J Psychopathol Behav Assess (2012) 34:116–125
DOI 10.1007/s10862-011-9246-5

Dimensionality Analysis of the Thought Suppression Inventory: Combining EFA, MSA, and CFA

Andreas A. J. Wismeijer

Published online: 12 August 2011

Abstract The Thought Suppression Inventory (TSI; Rassin, European Journal of Personality 17: 285-298, 2003) was designed to measure thought intrusion, thought suppression and successful thought suppression. Given the importance to distinguish between these three aspects of thought control, the aim of this study was to scrutinize the dimensionality of the TSI. In a sample of 333 Dutch senior citizins, we examined (1) the dimensionality of the TSI using various procedures such as PAF, Mokken scale analysis (MSA) and CFA, and (2) the scale properties of the TSI. PAF favored a two factor solution, however, MSA and CFA suggested that three dimensions most adequately capture the structure of the TSI. Although all scales obtained at least medium scalability coefficients, several items were identified that are psychometrically unsound and may benefit from rewording or replacement. The findings suggest that the TSI is a three-dimensional questionnaire as originally proposed by Rassin (European Journal of Personality 17: 285-298, 2003) measuring thought intrusion, thought suppression, and successful thought suppression.

Keywords Thought suppression · Thought intrusion · Dimensionality · Mokken scale analysis · Confirmatory factor analysis

Intrusive thoughts, or experiencing unwanted thoughts that are difficult to control, is an important theme in the clinical psychological literature. Not only because intrusive thoughts can place a burden on one's cognitive and emotional functioning, but in particular because experiencing intrusive thoughts is a hallmark of many psychopathologies such as anxiety disorders, obsessions and depression (Julien et al. 2007; Shipherd and Salters-Pedneault 2008).

Early work by Wegner et al. (1987) and Wegner (1992, 1994) showed that attempts to suppress thoughts may have the paradoxical effect of increasing the number of intrusive thoughts one is having (rebound effect). This increase in intrusive thoughts may persist after suppression is no longer actively pursued and is even thought to cause undesirable outcomes such as obsessions (Lane and Wegner 1995). Although this ironic consequence of unsuccessful thought suppression has repeatedly been documented in experimental settings using predominantly neutral stimuli (such as the infamous white bear), empirical support for a ubiquitous negative role of suppression and its rebound effect in the persistence of thoughts has been mixed. For example, several factors have been identified that may limit the generalizability of suppression research such as whether the material to be suppressed is positive, negative, or neutral and self-relevant (Salkovkis and Campbell 1994). In addition, unsuccessful suppression of obsessive thoughts in particular may lead to an increase in intrusive thoughts, and unsuccessful suppression of neutral or positive thoughts are considerably less affected by Wegner's (1994) thought suppression paradox (Purdon and Clark 2001). Finally, successful thought suppression has also been documented (e.g., Brewin and Beaton 2002), in particular when one has positive beliefs about having control over one's thoughts (Wenzlaff and Wegner 2000). Given this unresolved debate on whether thought suppression is a successful or unsuccessful mechanism of thought control, it is important that instruments that aim to assess thought suppression distinguish successful from unsuccessful thought suppression in order to better understand the

A. A. J. Wismeijer (✉)
Department of Clinical Psychology, Tilburg University,
PO Box 90153, 5000 LE Tilburg, The Netherlands
e-mail: a.a.j.wismeijer@uvt.nl

dynamics of thought suppression as thought control mechanism.

Wegner and Zanakos (1994) developed a unidimensional thought suppression measure to identify individuals that use thought suppression as the default thought control strategy and are therefore more susceptible to experience intrusive thoughts. This measure, the White Bear Suppression Inventory (WBSI), is since its development the most widely used instrument to assess thought suppression. However, several authors suggest the WBSI also measures the experience of intrusive thoughts (e.g., Höping and De Jong-Meyer 2003). Rassin (2003) points out the importance to distinguish thought suppression and thought intrusion as the frequency of thought intrusions that one experiences may not only be the result of unsuccessful thought suppression but also the *cause* for thought suppression in the first place. In other words, it is unclear how to interpret a high WBSI score: as a measure of thought suppression or of the number of intrusive thoughts.

To address these limitations of the WBSI, Rassin (2003) developed the Thought Suppression Inventory (TSI) to assess thought suppression attempts as well as thought intrusion and successful thought suppression. The TSI is a brief self-report measure that consists of 15 items that contribute to three 5-item subscales measuring thought intrusion (Int), thought suppression (Sup) and successful (effective) thought suppression (in this study abbreviated as "Eff" to better distinguish from Sup). Table 1 (first column) shows the items and how they are distributed among the three scales. In addition to its dimensionality, Rassin also reported satisfactory test-retest reliability for Int and Eff (a lower than desired test-retest reliability was found for Sup, given its supposed trait-like character) and external validity in two student samples. Given the earlier mentioned importance to adequately disentangle Int, Sup, and Eff, the aim of this study is to investigate the factorial structure of the TSI using PAF, CFA and the less well-known but more sophisticated Mokken scale analysis (MSA). To avoid on one side the significantly distorted frequency distribution and skew often present in data sampled from a clinical population and the low generalizability to clinical populations of student data on the other side, a senior citizen sample was used for this study. Epidemiological studies show that community-dwelling elderly experience substantial rates of relatively common mental disorders related to the occurrence of intrusive thoughts such as depression, obsessive-compulsive disorder, and generalized anxiety disorder (Kessler et al. 2005). Hence, where the occurrence of thought intrusion is concerned, senior citizen data (in comparison with student data) may be more representative for a clinical population, without being subject to significantly distorted answering tendencies that are indicative of a clinical population.

