students, who are a commonly examined group. Additionally, we hypothesized that, as a consequence of reduced inhibitory control with age, older groups experience less effective suppression and more intrusions. Following Wismeijer (2012), modern test theory was used to explore the dimensionality of TSI-R. Modern test theory offers rich and sophisticated tools to scrutinize the psychometric properties by focusing on item responses (Reise, Ainsworth, & Haviland, 2005).

2. Method

2.1. Participants

The majority of the 784 participants in our study indicated being Dutch nationals (96.4%), while a minority had non-Dutch or dual nationalities (3.6%), though all participants spoke Dutch. The majority of our sample indicated their highest level of completed education was at undergraduate or graduate level (71.3%), followed by vocational training (17.0%), high school (10.3%), or another type of education (1.4%). Participants were divided over three age categories to ensure a divers sample; 25 years and younger (n = 351, mean age = 20.2 years, SD = 2.19, 28.5% male), between 26 and 50 years (n = 202, mean age = 36.1 years, SD = 7.79, 33.7% male), 51 years or older (n = 231, mean age = 62.4 years,[3] SD = 7.61, 42.9% male). All subjects that finished at least the TSI-R at the first testing time were requested to participate in the second testing period. Consequently, the sample measuring test-retest reliability consisted of 427 subjects (25 years or younger: n = 82, mean age = 22 years, SD = 2.21, 32.9% male; between 26 and 50 years: n = 156, mean age = 36.7 years, SD = 7.87, 32.7% male; 51 years or older: n = 189, mean age = 62.7 years[3], SD = 7.68, 43.1% male). All subjects participated voluntarily.

2.2. Questionnaires

2.2.1. Thought Suppression Inventory-Revised

The Thought Suppression Inventory-Revised (TSI-R) is a revision of the TSI (Rassin, 2003), a Dutch 15-item self-report instrument to measure successful and unsuccessful thought suppression. Items are scored from 1 *strongly disagree* to 5 *strongly agree* indicating agreement with statements such as 'I have many unpleasant thoughts'. Total scale scores are calculated by adding item scores for each of the three scales independently. Total scores range from 5 to 25, with higher scores indicating more intrusions, more suppression attempts, or successful suppression (compared to non-successful suppression). for the revised version, TSI items critiqued by Wismeijer (2012) were either rephrased or replaced, and several new items were added. Consequently, the TSI-R consisted of 21 items at the moment of testing (see Table 1, for TSI-R items)

2.2.2. White Bear Suppression Inventory

The WBSI (Wegner & Zanakos, 1994) is a 15-item self-report questionnaire measuring thought suppression. Items are scored from 1 *strongly disagree* to 5 *strongly agree* and total scores vary from 15 to 75, where higher total scores reflect a stronger tendency towards thought suppression. See Appendix A for reliability estimates (internal consistency) of our three age groups for all questionnaires besides the TSI-R.

[2] Note that this lack of correlation might also be attributed to relatively large measurement error – shown by the low reliability estimates (test-retest reliability = .43, Cronbach's alpha = .64) – of suppression attempts in comparison with the other constructs (Furr & Bacharach, 2014).

[3] Mean age was calculated based on 230 participants for the first testing time and on 188 participants for the second testing time; one participant did not report age, only age group.

Table 1
Twenty-one translated items of the Dutch TSI-R.

Item	Construct	Description
1.	Int	I have thoughts, which I would rather not have.
2.	Sup Att	Sometimes I stay busy just to prevent having certain thoughts.
3.	Eff Sup	I am able to suppress unpleasant thoughts.
4.	Int	I have many unpleasant thoughts.
5.	Sup Att	I try to suppress certain thoughts.
6.	Eff Sup	I am able to put aside problems and worries.
7.	Int	I experience many emotions that are too intense to control.
8.	Sup Att	I try to think about something else, if I have an unpleasant thought.
9.	Eff Sup	I am able to keep a problem out of mind until I have time to deal with it.
10.	Int	I am often preoccupied with certain unwanted thoughts or ideas.
11.	Sup Att	I try to put unpleasant thoughts out of mind as quickly as possible.
12.	Eff Sup	I am able to suppress an unpleasant experience to the point that it hardly comes to mind.
13.	Int	Some unwanted thoughts enter my mind without me being able do anything about it.
14.	Sup Att	I try not to think of unpleasant events.
15.	Eff Sup	I succeed in controlling unwanted thoughts whenever it is necessary.
16.	Int	I am unable to concentrate because certain unpleasant thoughts dominate my mind.
17.	Sup Att	Sometimes I decide to keep certain memories out of mind.
18.	Eff Sup	I am able to stop unpleasant memories from coming to mind whenever I want to.
19.	Int	My thoughts are often about the same unpleasant idea.
20.	Sup Att	Certain things I try not to think of.
21.	Eff Sup	I am able to put aside unpleasant thoughts and/or images effortlessly.

Note. Int = Intrusion; Sup Att = Suppression Attempts; Eff Sup = Effective Suppression.

2.2.3. Thought Control Ability Questionnaire

The Thought Control Ability Questionnaire (TCAQ; Luciano, Algarabel, Tomás, & Martínez, 2005) is a 25-item self-report measure that measures the ability to control unwanted thoughts. Items are scored from 1 *strongly disagree to* 5 strongly *agree*. Total sum scores range from 25 to 125, where high scores indicate more control over unwanted intrusive thoughts.

2.2.4. Obsessive–Compulsive Inventory-Revised

The Obsessive–Compulsive Inventory-Revised (OCI-R; Foa et al., 2002) is an 18-item self-report measuring obsessions and compulsions. Items are from 0 *not at all* to 4 *extremely*, and total scores range from 0 to 72, where a higher score is indicative of more OCD related behavior. Additionally, subscale scores can be calculated for six three-item factors: washing, checking, ordering, hoarding, obsessing, and mental neutralizing.

2.2.5. PTSD Symptom Scale Self-Report

The PTSD Symptom Scale-Self Report (PSS-SR; Foa, Riggs, Dancu, & Rothbaum, 1993) is a 17-item self-report in which participants rate how much they experienced DSM-IV specified PTSD symptoms. Items are scored from 0 *not at all* to 3 *five times per week or more*, and higher sum total scores (ranging from 0 to 51) indicate higher severity of PTSD symptoms. Additionally, the PSS-SR consists of three clusters; re-experiencing (4 items), avoidance (7 items), and arousal (6 items).

