

H. Merckelbach, P. Muris, I. Wessel & P.J. van Koppen  
(1998) The Gudjonssons Suggestibility Scale (GSS):  
Further data on its reliability, validity, and metacognition  
correlates. *Social Behavior and Personality*, 26, 203-210

## THE GUDJONSSON SUGGESTIBILITY SCALE (GSS): FURTHER DATA ON ITS RELIABILITY, VALIDITY, AND METACOGNITION CORRELATES

---

HARALD MERCKELBACH, PETER MURIS, INEKE WESSEL,  
AND PETER J. VAN KOPPEN

This article presents two studies in which the psychometric properties and validity of the Gudjonsson Suggestibility Scale (GSS) were further investigated. Results of the first study ( $N = 40$ ) indicate that the GSS has reasonable internal consistency. Additionally, a modest, but significant test-retest stability was found for the GSS. As to the association between suggestibility and self-reported cognitive efficiency (i.e., metacognition measures), scores on the Yield dimension of the GSS were positively and significantly related to scores on the Dissociation Experiences Scale (DES), but not to scores on the Cognitive Failures Questionnaire (CFQ). In a second, experimental study ( $N = 53$ ), evidence was found for the predictive validity of the GSS. In that study, subjects saw a slide series and were then confronted with leading questions about the critical (emotional) slide. In addition, they completed the Yield scale of the GSS. A small but significant correlation was found between subjects' Yield scores on the GSS and their susceptibility to leading questions about the slide series.

The Gudjonsson Suggestibility Scale (GSS) is an instrument that intends to measure individual differences in interrogative suggestibility (Gudjonsson, 1984a; 1992; 1997). More specifically, the GSS taps two different aspects of interrogative suggestibility, namely the tendency to give in to leading questions (Yield) and the tendency to shift responses under conditions of interpersonal pressure (Shift). In short, the GSS consists of a narrative paragraph that is read out to the subject, who then reports all he or she recalls about the story. Following this, the subject is asked a number of questions about the story, some of which are (mis)leading. Next, the subject is told in an authoritative manner that he or she

---

Harald Merckelbach, Peter Muris, Ineke Wessel, and Peter J. van Koppen, Department of Psychology, Maastricht University, The Netherlands.

Peter J. van Koppen, Netherlands Institute for the Study of Criminality and Law Enforcement, Leiden, The Netherlands.

Please address all correspondence and reprint requests to: Prof. dr. Harald Merckelbach, P.O. Box 616, 6200 MD Maastricht, The Netherlands.

has made a number of errors and must answer the questions for a second time. Yield refers to susceptibility to suggestive questioning, while Shift pertains to pressured suggestibility, i.e., the tendency to change answers as a result of social pressure.

Several studies have evaluated the psychometric properties of the GSS. In general, these studies found that the GSS Yield and Shift scales have sufficient internal reliability and test-retest stability (Gudjonsson, 1992), although it is also clear that Shift scores critically depend on whether the pressured instructions are presented in a convincing manner. Furthermore, factor analytic studies have revealed a factor structure that fits the Yield and Shift dimensions (e.g., Gudjonsson, 1984a; 1992). In addition, there are field and case studies that have found evidence to support the validity of the GSS (see, for a review, Gudjonsson, 1992; see also, Gudjonsson, 1995). For example, in a study by Gudjonsson and Singh (1984), GSS scores of delinquent boys were found to be correlated with judgements of their suggestibility made by their teachers. As another example, Gudjonsson (1984b) reported that suspects who retract their confession ("false confessors") have higher GSS scores than suspects who consistently deny any involvement in the crime they are charged with ("deniers").

Secondly, a number of studies have examined psychological and psychopathological correlates of GSS Yield and Shift scores. By and large, this research has revealed a theoretically meaningful pattern of correlations. For example, age and memory are both negatively related to GSS Yield and Shift scores (e.g., Danielsdottir, Sigurgeirsdottir, Einarsdottir & Haraldsson, 1993; Gudjonsson, 1984a), while correlations with anxiety measures are consistently higher for Shift than for Yield suggestibility (Gudjonsson, 1988; but see Gudjonsson, Rutter & Clare, 1995).

The GSS is increasingly used for practical forensic purposes (e.g., Kassin, 1997). Forensic application of the GSS can be expected to increase even more now that trauma therapists (e.g., Brown, 1995) have recommended the GSS as an instrument for identifying patients who are vulnerable to develop pseudomemories.

The current article presents data from two studies in which the GSS was further investigated. Note that most work on the psychometric properties of the GSS has been done by Gudjonsson and co-workers (see, for a summary, Gudjonsson, 1992). The first study sought to replicate their basic findings. Thus, the first study was concerned with the psychometric properties (i.e., internal consistency and test-retest stability) of the GSS. In addition, this study examined the connection between suggestibility as indexed by the GSS and some widely used metacognition self-report scales, namely the Dissociative Experiences Scale (DES) and the Cognitive Failures Questionnaire (CFQ). These scales were included because they provide important information about how a person evaluates his or her own cognitive capabilities. One could argue that uncertainty about one's own

cognitive efficiency (e.g., “memory distrust”; see e.g., Gudjonsson, 1992) contributes to suggestibility. The second study sought to validate the Yield scale of the GSS against an experimental standard. That is, GSS Yield scores of subjects were related to their tendency to give in to leading questions about a slide series.