Method

Participants and Procedure

The sample consisted of 333 senior citizens of which 153 were men (46%) and 180 were women (54%). Men (M=66.45, SD=9.71) were on average slightly older than women (M=63.80, SD=9.32) [$t(331)=2.54$, $p=0.012$, two-tailed]. The participants were sampled from the ProAging panel in the Netherlands, a large internet-based panel of senior citizens. Senior citizens may register voluntarily to the ProAging panel and are asked to complete a limited number of online questionnaires yearly. After having completed a questionnaire the respondents have the option to close the questionnaire, or to modify it at a later time. They also have the option to stop the completion temporarily and to continue at a later time. The respondents are guaranteed anonymity and do not receive financial compensation for participating.

Measures

The Thought Suppression Inventory (TSI; Rassin 2003) is a 15-item self-report measure, aimed to assess three dimensions of thought suppression: 1) intrusion (Int), 2) suppression attempts (Sup), and 3) successful suppression (Eff). Each subscale consists of 5 items that are answered on a 5-point Likert scale running from 1 to 5 (1 = *strongly disagree*; 2 = *disagree*; 3 = *neutral*; 4 = *agree*; and 5 = *strongly agree*). Total-score ranges run from 5 to 25 for each subscale. All items are positively worded with respect to the construct of interest, so that higher total scale scores suggest higher levels of Int, Sup or Eff. Example items are "I have thoughts which I would rather not have" (7Int), "I try to avoid certain thoughts" (14Sup), and "I am able to put aside problems and worries" (9Eff). We used the original Dutch version of the TSI which was developed and validated by Rassin in three samples of undergraduate students aged 18-42, totaling 350 individuals. Rassin reported absolute inter-factor correlations between .21 and .25 (all $p<0.01$) and Cronbach's alphas of .71, .64 and .67. In addition, three-week test-retest reliability was .80 for Int, .43 for Sup, and .83 for Eff.

Data Analytic Strategy

Rassin's (2003) study is the only available study that presented data on the TSI, using PCA with Varimax rotation. However, psychological data often violate the assumption of multivariate normality. Hence, it was decided to start with exploratory factor analysis using principal axis factoring (PAF), using IBM SPSS Statistics 19 (2010). PAF is one of the most widely used and recommended methods for factor extraction of data that severely violate

Table 1 Item content, pattern coefficients of the 2-factor model and the 3-factor model after oblimin rotation using principal axis factoring, and results of the automated item selection procedure in exploratory factor analysis

Item	Pattern coefficients of 2-factor model		Pattern coefficients of 3-factor model			Selected clusters from AISP, item coefficients H_j, total scale coefficients H per cluster										
	F1	F2	F1	F2	F3	c=.0		c=.3			c=.4			c=.5		
						Cl.1	Cl.2	Cl.1	Cl.2	Cl.3	Cl.1	Cl.2	Cl.3	Cl.1	Cl.2	Cl.3
1.I have many unpleasant thoughts (I)	.76	-.08	.72	-.14	-.02	.38		.46			.52			.55		
4. I experience many emotions that are too intense to control (I)	.55	-.13	.54	-.12	.06	.27		.33			us	us	us	us	us	us
7.I have thoughts which I would rather not have(I)	.75	.06	.75	-.01	-.05	.43		.50			.54			.56		
10.I regularly 'hear' unexplainable things inside my head, such a s my own voice, or the voices of people who are not present(I)	.39	-.08	.44	.01	.12	.25		.32			us	us	us	us	us	us
13.I am unable to concentrate(I)	.53	-.01	.45	-.15	-.13	.30		.34			.41			us	us	us
2.I always try to forget unpleasant events as fast as possible(S)	.27	.54	.08	.06	-.63	.25			.41			.53			.57	
5.I always try to put problems out of mind(S)	.14	.57	-.13	.01	-.77		.25		.45			.49			.57	
8.If I have an unpleasant thought, I try to think about something else(S)	.33	.59	.34	.36	-.31	.32		.29				.44		us	us	us
11.Sometimes I stay busy just to prevent having certain thoughts(S)	.64	.25	.70	.18	-.08	.41		.45			.49			.53		
14.I try to avoid certain thoughts(S)	.62	.36	.47	-.02	-.44	.42		.43			.47			.50		
3.I am able to keep a problem out of mind until I have time to deal with it(E)	-.27	.50	.01	.72	.14		.36			.45			.51			.57
6.I am able to suppress unpleasant thoughts(E)	-.24	.75	-.07	.69	-.17		.46		.37				.53			.57
9.I am able to put aside problems and worries(E)	-.43	.50	-.23	.59	.02		.36			.45			.47	us	us	us
12-I am able to suppress unpleasant experiences to the point that I hardly remember them(E)	.11	.46	.12	.29	-.25	.16			.35		us	us	us	us	us	us
15.I can keep the lid on my feelings if letting them out would interfere with what I am doing(E)	-.12	.39	-.04	.35	-.10		.28	us	us	us	us	us	us	us	us	us
Scale H					.32	.34	.39	.40	.45	.40	.49	.50	.54	.57	.57	