2.2.6. The Marlowe–Crowne Social Desirability Scale

The Marlowe–Crowne Social Desirability Scale (MC-SDS; Crowne & Marlowe, 1960) is 33-item self-report instrument, which measures social desirable response tendencies. Items are scored dichotomously (yes/no) and sum total scores range from 0 to 33; higher scores indicating social desirable response tendencies.

2.3. Procedure

Participants were recruited from volunteer research pools and the nearby community via e-mail lists and snowball sampling. Contacted participants were invited via e-mail to take part in a two-phase online questionnaire programmed in Qualtrics (Qualtrics, Provo, UT). They were requested to fill out all questionnaires in a quiet space or room where they would not be disturbed. At the first time of testing participants provided demographical data (age, sex, level of education and ethnicity), and filled out TSI-R, WBSI, TCAQ, OCI-R, PSS-SR and MC-SDS (fixed order), which took approximately 25 min to complete. After two weeks participants who at least completed the TSI-R were requested to fill out the TSI-R a second time, which took no longer than 5 min to complete.

2.4. Statistical analyses: the monotone homogeneity model

Like Wismeijer (2012), we used a nonparametric model stemming from item response theory to assess the dimensionality and the scalability of the data (for an introduction into nonparametric IRT, see Sijtsma & Molenaar, 2002). Specifically, we applied one of the Mokken models to the data, namely the monotone homogeneity model (MHM; Mokken & Lewis, 1982). One advantage of using item response theory in exploring the dimensionality of the data is that it enables us to take the measurement level of the data into account. For the TSI-R, it is reasonable to suppose that the psychological constructs – Intrusion, Suppression Attempts, and Effective Suppression – have an ordinal measurement level. Therefore, the MHM is particularly suited as it imposes only ordinal restrictions on the data, while other models (e.g., 2PLM) might be unduly restrictive (see Sijtsma & Molenaar, 2002, for an in-depth comparison of the NIRT models and parametric IRT models). Technically, this implies that the item response function (IRF) that describes the relationship between the item score and the underlying latent trait or scale has to be monotonously non-decreasing, meaning that the steeper the IRF of an item, the better that item discriminates between different scale values and the better the item fits into the scale. Therefore, one might say that an adequately discriminating item measures the underlying latent trait well. In turn, in an exploratory setting, the interpretation of the latent trait is determined by well discriminating items, just like this is done by using high factor loadings to interpret the factors in a principal component analysis.

Mokken scale analysis (MSA) offers some sophisticated tools to investigate scalability of items of a polytomous item set and the unidimensionality of those items (Hemker, Sijtsma, & Molenaar, 1995). First, the H coefficient (Loevinger, 1948) of a scale gives information about the scalability of a set of items as a scale. Concurrently, H scale coefficients express model fit, where H has a maximum of 1. Values between .30 and .40 are considered to be a weak scale, values between .40 and .50 constitute a medium scale, and values from .50 are interpreted as a strong scale (Mokken, 1971, p.185). A strong scale implies that the items constitute a uniform set and that these items are good in discriminating different values on the latent trait. The H coefficient can also be determined for individual items, the H_i coefficients. These show the scalability of individual items within a scale. According to Mokken individual items require H_i of at least .3. The higher the H_i, the steeper the IRF of an item (i.e., better discriminability).

Furthermore, MSA can be used to assess the assumption of monotonously non-decreasing IRFs by testing whether decreases on some part(s) of the scale are significant. These tools enable us to explore the scalability of items on a more appropriate and sophisticated way then would have been possible when using exploratory factor analysis, such as principal component analysis.

MSA can be used in a confirmatory and exploratory way. In an exploratory approach we wish to explore the dimensionality of the data since 8 of 15 original TSI items were adjusted according to the recommendations of Wismeijer (2012), new items were added, and

because we use a different (viz., broader) population. In the exploratory analyses we will search for scales that consist of items that have at least weak scale quality ($H_i > .3$). Following Wismeijer (2012) as much as possible, the items 1, 4, 7, 10, 13, 16, and 19 should fit in the Intrusion scale, the items 2, 5, 8, 11, 14, 17, and 20 in the Suppression Attempts scale, and the items 3, 6, 9, 12, 15, 18, and 21 should fit in the Effective Suppression scale. Subsequently, in confirmatory analyses we wish to cross validate the dimensionality (i.e., scales) that resulted from the exploratory analysis. To accommodate both exploratory and confirmatory analyses, we split the dataset in two by assigning the even rows to the train set and the uneven rows to the test set. To ensure age is evenly distributed we sorted the data by age before splitting the data. Specifically, confirmatory analyses were performed on three different age groups [25 and younger ($n = 176$); between 26 and 50 ($n = 101$); and 51 years and older($n = 115$)] and the quality of the scales for these three groups was compared. The MSA were performed using the Mokken package (Van der Ark, 2012) in R (R Core Team, 2013).

3. Results

3.1. Exploratory MSA

In the exploratory MSA we started with a lower bound value of $H = .30$ and gradually increased this lower bound value to assess how the items group together. In the standard item selection algorithm of MSA – automated item selection procedure (AISP) – items are sequentially added to the scale as long as the H coefficient of the scale including the item is above .30 (when using a lower bound of .30). When no items are left that fulfill the criteria, a second, third, etc. scale is evaluated according to the same criteria. The exploratory MSA predominantly revealed three scales with H values ranging between $H = .44$ and $H = .53$ for different values of the lower bound. Table 2 shows the grouping of the items for different values of the lower bounds. Note that only the results of the lower bounds .30, .35, .40 and .45 are reported as higher lower bounds were undesirably strict and many items were excluded.

Table 2
Selected clusters from AISP, H_i coefficients, H coefficients per cluster in exploratory Mokken scale analysis.

Item	$c = .3$			$c = .35$			$c = .40$			$c = .45$			
	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C4
Int1	.50			.52			.52			.52			
Int4	.59			.62			.62			.62			
Int7	.46			.48			.48			.48			
Int10	.58			.61			.61			.61			
Int13	.48			.51			.51			.51			
Int16	.48			.50			.50			.50			
Int19	.46			.48			.48			.48			
Sup Att2	.49			.51			.51			.51			
Sup Att5	.51			.52			.52			.52			
Sup Att8			.42			.42			.42				
Sup Att11			.48			.47			.47			.51	
Sup Att14			.44			.44			.44			.51	
Sup Att17			.43			.45			.45				.49
Sup Att20	.32					.45			.45				.49
Eff Sup3													
Eff Sup6		.46			.46			.50			.50		
Eff Sup9		.46			.46			.49			.49		
Eff Sup12		.38			.38								
Eff Sup15		.49			.49			.51			.51		
Eff Sup18		.52			.52			.55			.55		
Eff Sup21		.56			.56			.58			.58		
Scale	.49	.48	.44	.53	.48	.47	.53	.53	.47	.53	.53	.51	.49

Note. c = lower bound; C = cluster; Int = Intrusion; Sup Att = Suppression Attempts; Eff Sup = Effective Suppression.

For all lower bounds, the first component can be interpreted as the Intrusion scale, the second component as the Effective Suppression scale, and the third component as the Suppression Attempts scale. Remarkably, item 2 ('Sometimes I stay busy just to prevent having certain thoughts') and item 5 ('I try to suppress some thoughts') were meant to load on the Suppression Attempts scale but consistently loaded on the Intrusion scale. Additionally, at a lower bound of $H = .3$, item 20 ('Certain things I try not to think of') also loads on the Intrusion component. However, item 20 does fit into the Suppression Attempts scale as intended when the lower bound is increased. Furthermore, item 3 ('I am able to suppress unpleasant thoughts') does not fit any of the scales when a minimum lower bound of $H = .30$ is required. The same holds for item 12 ('I am able to suppress unpleasant experiences to the point that I hardly remember them') if the lower bound equals $H = .40$, and for item 8 ('I try thinking of something else if I have an unpleasant thought') if the lower bound equals $H = .45$. When $H = .45$, four components are found in which the Suppression Attempts scale is further broken down into two scales; one regarding suppression attempts of unpleasant experiences or thoughts, and one concerning suppression attempts of memories.

The results from using a lower bound of $H = .35$ fit best given our expectations. Item 20 does fit into the Suppression Attempts scale and item 12 and 8 are still included. Consequently, this structure will be used in further confirmatory analyses. In this scale structure, the Intrusion scale has the highest H coefficient; $H = .53$, with H_i values ranging from .48 (item 19) to .62 (item 4), and a reliability of $\lambda_2 = .88$. The Effective Suppression scale has an H coefficient of $H = .48$, with H_i ranging from .38 (item 12) to .56 (item 21). For this scale reliability was $\lambda_2 = .81$. Finally, the Suppression Attempts scale has an H coefficient of $H = .47$, with H_i values ranging from .42 (item 8) to .47 (item 11). The scale reliability is $\lambda_2 = .77$. For all items in these three scales the monotonicity was not violated (i.e., none of the IRFs decreased). Furthermore, for semantic reasons, items 2 and 5 were removed from the Intrusion scale to enable a clear interpretation of the scale as Intrusions. Consequently, the final Intrusion scale consisted of 7 items, had an H coefficient of $H = .54$, in which H_i values ranged from .48 (item 19) to .62 (item 10), and a scale reliability of $\lambda_2 = .86$. Overall, from the exploratory MSE we can conclude that the hypothesized three-factor structure is present, with strong to medium scales.

3.2. Confirmatory MSA

Subsequently, a confirmatory MSA was performed on the test dataset ($n = 392$) to confirm the three-factor structure found in the exploratory MSA (see Table 3). The Intrusion scale has an H value of .55, indicating strong scale strength. Individual item H_i values range from .49 (item 19) to .60 (item 4), which indicates that the items belonging to this scale discriminate quite well between different values on the Intrusion scale. The reliability of the scale is $\lambda_2 = .87$, which is assumed to be sufficient. Secondly, the H coefficient of the Effective

Table 3
H_i and H coefficients and standard errors for the confirmatory TSI-R scales for all groups combined.

Intrusion		Effective suppression		Suppression attempts	
Item	H_i (*SE*)	Item	H_i (*SE*)	Item	H_i (*SE*)
1	.56 (.03)	6	.47 (.04)	8	.36 (.04)
4	.60 (.03)	9	.51 (.03)	11	.30 (.04)
7	.54 (.03)	12	.41 (.04)	14	.40 (.04)
10	.57 (.03)	15	.49 (.03)	17	.35 (.04)
13	.55 (.03)	18	.48 (.04)	20	.41 (.04)
16	.53 (.03)	21	.54 (.03)		
19	.49 (.03)				
H (*SE*)	.55 (.03)	H (*SE*)	.48 (.03)	H (*SE*)	.37 (.03)

Suppression scale is .48, which indicates medium scale strength. The H_i values are somewhat lower than the Intrusion scale and range from .47 (item 6) to .54 (item 21). Additionally, the reliability of the scale is $\lambda_2 = .81$, which is assumed to be sufficient as well. Finally, the Suppression Attempts scale had the lowest H value ($H = .37$), which expresses weak scale strength. Here the H_i values vary between the lower bound value .30 (item 11) to .41 (item 20), which indicates that the IRFs of the items in this scale discriminate less adequately compared to items in the other scales. Furthermore, the reliability of the Suppression Attempts scale is $\lambda_2 = .71$, which is low when the scale is used to measure individual scale levels. Overall, none of the 18 items showed a significant decrease in the item response function. This result indicates that the items within each scale form a unidimensional item set. Overall, the confirmatory MSA verified the three-factor structure found in the exploratory MSA.

3.3. Confirmatory MSA on separate age groups

The results of the confirmatory MSA on separate age groups are shown in Table 4. The H coefficients of the scales show that the strongest scales are found in the youngest age group with H values ranging from .47 (Suppression Attempts) to .59 (Intrusion), indicating strong scales for Intrusions and Effective Suppression, and medium scales for Suppression Attempts. The H values of scales in the older age groups were strong (Intrusion), medium (Effective Suppression), and weak (Suppression Attempts). The H_i values show that there are some problematic items in the middle age and oldest age group in the Suppression Attempts scale. All item H_i values are relatively low for these groups and in particular items 8 (26–50), item 11 (both 26–50, and ≥51), and item 17 (both 26–50, and ≥51) have a bad quality (i.e., these items discriminate poorly). In addition, Table 4 provides the standard errors for each of the item parameters. This gives an indication of the stability of the H coefficients per item. As can be seen, the standard errors of the scale H coefficients are not very large and similar across age groups. Comparing those of the items across the age groups shows that the standard errors for the older age groups are slightly larger than the youngest age group. Looking at both the item parameters and standard error shows that there is no overlap in the H coefficients of the suppression attempt scale of the youngest age group compared to the older two age groups (see also Appendix B). This indicates that there may be measurement invariance for this scale and it should therefore be used with caution in older age groups.