I Intrusion, *S* Suppression, *E* Successful Suppression, *Cl.* Cluster, *us* unscalable the assumption of multivariate normality (Costello and Osborne 2005; Fabrigar et al. 1999). In addition, the Scree test and Kaiser's criterion (also known as the 'eigenvalue-greater-than-1-rule') were used to determine the amount of factors to be extracted. However, as the dimensionality of the instrument was the main focus of the present study, in addition two alternative methods of factor retrieval were used so as to rule out inconsistency in results due to the particular method chosen: parallel analysis (PA) and Velicer's minimum average partial (MAP) analysis (e.g., Zwick and Velicer 1986). PA and MAP tend to result in the same decision about the number of factors to retain but when they disagree the procedures complement each other: When MAP errs it tends to underextract, and when PA errs it tends to overextract (O'Connor 2000, p. 398). Therefore, it is generally recommended to use PA and MAP simultaneously (O'Connor 2000). As the factors were likely to correlate, Oblimin rotation was used. Note that, if the first eigenvalue is large, which typically occurs in oblique factor structures, PA results in underextraction (Beauducel 2001). Both PA and MAP were run using the SPSS macros of O'Connor (2000). For PA, 1,000 randomly generated datasets were used.

Second, exploratory Mokken scale analysis (MSA) was executed to investigate the dimensionality and the monotonicity of the TSI items using MSPWIN 5.0 (Molenaar and Sijtsma 2000). MSA is a method from item response theory (Embretson & Reise, 2000; Van der Linden and Hambleton 1997) and offers a technique for exploring and testing hypotheses about dimensionality. MSA can be used to identify one or more dimensions in the data and does this in such a way that the items selected in one cluster satisfy a measurement model known as the monotone homogeneity model (MHM; Mokken and Lewis 1982; Sijtsma and Molenaar 2002). This non-parametric item response model implies that persons can be ordered on a scale using the items in a selected cluster. Thus, MSA provides a method for dimensionality investigation and a measurement model in one technique (for a practically orientated introduction to MSA see Wismeijer et al. 2008 and Emons et al. 2010). Dimensionality was investigated using MSPWIN's automated item selection procedure (AISP) that aims to find unidimensional clusters of items. Clusters were identified running consecutive AISPs with increasing lower bound scalability criteria. Monotonicity, that is, the higher a respondent's disposition on the latent trait the more likely it is that (s)he obtains higher scores on the items measuring that latent trait, was assessed by testing observed decreases in item response functions (IRFs) for significance in MSPWIN.

Third, confirmatory MSA was executed using MSPWIN for the three-factor solution suggested by Rassin and for alternative solutions should these be ad hoc proposed by EFA and/or exploratory MSA. Confirmatory MSA assesses the scalability of item sets that the researcher a priori defined as scales, and uses scalability coefficient H for this purpose (Sijtsma and Molenaar 2002, chap. 4). Different intervals of H values correspond with different qualifications of the quality of the scale: The higher the H value, the more accurate the ordering of person on the scale by means of their total scores. Positive values of H are interpreted using the following rules of thumb: if $H<.3$ the itemcluster is considered unscalable for practical purposes, $.3\leq H\leq.4$ indicates a weak scale, $.4\leq H\leq.5$ a medium scale, and $H\geq.5$ indicates a strong scale (Sijtsma and Molenaar 2002, p. 60).

Fourth, CFA was executed using AMOS 7.0 (Arbuckle 2006) to test Rassin's three-dimensional solution and, as with confirmatory MSA, to test ad hoc proposed alternative models. Latent modeling allows for controlling effects of measurement error (Byrne 2001) which is thought to be present in most psychological data, in particular when psychopathological processes (or subclinical levels such as assessed with the TSI) are assessed (Thornton and Gilden 2005). The RMSEA, CFI, TLI, and GFI were used to assess goodness-of-fit of the factorial model. The RMSEA was used as the main fit index. RMSEA values are interpreted as follows: RMSEA smaller than .05 indicates good fit, ranging from .05 to .08 reasonable fit, .08 to .10 medium fit, and larger than .10 poor fit (Byrne 2001, pp. 84-85). Values of CFI, TLI, and GFI that exceed .9 were interpreted as indicating adequate model fit (Hu and Bentler 1999).

Finally, the subscale properties are provided for the scales of both the three-factor solution and of possible alternative models. Four estimates of scale-score reliability are reported: Cronbach's alpha, Guttman's lamba2, the greatest lower bound (GLB; computed using MRFA2) and the total-scale coefficient H (using MSPWIN). There were no missing values.