3.4. Additional analyses

To further assess the validity of the TSI-R, mean scale scores were calculated for the factor structure that was used in the confirmatory analyses and were correlated with other study measures in SPSS (version 22) by means of Spearman correlations given the ordinal nature of the measurements (see Table 5). Additionally, TSI-R scale intercorrelations are presented (see Table 6). The mean scores were calculated for all participants ($N = 784$). Overall, both the Intrusions and Suppression Attempts scale correlated positively with psychopathology such as PTSD and OCD, while the Effective Suppression scale was inversely related to psychopathology measures. Interestingly, the WBSI, which intends to measure thought suppression exclusively, indeed correlated with the Effective Suppression scale, but inversely ($\rho = -.37$ to $\rho = -.38$). Moreover, the WBSI displayed the strongest correlations with the Intrusions scale ($\rho = .73$ to $\rho = .79$). To assess the temporal stability of the TSI-R, test-retest correlations were calculated. For the Intrusion scale reliability is good ($\rho = .81$ to $\rho = .84$), for the Effective Suppression scale it is acceptable ($\rho = .73$ to $\rho = .78$), while for the Suppression Attempts scale, it is insufficient for use at an individual level $\rho = .63$ to $\rho = .68$). We would like to stretch that reliability is an important aspect of a test, because it determines how much the observed score is influenced by measurement error. The higher the reliability, the less the observed score consists of measurement error. For individual high-stakes testing a reliability coefficient of approximately .8 is desirable (Streiner, 2003), otherwise test scores may not be accurately interpreted.

Lastly, we investigated if the age groups differed on the TSI-R's subscales. Analysis of Variance (ANOVA) revealed significant effects for intrusions ($F(2, 781) = 10.26$, $p < .001$, $\eta p2 = .03$). Bonferroni corrected post-hoc follow-up showed that only the youngest group displayed more intrusions compared to the other groups ($ps < .001$). For effective suppression, we found a trend ($F(2, 781) = 2.90$, $p = .056$, $\eta p2 = .01$), which remained a trend in follow-up, showing that the youngest group displayed less effective suppression compared to the oldest group ($p = .06$, see Table 7). Note that group differences for the Suppression Attempts scale (and correlations with other questionnaires) have to be interpreted with caution. Group differences were not compared statistically, because there may be a violation of measurement invariance (see Appendix B).

4. Discussion

The aim of this study was to develop and evaluate the structural, convergent/discriminant, and criterion validity of a revised version of the TSI. The TSI-R distinguishes between intrusions, thought suppression attempts, and effective thought suppression. To test if the TSI-R is psychometrically sound across all age groups, and not only in the commonly examined group of students, we included individuals from different age groups and analyzed the data using nonparametric item response theory. The results show that the three unique

Table 4
H_i and H coefficients, standard errors, and reliability estimates (λ_2) for the confirmatory TSI-R scales within the age groups.

H_i (*SE*)											
Intrusion				Suppression attempts				Effective suppression			
Item	≤25	26–50	≥51	Item	≤25	26–50	≥51	Item	≤25	26–50	≥51
1	.59 (.04)	.54 (.06)	.57 (.05)	8	.47 (.07)	.25 (.08)	.32 (.07)	6	.52 (.05)	.49 (.07)	.39 (.09)
4	.63 (.04)	.61 (.04)	.58 (.06)	11	.49 (.06)	.23 (.07)	.12 (.08)	9	.55 (.05)	.52 (.06)	.49 (.06)
7	.58 (.05)	.49 (.06)	.50 (.06)	14	.49 (.06)	.35 (.06)	.35 (.06)	12	.45 (.05)	.45 (.09)	.37 (.08)
10	.61 (.04)	.57 (.06)	.52 (.08)	17	.44 (.06)	.22 (.08)	.32 (.06)	15	.51 (.05)	.54 (.06)	.41 (.07)
13	.60 (.05)	.52 (.07)	.47 (.07)	20	.50 (.05)	.35 (.06)	.34 (.07)	18	.55 (.05)	.42 (.07)	.45 (.06)
16	.57 (.04)	.50 (.07)	.53 (.06)					21	.60 (.05)	.53 (.07)	.51 (.05)
19	.53 (.05)	.48 (.06)	.45 (.06)								
H	.59 (.04)	.53 (.05)	.52 (.05)	H	.47 (.05)	.28 (.05)	.29 (.05)	H	.52 (.04)	.49 (.06)	.44 (.05)
λ_2	.89	.86	.85	λ_2	.79	.63	.66	λ_2	.84	.82	.77

Note. $N_{\leq 25}$ = participant group 25 years and younger; N_{26-50} = participant group between 26 and 50 years; $N_{\geq 51}$ = participant group 51 years and older. $N_{\leq 25} = 176$, $N_{26-50} = 101$, $N_{\geq 51} = 115$.

Table 5
Means (M), standard deviations (SD), and Spearman correlations for all study measures with the confirmatory TSI-R scales for the three age groups.

		≤25			26–50			≥51		
	M (*SD*)	Int	Supp Att	Eff Supp	Int	Supp Att	Eff Supp	Int	Supp Att	Eff Supp
WBSI	42.66 (11.43)	.78**	.47**	−.38**	.79**	.30**	−.37**	.73**	.48**	−.37**
TCAQ	83.68 (14.73)	−.78**	−.25**	.65**	−.76**	−.13*	.62**	−.72**	−.23**	.63**
OCI-R	27.93 (8.64)	.45**	.24**	−.22**	.43**	.07	−.27**	.55**	.21**	−.31**
–Washing	3.70 (1.62)	.13*	.11*	−.002	.12	.04	−.11	.25**	.15*	−.12
–Obsessing	4.82 (2.12)	.68**	.26**	−.43**	.68**	.10	−.41**	.68**	.22**	−.49**
–Hoarding	5.47 (2.44)	.25**	.13*	−.10	.27**	.01	−.18*	.33*	.14*	−.08
–Ordering	5.10 (2.40)	.23**	.10*	−.08	.19**	.08	−.12	.41**	.19*	−.26**
–Checking	5.14 (2.13)	.23**	.18**	−.10	.08	.06	−.07	.33**	.14*	−.18**
–Neutraliz.	3.70 (1.57)	.13*	.08	−.01	.14*	.08	−.07	.18**	.10	−.09
PSS-SR	23.34 (7.47)	.54**	.27**	−.30**	.59**	.20**	−.33**	.57**	.24**	−.35**
–Re-exp.	5.46 (2.17)	.40**	.26**	−.25**	.45**	.15*	−.30**	.43**	.16*	−.26**
–Avoidance	7.72 (2.73)	.50**	.25**	−.23**	.54**	.15*	−.27**	.45**	.22**	−.24**
–Arousal	8.67 (2.97)	.53**	.23**	−.30**	.47**	.18**	−.29**	.50**	.20**	−.35**
MC-SDS	19.17 (5.00)	−.25**	−.09	.24**	−.34**	−.04	.23**	−.39**	−.07	.33**

Note. $N = 784$, except for WBSI ($N = 777$). Statistical significance levels: $^{*}p < .05$, $^{**}p < .001$; WBSI = White Bear Suppression Inventory; TCAQ = Thought Control Ability Questionnaire; OCI-R = Obsessive–Compulsive Inventory-Revised; PSS-SR = PTSD Symptom Scale Self-Report; MC-SDS = Marlowe–Crowne Social Desirability Scale; Int = Intrusion; Sup Att = Suppression Attempts; Eff Sup = Effective Suppression.

constructs of the thought suppression process, as has been found in previous studies, are replicated in the TSI-R in the total sample (Rassin, 2003; Wismeijer, 2012). The confirmatory MSA showed a three-factor model that consisted of the components intrusions, suppression attempts, and effective suppression with strong, weak, and medium scales respectively. Furthermore, confirmatory analyses for the three age groups showed that the hypothesized factor structure is moderately or well represented depending on the specific age group. That is, in the youngest age group, the strongest scale and scale consistency is found. In the middle and old age groups the Intrusion and Effective Suppression scales appear to be sound and usable scales, but caution is required in using the Suppression Attempts scale for these age groups since the scalability of the scale was low.

A possible explanation for stronger scales in the youngest group is the uniformity of this group. These subjects were more alike in terms of age range (16–25) and educational background (mostly psychology students) in comparison to the two older age groups. Moreover, in Wismeijer's (2012) senior citizen population, which is comparable to our oldest group, suppression attempts were also difficult to measure; 4 out of 5 items had unsatisfactory psychometric properties, which may be a result of the aging process. Since general inhibitory control decreases with age, conscious control of unwanted memories may also decline with age (Anderson et al., 2011; Hasher & Zacks, 1988). A potential consequence is that the participants *gradually* start to abandon unsuccessful suppression strategies (i.e., suppression attempts). Perhaps eventually even effective suppression strategies become less successful and stable. This suggests that specifically the construct suppression attempts – but with age possibly also effective suppression – may indeed be more difficult to measure consistently in older populations.

However, suppression attempts could be a difficult construct to measure uniformly in general. As the results showed, this scale only reached medium scale strength in the youngest age group and the measurement invariance may have been violated. This may be explained by the fact that an attempt can be either successful or unsuccessful. Different individuals may alternate between attempts dependent on the outcome of that suppression attempt, or may know that some attempts are unsuccessful and lead to intrusions. This may also explain why in the exploratory MSA suppression attempts items 2 and 5 were consistently clustered with intrusions items; these suppression attempts seem to consistently fail and lead to intrusions.

Regarding further construct validity, the TSI-R showed appropriate convergent and discriminant scale validities that were comparable for the different age groups. Effective suppression correlated positively with a questionnaire measuring thought control specifically, while inverse correlations with this questionnaire were present for the Intrusion and Suppression Attempts scales. Furthermore, there were medium to strong correlations of all TSI-R scales with the WBSI, and remarkably, the Intrusion scale showed the strongest positive correlation with the WBSI, while the Effective Suppression scale was correlated inversely to the WBSI. The latter result not only strengthens the idea that the WBSI specifically measures *failing* suppression attempts, but also that our Effective Suppression scale measures the opposite: attempted suppression that was successful.

Generally, the TSI-R displayed appropriate criterion validity. Nearly all scales of the TSI-R were significantly correlated with PTSD symptom severity, and specifically intrusions and thought suppression attempts were related to more PTSD symptoms, which is in accordance with models of PTSD (e.g., Ehlers & Clark, 2000). However, inconsistent with these models is the observation that effective suppression was related to a *reduction* in PTSD symptomatology.

Table 6
Spearman intercorrelations for the TSI-R scales.

	≤25			26–50			≥51		
	Int	Supp Att	Eff Supp	Int	Supp Att	Eff Supp	Int	Supp Att	Eff Supp
Int	–	–	–	–	–	–	–	–	–
Supp Att	.34**	–	–	.13	–	–	.33**	–	–
Eff Supp	−.47**	.02	–	−.44**	−.003	–	−47**	.07	–

Note. Int = Intrusion; Sup Att = Suppression Attempts; Eff Sup = Effective Suppression.

Table 7
Means and standard errors for the confirmatory TSI-R scales for the three age groups.

	Intrusion	Suppression attempts	Effective suppression
≤25	18.69 (.30)	16.97 (.17)	17.88 (.22)
26–50	16.92 (.37)	15.64 (.23)	18.39 (.27)
≥51	16.97 (.34)	16.01 (.21)	18.65 (.25)

Note. The Intrusion scale consists of 7 items, the Suppression Attempts scale of 5 items, and the Effective Suppression scale of 6 items. $N_{\leq 25} = 351$, $N_{26-50} = 202$, $N_{\geq 51} = 231$.

Thought suppression is mostly considered a maladaptive strategy in dealing with threat, because it is assumed to directly lead to an increase in symptoms (e.g., intrusions; Ehlers & Clark, 2000; Shipherd & Salters-Pedneault, 2008). In fact, these models implicitly seem to consider any form of thought suppression to be maladaptive. Therefore they disregard the idea that successful thought suppression is possible, and potentially beneficial.

Interestingly, our study shows that effective suppression was not only associated with lower symptomatology for PTSD, but also with fewer OCD symptoms even despite the limited range of scores on these questionnaires. Effective suppression indeed seems to be a construct that relates uniquely to intrusions. TSI-R scale intercorrelations specifically show that effective suppression is associated with fewer self-reported intrusions, while suppression attempts were correlated with more intrusions. The relationship between effective suppression and intrusions is also consistent with recent studies on suppression-induced forgetting. In these studies participants successfully forget unwanted memories (e.g., Anderson & Huddleston, 2011) and reduce the number of intrusions of these unwanted memories over time (e.g., Levy & Anderson, 2012). Moreover, a recent experimental study corroborates our correlational findings and shows in a patient sample that failing to forget unwanted memories is indeed related to more severe PTSD symptomatology (Catarino, Küpper, Werner-Seidler, Dalgleish, & Anderson, 2015).