Results

Preliminary Analyses

The items and their distribution among the scales are shown in the first column of Table 1. Women scored higher than men on Int and Sup; Cohen's d equalled .42 for Int and .43 for Sup, which represent small to medium effect sizes (Cohen 1988). Hence, all analyses were also executed separately for men and women. However, as the results were almost identical for men and women (and the differences did not appear to be systematic), only the results for the total sample are presented in the main text. Summaries of the principal analyses by gender are given in Table 2 and Table 3. The scale scores are presented in the fourth row from below of Table 2.

Table 2 Confirmatory MSA (Hjk and H values) for the 2-factor model and the 3-factor model for the total sample and men and women separately

Item	Mean		2-factor model				3-factor model					
			Intrusion		General suppression		Intrusion		Suppression		Effective suppression	
			Hj		Hj		Hj		Hj		Hj	
	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀
Int1	2.68		.49				.49					
	2.50	2.83	.53	.42			.53	.42				
Int4	2.69		.40				.40					
	2.45	2.89	.43	.33			.43	.33				
Int7	3.11		.53				.53					
	2.90	3.29	.52	.50			.52	.50				
Int10	1.62		.32				.32					
	1.56	1.68	.38	.26			.38	.26				
Int13	2.56		.37				.37					
	2.41	2.69	.39	.33			.39	.33				
Sup2	3.22				.32				.44			
	3.18	3.24			.31	.32			.52	.35		
Sup5	2.95				.32				.38			
	2.84	3.06			.29	.32			.42	.31		
Sup8	3.41				.36				.47			
	3.20	3.58			.34	.36			.50	.38		
Sup11	2.76				.19				.38			
	2.52	2.96			.25	.12			.46	.27		
Sup14	3.07				.24				.47			
	2.84	3.26			.25	.21			.53	.37		
Eff3	2.83				.22						.40	
	2.84	2.83			.25	.20					.47	.33
Eff6	2.92				.33						.45	
	2.88	2.96			.33	.33					.50	.40
Eff9	2.85				.18						.37	
	3.01	2.72			.20	.20					.43	.31
Eff12	2.44				.29						.26	
	2.35	2.52			.31	.26					.32	.22
Eff15	3.47				.21						.28	
	3.36	3.57			.19	.21					.35	.21
Scale Mean (SD)			12.66 (3.77)		29.93 (5.61)		12.66 (3.77)		15.40 (3.61)		14.53(3.34)	
♂ ♀			11.82 (3.87)	13.38 (3.54)	29.01 (5.82)	30.71 (5.33)	11.82 (3.87)	13.38 (3.54)	14.58 (3.87)	16.10 (3.21)	14.44 (3.57)	14.61 (3.14)
Scale *H*			.42		.26		.42		.42		.35	
			.45	.37	.27	.25	.45	.37	.49	.33	.42	.30

Exploratory Factor Analysis

PCA yielded a first component which explained 25.4% of the total variance (eigenvalue $\lambda_1=3.82$), a second component which explained 20.9% ($\lambda_2=3.14$) and a third component (eigenvalue $\lambda_3=1.04$) which explained 6.9% of the total variance, totalling 53.3% of explained variance. The other components had eigenvalues smaller than 1. The scree plot showed a sharp bend as off λ_3, suggesting that only the first two components should be retained. Kaiser's criterion identified three components to be retained, although it must be noted that the third eigenvalue is only just larger than 1. In addition, the MAP and PA results suggested retaining two components. Although these results tend to favor a two-factor solution, there was also some evidence in favor of the three-factor solution originally intended by Rassin (2003). Hence, both solutions were investigated using PAF. The pattern coefficients of the hypothesized three-factor solution following Oblimin rotation (see Table 1, colums 4-6) revealed a solution that closely resembled Rassin's three factors. One exception was item 11Sup, that had its highest loading on a factor further consisting entirely of Int items. In addition, items 8Sup, 12Ef, and 14Sup loaded on more than one factor with loadings that differed less than .10 and therefore did not discriminate sufficiently between the factors.

Table 3 Factor structure and fit indices (RMSEA, CFI, TLI and GFI) of Confirmatory Factor Analysis (CFA) for the 2-factor model and the 3-factor model for the total sample and for men and women separately.

Number of factors	Scale labels	Items	CFA				Confirmatory MSA	Scale-score reliability		
			RMSEA (♂/♀)	CFI (♂/♀)	TLI (♂/♀)	GFI (♂/♀)	H (Rho) [$H_♂/H_♀$(Rho$_♂$/Rho$_♀$)]	α (♂/♀)	λ2 (♂/♀)	GLB (♂/♀)
2	Intrusion	1,4,7,10,13	.11 (.13/.10)	.77 (.72/.75)	.71 (.66/.70)	.85 (.55/.85)	.42 (.76) [.45/.37(.78/.72)]	.75 (.77/.70)	.76 (.78/.71)	.77 (.80/.74)
	General Suppression	2,3,5,6,8,9,11,12,14,15					.26 (.76) [.27/.25(.78/.73)]	.75 (.77/74)	.77 (.79/75)	.86 (.88/.86)
3	Intrusion	1,4,7,10,13	.08 (.08/.11)	.87 (.88/.71)	.83 (.86/.65)	.90 (.86/.82)	.42 (.76) [.45/.37(.78/.72)]	.75 (.77/.70)	.76 (.78/.71)	.77 (.80/.74)
	Suppression	2,5,8,11,14					.42 (.75) [.49/.33(.81/.68)]	.75 (.80/.68)	.76 (.80/.69)	.82 (.84/.79)
	Effective Suppression	3,6,9,12,15					.35 (.70) [.42/.30(.76/.65)]	.69 (.76/.63)	.70 (.77/.65)	.74 (.80/.70)