To our surprise, older participants had the fewest intrusions and displayed the most effective suppression. Although this was unforeseen, the scale interrelationships within age groups are expected based on theory (e.g., higher effective suppression with lower intrusions). Why older participants report performing better than younger adults is currently unclear, though explanations can be put forth. A substantial body of research now shows that despite decreases in biological and psychological functioning, people in old-age generally report high levels of emotional well-being, which seems to be a result of changes in emotional processing (see Charles & Carstensen, 2010 for a review). Older adults generally show positivity biases in memory; they focus more attention on positive information compared to negative information, recall positive information better, and tend to interpret negative memories as neutral or positive (Charles, Mather, & Carstensen, 2003; Mather & Carstensen, 2005; Schlagman, Schulz, & Kvavilashvili, 2006).

Perhaps, age-related differences in emotional processing may affect how individuals of different ages deal with unwanted memories. As a consequence, unwanted memories may in part simply pass by unattended, because older individuals pay less attention to these memories or because an intrusive memory is not necessarily unwanted because it is not considered as a negatively valenced memory. This is in line with research showing that older adults demonstrate more self-reported mindfulness compared to younger adults (Hohaus & Spark, 2013), and also with some prevalence studies that show intrusion-related disorders plateau or even decline with age (e.g., Kessler et al., 2005). Perhaps changes in emotional processing are even able to compensate for inhibitory deficits. As a result the number of consciously experienced intrusions decreases. However, in spite of possible declines in inhibitory abilities compared to younger adults, older adults may actually still be more effective in suppressing unwanted memories, because overall they experience fewer negative memories they do not want to think about.

It is currently still largely unclear how age affects thought suppression or even how the theoretical constructs of the thought suppression themselves are exactly related; they could be intertwined, strongly linked empirically, or maybe even parts of the same mechanism. Some sort of relationship between intrusion and suppression attempts may even be expected, as it is a frequently made assumption in theory and in cognitive models. Clark and Rhyno (p.4, 2005) for instance define intrusive thoughts as unintended, recurrent events that are *difficult to control*. Theoretically, this could mean that an intrusion is often accompanied by a suppression attempt that failed. Alternatively, an intrusion could trigger an attempted suppression.

This study has some limitations. Although we found a three-factor solution using self-reports, we did not measure any of those factors behaviorally, though there are sophisticated tasks available to measure these factors (see Benoit et al., 2014; Levy & Anderson, 2012). This would also provide a solution for participants responding in a socially desirable manner, which is an unwanted influence in questionnaire research in general. With respect to this response bias, our study is no exception; there were medium strength correlations between social desirability and some of the scales in our research. Though our study employed a population-wide sample in terms of age, our data was not fully equally distributed over age groups; compared to the youngest group the older groups had lower numbers of observations. Also, even though we used a large stratified sample of Dutch nationals, a larger sample would have further increased the accuracy of our analyses, since MSA, specifically AISP, requires considerable sample sizes (Straat, van der Ark, & Sijtsma, 2014). Finally, even with healthy individuals suppressing everyday intrusive thoughts, our study can be considered as limited because of its relative homogeneity, in the sense that we used healthy participants only. In a more heterogeneous sample, for instance consistent of individuals with psychological problems or patients with PTSD or OCD, it is conceivable that a clearer three-factor structure might be present with items that have higher discriminability. This may also shed some light on those individuals who report intrusive thoughts without suppression attempts or those who report suppression attempts without intrusions.

All in all, the Mokken scale analyses showed that in all three age groups the TSI-R measures three factors of the thought suppression process: intrusions, suppression attempts, and effective suppression. Though, use of the TSI-R is not yet suited for administration in the general population, but is best suited for research in undergraduates, which are frequently used for psychological research and theory building. Further research should aim to address an unexpected finding; why older individuals show less intrusions and more effective suppression compared to younger individuals. Additionally, it should investigate a recurring and still unresolved phenomenon in thought suppression studies: the link between different processes within the thought suppression mechanism, especially how suppressions and intrusions are related. These processes could be strongly intertwined, causally linked, or continuously reinforcing each other. The factor analytical approach used is the common approach to investigate the dimensionality. It sheds light on the underlying factors involved in the mechanism, but it cannot reveal the reciprocal intertwinement of the relationship that is hypothesized by the theory. This would be the next step in future research: revealing this reciprocal intertwinement and investigating the potentially moderating role of age or clinical diagnosis. Accordingly, the specific relationships of these processes should be investigated in populations of different age or clinical status, as it is very likely that these processes and the type of relationship they display, could vary with age or clinical diagnosis.

Acknowledgments

We are grateful to Pim Renkema, Pasqualle Seijkens, Anne Klein Schiphorst, Eva Klaazen, Judith van Velden, Judith Tommel, and Dounia Boujemaaoui for their assistance in testing. We thank Arne Leer and Iris Engelhard for their insightful suggestions and manuscript feedback, and Lianne Keemink and Ethan Alessandro for proofreading.

Appendix A

Internal consistency reliability estimates (λ_2) for WBSI, TCAQ, OCI-R, PSS-SR and MC-SDS for the three age groups.

	≤25	26–50	≥51
WBSI	.92	.92	.92
TCAQ	.92	.93	.92
OCI-R	.88	.87	.88
–Washing	.80	.74	.78
–Obsessing	.83	.75	.71
–Hoarding	.73	.70	.82
–Ordering	.84	.85	.79
–Checking	.66	.73	.72
–Neutralizing	.74	.73	.68
PSS-SR	.92	.90	.90
–Re-experiencing	.82	.81	.80
–Avoidance	.81	.77	.82
–Arousal	.80	.71	.71
MC-SDS	.72	.76	.71

Note. $N = 784$, except for WBSI ($N = 777$; WBSI = White Bear Suppression Inventory; TCAQ = Thought Control Ability Questionnaire; OCI-R = Obsessive–Compulsive Inventory-Revised; PSS-SR = PTSD Symptom Scale Self-Report; MC-SDS = Marlowe–Crowne Social Desirability Scale.

Appendix B

At this moment MSA does not provide a direct way to compare H_i coefficients and test for measurement invariance. A potential, but admittedly not perfect, solution is to calculate a confidence interval for each item. If item confidence intervals for the three groups overlap, there is an indication for measurement invariance, which shows that the same construct is measured in these groups. If one or more of the confidence intervals does not overlap with the other(s), there *may* be a violation of measurement invariance, which may invalidate that particular item's scale. This is currently only the case for item 11. Therefore, mean group differences for the Suppression Attempts scale were not compared statistically.

Confidence intervals of H_i coefficients and H coefficients for the confirmatory TSI-R scales within the age groups.

Item	≤25	26–50	≥51
Intrusion			
1	.59 [.51, .67]	.54 [.42, .66]	.57 [.47, .67]
4	.63 [.56, .70]	.61 [.53, .69]	.58 [.46, .70]
7	.58 [.49, .67]	.49 [.36, .62]	.50 [.38, .62]
10	.61 [.53, .69]	.57 [.46, .68]	.52 [.36, .67]
13	.60 [.50, .69]	.52 [.39, .65]	.47 [.33, .61]
16	.57 [.49, .65]	.50 [.37, .63]	.53 [.41, .65]
19	.53 [.44, .62]	.48 [.37, .59]	.45 [.33, .57]
Scale	.59 [.52, .66]	.53 [.44, .62]	.52 [.42, .62]
Suppression attempts			
8	.47 [.34, .60]	.25 [.09, .41]	.32 [.18, .46]
11	.49 [.38, .60]	.23 [.09, .36]	.12 [−.04, .28]
14	.49 [.37, .61]	.35 [.23, .47]	.35 [.24, .46]
17	.44 [.33, .55]	.22 [.06, .38]	.32 [.20, .44]
20	.50 [.40, .60]	.35 [.23, .47]	.34 [.20, .48]
Scale	.48 [.38, .58]	.28 [.17, .38]	.29 [.20, .38]
Effective suppression			
6	.52 [.42, .62]	.49 [.36, .62]	.39 [.22, .56]
9	.55 [.45, .65]	.52 [.40, .64]	.49 [.38, .60]
12	.45 [.35, .55]	.45 [.27, .63]	.37 [.22, .53]
15	.51 [.41, .61]	.54 [.42, .66]	.41 [.27, .54]
18	.55 [.44, .66]	.42 [.28, .56]	.45 [.34, .56]
21	.60 [.51, .69]	.53 [.39, .67]	.51 [.41, .61]
Scale	.53 [.45, .61]	.49 [.37, .61]	.43 [.33, .53]

Note. Confidence intervals are presented within brackets.

References

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*, 366–369. http://dx.doi.org/10.1038/35066572.

Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences*, *18*(6), 279–292. http://dx.doi.org/10.1016/j.tics.2014.03.002.

Anderson, M. C., & Huddleston, E. (2011). Towards a cognitive and neurobiological model of motivated forgetting. In R. F. Belli (Ed.), *Nebraska symposium on motivation. True and false recovered memories: Toward a reconciliation of the debate*, *Vol. 58*. (pp. 53–120). New York, NY: Springer. http://dx.doi.org/10.1007/978–1-4614-1195-6_3.

Anderson, M. C., Reinholz, J., Kuhl, B. A., & Mayr, U. (2011). Intentional suppression of unwanted memories grows more difficult as we age. *Psychology and Aging*, *26*(2), 397. http://dx.doi.org/10.1037/a0022505.

van der Ark, L. A. (2012). New developments in Mokken scale analysis in R. *Journal of Statistical Software*, *48*(5), 1–27.

Benoit, R. G., Hulbert, J. C., Huddleston, E., & Anderson, M. C. (2014). Adaptive top-down suppression of hippocampal activity and the purging of intrusive memories from consciousness. *Journal of Cognitive Neuroscience*. http://dx.doi.org/10.1162/jocn_a_00696.

Biss, R. K., Campbell, K. L., & Hasher, L. (2013). Interference from previous distraction disrupts older adults' memory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *68*(4), 558–561.

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

Brewin, C. R., Christodoulides, J., & Hutchinson, G. (1996). Intrusive thoughts and intrusive memories in a nonclinical sample. *Cognition and Emotion*, *10*, 107–112. http://dx.doi.org/10.1080/026999396380411.

Catarino, A., Küpper, C. S., Werner-Seidler, A., Dalgleish, T., & Anderson, M. C. (2015). Failing to forget: Inhibitory-control deficits compromise memory suppression in posttraumatic stress disorder. *Psychological Science*, 1–3. http://dx.doi.org/10.1177/0956797615569889.

Charles, S., & Carstensen, L. L. (2010). Social and emotional aging. *Annual Review of Psychology*, *61*, 383. http://dx.doi.org/10.1146/annurev.psych.093008.100448.

Charles, S. T., Mather, M., & Carstensen, L. L. (2003). Aging and emotional memory: the forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General*, *132*(2), 310. http://dx.doi.org/10.1037/0096–3445.132.2.310.

Clark, D. A., & Rhyno, S. (2005). Unwanted intrusive thoughts in nonclinical individuals: Implications for clinical disorders. In D. A. Clark (Ed.), *Intrusive thoughts in clinical disorders: Theory, Research, Treatment* (pp. 1–29). New York: Guilford Press.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349–354. http://dx.doi.org/10.1037/h0047358.

Depue, B. E., Curran, T., & Banich, M. T. (2007). Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. *Science*, *317*(5835), 215–219. http://dx.doi.org/10.1126/science.1139560.

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy*, *38*, 319–345. http://dx.doi.org/10.1016/S0005-7967(99)00123–0.

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. http://dx.doi.org/10.1037/1040–3590.14.4.485.

Foa, E. B., Riggs, D. S., Dancu, C. V., & Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress*, *6*(4), 459–473. http://dx.doi.org/10.1002/jts.2490060405.

Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics: An introduction* (2nd ed.). Sage.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation*, *22*, 193–225. http://dx.doi.org/10.1016/S0079-7421(08)60041–9.

Hemker, B. T., Sijtsma, K., & Molenaar, I. W. (1995). Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. *Applied Psychological Measurement*, *19*(4), 337–352. http://dx.doi.org/10.1177/014662169501900404.

Hohaus, L. C., & Spark, J. (2013). 2672–Getting better with age: do mindfulness & psychological well-being improve in old age? *European Psychiatry*, *28*(1). http://dx.doi.org/10.1016/S0924-9338(13)77295-X.