The two-factor solution (second and third column in Table 1) resulted in one factor that was more difficult to interpret and consisted of all Int items and 2 Sup items, and a second factor consisting of the remaining three Sup items and all Eff items. Hence, EFA suggested either a two-factor solution consisting of an Intrusion factor and a General Suppression factor of combined Sup and Eff items or a three-factor solution that, in addition to an Intrusion-factor, distinguished between a Suppression-factor and an Effective Suppression-factor. [1]

Exploratory Mokken Scale Analysis

An AISP was executed for the entire group of subjects and for men and women separately, starting with lower bound c=0 and progressively increasing c with steps of .05 in each next analysis, until c=.6, following advice from Hemker et al. (1995). Table 1 (seventh column and further) shows the results for the entire sample for lower bounds c=.0, .3, .4, and .5 (other c values are excluded because these results are highly similar to the ones shown). For c=0, two scales were found: The first scale consisted of 10 items (5 Int, 4 Sup and 1 Eff) , the second scale of the remaining 5 items (1 Sup and 4 Eff). For c=.3, three scales were found. The first scale consisted of 8 items (5 Int and 3 Sup), the second scale of 4 items (2 Eff and 2 Sup), and the third scale consisted of 2 Eff items. One item (15Eff) was unscalable. For c=.4 also three scales were found. The first scale lost one Int item, the second scale lost 2 Eff items but gained one Sup item resulting in 3 Sup items, and the third scale gained an additional Eff item. In addition, 4 more items proved unscalable (2 Int and 2 Eff). Finally, for c=.5 again three scales were found. The first scale lost an additional Int item, the second scale an Sup item and the third scale a Eff item. Three more items (one from each scale) were unscalable. Inspection of the cluster pattern obtained across the different c values (following guidelines from Sijtsma and Molenaar 2002, p. 81) suggested that the TSI consists of one strong scale (H=.50) of three Eff items, one medium scale (H=.49) of three Sup items, and one strong scale (H=.70) of two Int items. The analyses for men and women separately showed several differences (43 over all c levels together) with respect to how items were scaled. However, in the majority of cases (35) this diffence

[1] For men, the eigenvalues larger than 1 were 4.25, 3.15, and 1.09 and for women 3.45, 3.10, and 1.11. For men, the hypothesized three-factor solution was almost completely replicated using PAF with Oblimin rotation with the exception of item Sup11. In addition, item 12Eff did not sufficiently differentiate between Eff and Sup. For women, the hypothesized pattern matrix was largely replicated, with the exception of Sup11, Sup 8 and Sup 14. Item 15Eff did not differentiate well between Eff and Sup.

concerned a maximum of two items. In addition, the cluster patterns of both genders corroborated the three-dimensional solution found using the entire sample.

Confirmatory Mokken Scale Analysis

Dimensionality Analysis

Table 2 shows the H values resulting from confirmatory MSA for the hypothesized 3-factor model (eight column and further) and the 2-factor model suggested by EFA (fourth until seventh column). The a priori 3-factor model resulted in two medium scales Int and Eff (H=.42 for both scales), and one weak Eff scale (H=.35). The 2-factor model resulted, apart from the above mentioned medium Int scale, in an unscalable General Suppression scale with H=.26. Although $.0 \leq H < .3$ agrees with the MHM, from a practical point of view such low values suggest that person ordering is inaccurate and therefore a set of items is considered unscalable for practical purposes if $H<.3$ (e. g., Sijtsma, and Molenaar 2002, p. 60).

An additional confirmatory analysis was executed to determine whether to regard the Sup and Eff items as forming two separate clusters or one General Suppression cluster as suggested by EFA. For this mean, an AISP was executed including only the 10 Sup and Eff items. The analysis started with lower bound c=.3 and progressively increased with steps of .05 until c=.6. For c=.3, two scales were found: The first scale consisted of 6 items (5 Sup and 1 Eff), the second scale of the remaining 4 Eff items. Subsequent analyses with successively increasing lower bound c showed that the number of items that could be clustered into either the Sup or Eff scale decreased and that Sup items only clustered with other Sup items, and that Eff items only clustered with other Eff items. Hence, combining the confirmatory MSA results strongly suggested that Sup and Eff must be regarded as two separate scales.