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*(6), 1049–1055. http://dx.doi.org/10.1016/S0191-8869(02)00089–2.

Julien, D., O'Connor, K. P., & Aardema, F. (2007). Intrusive thoughts, obsessions, and appraisals in obsessive–compulsive disorder: A critical review. *Clinical Psychology Review*, *27*, 366–383. http://dx.doi.org/10.1016/j.cpr.2006.12.004.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593–602. http://dx.doi.org/10.1001/archpsyc.62.6.593.

Levy, B. J., & Anderson, M. C. (2012). Purging of memories from conscious awareness tracked in the human brain. *The Journal of Neuroscience*, *32*(47), 16785–16794. http://dx.doi.org/10.1523/JNEUROSCI.2640–12.2012.

Loevinger, J. (1948). The technic of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychological Bulletin*, *45*, 507–529.

Luciano, J. V., Algarabel, S., Tomás, J. M., & Martínez, J. L. (2005). Development and validation of the thought control ability questionnaire. *Personality and Individual Differences*, *38*(5), 997–1008. http://dx.doi.org/10.1016/j.paid.2004.06.020.

Luciano, J. V., Belloch, A., Algarabel, S., Tomás, J. M., Morillo, C., & Lucero, M. (2006). Confirmatory factor analysis of the white bear suppression inventory and the thought control questionnaire: A comparison of alternative models. *European Journal of Psychological Assessment*, *22*(4), 250–258. http://dx.doi.org/10.1027/1015–5759.22.4.250.

Lustig, C., Hasher, L., & Tonev, S. T. (2001). Inhibitory control over the present and the past. *European Journal of Cognitive Psychology*, *13*, 107–122. http://dx.doi.org/10.1080/09541440126215.

Magee, J. C., Harden, K. P., & Teachman, B. A. (2012). Psychopathology and thought suppression: A quantitative review. *Clinical Psychology Review*, *32*(3), 189–201. http://dx.doi.org/10.1016/j.cpr.2012.01.001.

Mather, M., & Carstensen, L. L. (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in cognitive sciences*, *9*(10), 496–502. http://dx.doi.org/10.1016/j.tics.2005.08.005.

Mokken, R. J. (1971). *A theory and procedure of scale analysis: With applications in political research*, *Vol. 1*, Walter de Gruyter.

Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement*, *6*, 417–430. http://dx.doi.org/10.1177/014662168200600404.

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. http://dx.doi.org/10.1016/0005–7967(96)00005–8.

Pica, G., Pierro, A., & Giannini, A. (2014). The relationship between thought suppression and retrieval-induced forgetting: an analysis of witness memories. *Cognitive Processing*, 1–10. http://dx.doi.org/10.1007/s10339-014-0626-0.

Purdon, C. (1999). Thought suppression and psychopathology. *Behaviour Research and Therapy*, *37*(11), 1029–1054. http://dx.doi.org/10.1016/S0005-7967(98)00200–9.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rachman, S. (1998). A cognitive theory of obsessions: Elaborations. *Behaviour Research and Therapy*, *36*(4), 385–401. http://dx.doi.org/10.1016/S0005-7967(97)10041–9.

Rachman, S., & de Silva, P. (1978). Abnormal and normal obsessions. *Behaviour Research and Therapy*, *16*, 233–248. http://dx.doi.org/10.1016/0005–7967(78)90022–0.

Rassin, E. (2003). The white bear suppression inventory (WBSI) focuses on failing suppression attempts. *European Journal of Personality*, *17*(4), 285–298. http://dx.doi.org/10.1002/per.478.

Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item response theory: Fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science*, *14*, 95–101. http://dx.doi.org/10.1111/j.0963–7214.2005.00342.x.

Rodríguez, M. G., Delgado, P. A., Rovella, A. T., & Cubas León, R. (2008). Structural validity and reliability of the Spanish version of the white bear suppression inventory (WBSI) in a sample of the general Spanish population. *The Spanish Journal of Psychology*, *11*(2), 650–659. http://dx.doi.org/10.1017/S1138741600004650.

van Schie, K., Geraerts, E., & Anderson, M. C. (2013). Emotional and non-emotional memories are suppressible under direct suppression instructions. *Cognition and Emotion*, *27*(6), 1122–1131. http://dx.doi.org/10.1080/02699931.2013.765387.

Schlagman, S., Schulz, J., & Kvavilashvili, L. (2006). A content analysis of involuntary autobiographical memories: Examining the positivity effect in old age. *Memory*, *14*(2), 161–175. http://dx.doi.org/10.1080/09658210544000024.

Schmidt, R. E., Gay, P., Courvoisier, D., Jermann, F., Ceschi, G., David, M., ... Van der Linden, M. (2009). Anatomy of the white bear suppression inventory (WBSI): A review of previous findings and a new approach. *Journal of Personality Assessment*, *91*(4), 323–330. http://dx.doi.org/10.1017/S1138741600004650.

Shipherd, J. C., & Salters-Pedneault, K. (2008). Attention, memory, intrusive thoughts, and acceptance in PTSD: an update on the empirical literature for clinicians. *Cognitive and Behavioral Practice*, *15*, 349–363. http://dx.doi.org/10.1016/j.cbpra.2008.01.003.

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage.

Spinhoven, P., & Van der Does, A. J. W. (1999). Thought suppression, dissociation and psychopathology. *Personality and Individual Differences*, *27*(5), 877–886. http://dx.doi.org/10.1016/S0191-8869(99)00037–9.

Straat, J. H., van der Ark, L. A., & Sijtsma, K. (2014). Minimum sample size requirements for Mokken scale analysis. *Educational and Psychological Measurement*, *74*(5), 809–822. http://dx.doi.org/10.1177/0013164414529793.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103. http://dx.doi.org/10.1207/S15327752JPA8001_18.

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

Wenzlaff, R. M. (2005). Seeking solace but finding despair: The persistence of intrusive thoughts in depression. In D. A. Clark (Ed.), *Intrusive thoughts in clinical disorders* (pp. 54–85). New York: Guilford.

Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology*, *51*(1), 59–91. http://dx.doi.org/10.1146/annurev.psych.51.1.59.

Wismeijer, A. A. J. (2012). Dimensionality analysis of the thought suppression inventory: Combining EFA, MSA, and CFA. *Journal of Psychopathology and Behavioral Assessment*, *34*(1), 116–125. http://dx.doi.org/10.1007/s10862-011-9246-5.