Monotonicity Analysis

Monotonicity was investigated for the three 5-item subscales and for the 2-factor variant of one Int scale of five items and one General Suppression scale of 10 items. For both models, we counted for each item the number of significant decreases in its IRFs and investigated *Crit*, an item summary value (Molenaar & Sijtsma, 200, p.74). *Crit*<40 means that sampling error likely caused the violations of monotonicity, 40≤*Crit*≤80 indicates mild violations whereas *Crit*>80 suggests serious violations. For the three-factor model, for minsize=66, a total of 5 significant sample violations occurred in 75 ISRFs (i.e., five ISRFs for each of the five items per scale). However, none of these violations had *Crit* values≥40: the lowest *Crit* value was 6 and was found for item 1Int, the highest was 32 and was found for item 12Eff. Hence, the monotonicity assumption held for all items. For the two-factor model, 3 violations were found from which the lowest *Crit* value was 6 (1Int) and the highest was 24 (9Eff). Again, for all items monotonicity was not rejected.

Confirmatory Factor Analysis

As the results from EFA (in particular MAP and PA) suggested a two-dimensional solution, but the results from exploratory and confirmatory MSA favored a three-dimensional solution, the fit of both models to the data was tested using CFA on the covariance matrices. Both models fitted poorly (χ^2=368,69, df=87, p<.001 for the three-factor model and χ^2=661,48, df=89, p<.001 for the two-factor model) (See Table 3). Subsequently, for both models it was checked if measurement errors of items within the same scale were covarying. If so, these covariances were added but only between items from the same scale and if the modification indices exceeded 25. If the fit remained unsatisfactory, it was further checked if adding direct effects could improve the fit. Also here only effects with a modification index larger than 25 were added.

Incorporating covariances increased the fit of both models. For the three-dimensional model, adding 2 error covariances (one pair concerned Eff items, the other pair Sup items) and one direct effect (of latent variable Eff on 8Sup) improved the fit to χ^2=279,78 (df=84, p<.001). The RMSEA value also improved to .08 (90% CI ranging from .07 to .10), indicating by approximation a reasonable fit. CFI, TLI, and GFI indices are shown in Table 3. The fit of the two-factor model could be improved by adding 2 error covariances (one pair concerned Eff items, the other pair Sup items) and one effect (of latent variable Eff on 8Sup). This improved the fit to χ^2=428,48 (df=86, p<.001) with an RMSEA value of .11 ((90% CI ranging from .10 to .12), indicating a poor fit.

An additional analysis was executed to examine whether the Sup and Eff dimensions could be distinguished in a structural equation model. We fitted a model in which all Sup and Eff items loaded on the same (Gensup) factor to the data and compared that to the model considering Sup and Eff as two separate dimensions. After adding two error covariances (between itempairs 3Eff-6Eff and 11Sup-14Sup) the fit for this Gensup model was χ^2=237,72 (df=33, p<.001; RMSEA=.14, CFI=.76, TLI=.67, and GFI=.87), indicating poor fit. In contrast, the fit of the model that distinguished Sup and Eff, after adding one error covariance (between itempair 11Sup-14Sup), was χ^2= 168,91 (df=33, p<.001; RMSEA=.11, CFI=.84, TLI=.78, and GFI=.91), indicating reasonable fit. Hence, comparing the results of CFA shows that the three-factor model, thus distinguishing Supp and Eff, showed the best fit as indicated by smaller RMSEA, and higher CFI, TLI and GFI. Table 3 shows that the gender differences of the fit indices were minimal, with the exception of GFI of the Gensup model that was .55 for men and .85 for women.

Scale Analysis Results Combining EFA, Exploratory and Confirmatory MSA and CFA

Clustering the 15 items in their originally proposed factors (five items in each factor) yields the following combined results. MSA suggested a medium Int scale (H=.42, see Table 2) consisting of two weak items (H_{10} and H_{13}=.3≤H_j<.4), two medium items (H_1 and H_4=.4≤ H_j<.5), and one strong item (H_7=.52). CFA showed standardized regression weights for Int that ranged from .38 (item 10) to .79 (items 1 and 4). In addition, the medium Sup scale (H=.42) consisted of two weak items (H_{11} and H_5=.3≤H_j<.4) and three medium items (H_2, H_8, and H_{14}=.4≤H_j<.5). CFA showed standardized regression weights ranging from .50 (item 5) to .70 (item 14). Finally, the medium Eff scale (H=.35) consisted of two unscalable item (H_{12} and H_{15}<.3), one medium item (H_9=.37) and two strong items (H_3,and H_6=.4≤H_j<.5). For Eff the regression weights ranged from .39 (item 15) to .79 (item 6). Int correlated .31 with Sup and -.19 with Eff, Sup correlated .14 with Eff. Table 3 (last three columns) shows the reliability results for the total scores of both the 2-factor and the 3-factor model. Consistent with current critical views on the use of Cronbach's alpha for reliability (see for example Sijtsma 2009), for all total scores lambda2 was slightly higher than Cronbach's alpha, and in its turn for all total scores GLB was higher than lambda2.

Discussion

The aim of this study was to examine the dimensionality of the Thought Suppression Inventory, developed by Rassin (2003). Opposed to EFA (PAF, but MAP and PA in particular), it was suggested by exploratory and confirmatory MSA and CFA that the original 3-factor model as proposed by Rassin is to be preferred. These three factors are Intrusion, Suppression and Successful Suppression.

Although the three-dimensional structure is suggested, the analysis showed that several items turned out to have unsatisfactory psychometric properties. First, the pattern matrix following Oblimin rotation showed that item 14Sup had almost similar loadings on Int and Sup and that item 8Sup had almost similar loadings on all three scales. In addition, item 12Eff also loaded on all three factors with minimum differences between factorloadings of .10 and .13, insufficient to clearly discriminate between the factors. Second, exploratory and confirmatory MSA suggested that item 15Eff was unscalable even at the smallest lowerbound *c*. In addition, it suggested that Int consisted of a strong core around items 1, 4 and 7, that Sup consisted of a strong core of items 8, 11, and 14 and, finally, that Eff had a strong core of items 3, 6, and 9. This suggests that the weaker items 2, 5, 10, and 13 should at least be rephrased or even removed. Finally, Items 12 and 15 had H<.3 which indicates the items are unscalable. CFA showed that item 8Sup loaded on both Sup and Eff. Hence, from 15 items, 8 items were marked as unsound and should either be rephrased or removed.

It is important to consider whether low scalability means that the item is a weak indicator of the latent trait one aims to measure or that it is the result of poor wording of the items (Emons et al. 2010). Indeed, several items of the TSI might improve by rewording the items. For example, the wording of item 10Int, *"I regularly 'hear' unexplainable things inside my head, such as my own voice, or the voices of people who are not present"* may be too strong and disturbing (also evidenced by the lowest item mean of all TSI items) and in addition may assess the experience of intrusive perceptions rather than intrusive thoughts. Rewording the item by, for example, using a more daily life example of having intrusive thoughts (such as not being able to put the errands one must run that day out of mind), may benefit its scalability.

MSA provided a detailed analysis of the items' scalability and the dimensional structure of the TSI. By progressively increasing the lower bound *c* for scalability and thus placing stronger demands on the data structure, MSA provided alternating ways of forming scales. Studying this pattern of alternating cluster outcomes provides detailed information on the most appropriate conclusion with respect to scalability and dimensionality (Wismeijer et al. 2008). Various small differences between men and women with respect to how items were clustered were detected. However, in both genders the same conclusion was reached that a three-dimensional structure best fitted the TSI data, and that these three dimensions consisted of the same content. This is an asset that brings additional validity to the main finding of this study that the TSI indeed can be best regarded as a three-dimensional instrument.

The finding that the Suppression and Successful Suppression factors should be distinguished and can best be treated as two separate dimensions corroborates the literature that found that both may have separate dynamics and consequences for the number of subsequent thought intrusions (Blumberg 2000; Höping and de Jong-Meyer 2003; Muris et al. 1996; Rassin 2003). The controversy on whether unsuccessful thought suppression indeed leads to a rebound effect partially focuses on the potential determinants of the rebound effect such as unfocused distraction (Wegner et al. 1991), asymmetry of associate priming in suppression (Najmi and Wegner 2009), metacognitive beliefs about thought processes (Clark and Purdon 1993, 1995) and different types of intrusive thoughts such as neutral, positive or negative thoughts (Salkovkis and Campbell 1994), ego-dystonic thoughts and so forth. However, few studies have explicitly investigated both successful and unsuccessful suppression simultaneously and most research on thought suppression used the WBSI that implicitly assessed unsuccessful thought suppression instead of thought suppression (Rassin 2003). Including a measure of successful suppression in future investigations is desirable to better understand under what circumstances and why successful suppression is indeed possible and when it leads to the rebound effect.

In light of the above discussion it is relevant to note that the Suppression factor may benefit from being renamed. First, it is rather odd that Effective Suppression, which one intuitively would assume is subsumed under the broader label ‘Suppression’, contains entirely different items and hence measures something else than Suppression. Second, scrutiny of the Suppression items suggests that the items do not assess thought suppression per se, but rather one’s intention or efforts to suppress one’s thoughts. This is caused by the words ‘try to’ (e.g. “I try to avoid certain thoughts”) that is present in 4 of the 5 Suppression items. Renaming the Suppression factor into, for example, ‘Thought Suppression Efforts’ or ‘Intention to Suppress Thoughts’ may solve this somewhat confusing situation.

Most noticeably, clinical populations such as OCD patients may benefit from a nuanced understanding of the relation between thought suppression techniques and intrusive thoughts as the current debate and inconsistent findings on what causes the rebound effect may inadvertently downplay the importance of the thought suppression paradox in the maintenance of obsessional thoughts (Purdon and Clark 2000). Using a scale that explicitly distinguishes thought suppression attempts from successful thought suppression in clinical research on risk factors for experiencing obsessive intrusive thoughts may show if individuals who are at higher risk for obsessive intrusive thoughts are indeed less successful at thought suppression and more vulnerable to the rebound effect (Purdon and Clark 2001). To date this is an unresolved matter that lies at the heart of understanding the obsessive nature of OCD patients.

This study has some limitations. First, a sample of community-dwelling senior citizens was used, which limits generalizing to other populations. However, this particular sample was deemed appropriate to assess relatively active thought suppression processes without the disadvantages inherent of investigating a clinical sample. In addition, having replicated Rassin’s (2003) findings that used undergraduate students and less advanced statistical techniques, brings additional validity to the three-dimensional solution of the TSI and suggests the solution may be unsensitive to age differences. A second limitation is that the external validity of the present study is limited because only a self-report assessment was employed.

More replication studies are needed to confirm the dimensionality of the TSI, using samples drawn from multiple populations and clinical populations in particular, as the TSI may be of special importance to clinical psychological research. In addition, research on patient populations that are characterized by experiencing uncontrollable thoughts or inadequate thought control in general should use the TSI to be able to differentiate between thought suppression and successful thought suppression. This may shed light on when thought suppression may lead to the rebound effect, who is most at risk to experience the rebound effect and what role (unsuccessful) thought suppression plays in disorders related to intrusive thoughts and obsessions.

It is concluded that the TSI indeed consists of the three dimensions suggested by Rassin (2003). However, as several items appeared to have some shortcomings, it is advised that future research critically evaluates the weaker items in order to either reword, replace or remove them. In addition, the predictive validity of the three dimensions must be studied prior to considering adopting the TSI in clinical practice.

References

Arbuckle, J. L. (2006). *Amos (Version 7.0) [Computer Program]*. Chicago: SPSS.

Beauducel, A. (2001). Problems with parallel analysis in data sets with oblique simple structure. *Methods of Psychological Research, 6*, 141–157.

Blumberg, S. J. (2000). The White Bear Suppression Inventory: revising its factor structure. *Personality and Individual Differences, 29*, 943–950.

Brewin, C. R., & Beaton, A. (2002). Thought suppression, intelligence, and working memory capacity. *Behavior Research and Therapy, 29*, 943–950.

Byrne, B. M. (2001). *Structural equation modeling with Amos: basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Erlbaum.

Clark, D. A., & Purdon, C. (1993). New perspectives for a cognitive theory of obsessions. *Australian Psychologist, 28*, 161–167.

Clark, D. A., & Purdon, C. (1995). The assessment of unwanted intrusive thoughts: a review and critique of the literature. *Behaviour Research and Therapy, 33*, 967–976.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*, 1–9.

Emons, W. H. M., Sijtsma, K., & Pedersen, S. S. (2010). Dimensionality of the Hospital Anxiety and Depression Scale (HADS) in Cardiac Patients: Comparison of Mokken Scale Analysis and Factor Analysis. *Assessment*. Retrieved from http://asm.sagepub.com/content/early/2010/10/14/1073191110384951.full.pdf

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299.

Hemker, B. T., Sijtsma, K., & Molenaar, I. W. (1995). Selection of unidimensional scales from a multidimensional itembank in the polytomous Mokken IRT model. *Applied Psychological Measurement, 19*, 337–352.

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.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

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.

Kessler, R. C., Berglund, P., Demler, O., Jin, 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*, 593–602.

Lane, J. D., & Wegner, D. M. (1995). The cognitive consequences of secrecy. *Journal of Personality and Social Psychology, 69*, 237–253.

Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement, 6*, 417–430.

Molenaar, I. W., & Sijtsma, K. (2000). *User´s manual MSP5 for windows*. Groningen: iecProGAMMA.

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*, 501–513.

Najmi, S., & Wegner, D. M. (2009). Hidden complications of thought suppression. *International Journal of Cognitive Therapy, 2*, 210–223.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*, 396–402.

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

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*, 1163–1182.

Rassin, E. (2003). The White Bear Suppression Inventory (WBSI) focuses on failing suppression attempts. *European Journal of Personality, 17*, 285–298.

Salkovkis, P. M., & Campbell, P. (1994). Thought suppression induces intrusion in naturally occurring negative intrusive thoughts. *Behaviour Research and Therapy, 32*, 1–8.

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.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120.

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

Thornton, T. L., & Gilden, D. L. (2005). Provenance of correlations in psychological data. *Psychonomic Bulleton & Review, 12*, 409–441.

Van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory*. New York: Springer.

Wegner, D. M. (1992). You can´t always think what you want: Problems in the suppression of unwanted thoughts. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 193–225). New York: Academic.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review, 101*, 34–52.

Wegner, D. M., Schneider, D. J., Carter, S. R., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology, 53*, 5–13.

Wegner, D. M., Schneider, D. J., Knutson, B., & McMahon, S. R. (1991). Polluting the stream of consciousness: the effect of thought suppression on the mind's environment. *Cognitive Therapy and Research, 15*, 141–152.

Wegner, D. M., & Zanakos, S. (1994). Chronic thought suppression. *Journal of Personality, 62*, 615–640.

Wenzlaff, E. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology, 51*, 59–91.

Wismeijer, A. A. J., Sijtsma, K., van Assen, M. A. L. M., & Vingerhoets, A. J. J. M. (2008). A comparative study of the dimensionality of the self-concealment scale using principal components analysis and Mokken scale analysis. *Journal of Personality Assessment, 90*, 323–334.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432–442.