## STUDY 1

### METHOD

In the first study, 40 undergraduate students (37 women and 3 men; mean age = 20.6 years,  $SD = 2.2$ , range 18-26) completed the Dutch translation of the GSS. Gudjonsson (1992; 1997) described two versions of the GSS, namely GSS 1 and its parallel form GSS 2, the main difference being the content of the short stories of both versions. The data reported below pertain to the GSS 1 version. As we were primarily interested in suggestibility scores, no immediate or delayed free recall data were obtained in Study 1 and Study 2. Thus, in Study 1, subjects listened to the short story of the GSS, then engaged in a brief filler task which consisted of self-report scales not related to the short story (among which the DES and the CFQ; see below), and finally answered the 20 GSS questions, of which 15 were (mis)leading (Yield) and 5 “true” (i.e., non-misleading). Next, subjects were told that they had made a number of errors and they were asked to complete the questions for a second time (Shift). In passing, it should be noted that for the Shift scale, the original scoring system was used (i.e., only changes of answers to any of the suggestive questions were included in the calculation of Shift scores; Gudjonsson, 1984a; but see Singh & Gudjonsson, 1992). To get an impression of the stability of the GSS, the narrative of the GSS was read out to the students for a second time, four weeks later and they then completed the Yield scale again. On this second occasion, no Shift data were collected, simply because presenting the Shift instructions for a second time would have been unconvincing.

The DES is a 28-item self-report instrument. The items describe disturbances in the normal integration of identity, memory, awareness, and thought. Subjects indicate on a 100-mm visual analogue scale the degree to which the experience described in the item applies to them (0 = “never”; 100 = “always”). VAS scores are averaged over the 28 items to obtain a total DES score. Research has indicated that the DES has good test-retest stability and internal consistency (e.g., Bernstein & Putnam, 1986).

The CFQ comprises 25 items and measures self-reported frequency of everyday failures in perception/attention, memory, and action.

Subjects are asked to indicate on a 5-point scale (0 = “never”; 4 = “very often”) how often they have experienced each cognitive failure in the past months. Accordingly, total CFQ scores range between 0 and 100, with higher scores indi-

cating higher frequency of self-reported cognitive failures. The CFQ possesses reasonable internal consistency and test-retest stability (e.g., Broadbent, Cooper, Fitzgerald & Parkes, 1982; Merckelbach, Muris, Nijman & de Jong, 1996).

## RESULTS

Mean Yield, Shift and total suggestibility (Yield + Shift) scores were 5.65 ( $SD = 2.38$ ), 2.63 ( $SD = 2.06$ ), and 8.28 ( $SD = 2.34$ ), respectively. Reliability of the GSS in terms of internal consistency was sufficient, Cronbach's alphas being 0.79, 0.75, and 0.82 for Yield, Shift, and total suggestibility, respectively. As to the stability of the Yield scale, mean Yield scores for occasion 1 and 2 were 5.65 ( $SD = 2.38$ ) and 4.95 ( $SD = 2.91$ ), respectively. The test-retest correlation was 0.55 ( $P < 0.01$ ).

Mean DES and CFQ scores in this sample were 17.9 ( $SD = 11.5$ ) and 40.8 ( $SD = 9.1$ ), respectively. Table 1 shows Pearson product-moment correlations between suggestibility indices and metacognition scales (i.e., DES and CFQ). As can be seen, the only robust correlation was that between Yield and DES scores ( $r = 0.40$ ,  $P < 0.01$ ): the more subjects reported dissociative experiences, the stronger their tendency to give in to leading questions. The correlation between Yield and Shift did not attain significance ( $r = 0.28$ ,  $P = 0.09$ ), whereas the association between DES and CFQ was highly significant ( $r = 0.55$ ,  $P < 0.01$ ).

TABLE 1  
PEARSON PRODUCT-MOMENT CORRELATIONS BETWEEN GSS YIELD, GSS SHIFT, AND TOTAL SUGGESTIBILITY SCORES (YIELD + SHIFT), ON THE ONE HAND, AND DISSOCIATIVE EXPERIENCES SCALE (DES) AND COGNITIVE FAILURES QUESTIONNAIRE (CFQ) SCORES, ON THE OTHER HAND, IN AN UNDERGRADUATE SAMPLE ( $N = 40$ ).

	DES	CFQ
GSS Yield	0.40**	0.30
GSS Shift	0.09	-0.05
Total suggestibility	0.32*	0.17

Notes. Only relevant correlations are shown. \*\*  $P < 0.01$ ; \*  $P < 0.05$

## STUDY 2

### METHOD AND RESULTS

To examine the predictive validity of GSS Yield-scores, a laboratory study was conducted. A sample of 53 undergraduate students (41 women and 12 men; mean age = 20.0 years; range 18-27) completed the GSS and then were exposed to a slide series of which the middle, critical slide depicted an accident victim (see, for an extensive description, Wessel, 1997). More specifically, the critical slide showed a nine-year old girl who was lying on a pedestrian crossing, bleeding from a head injury. After a short time interval, subjects were given a number of memory questions about this critical slide, among which there were 10 leading

questions (see, for a similar approach, e.g., Barnier & McConkey, 1992). For example, subjects were asked whether the girl was wearing a green or a blue shawl, when in fact the girl was wearing no shawl at all. The average GSS Yield score for this sample was 4.0 ( $SD = 1.2$ ), while the mean score on the suggestive questions about the critical slide was 1.4 ( $SD = 1.4$ ). The Pearson correlation between subjects' GSS Yield scores and the extent to which they gave in to leading questions was small, but significant:  $r(53) = 0.22, P < 0.05$ .

## DISCUSSION

The current study further investigated the psychometric properties and validity of the GSS. The main results can be catalogued as follows. To begin with, the internal consistency of the GSS subscales proved to be sufficient. Furthermore, reliability of the Yield scale in terms of test-retest stability was modest, but significant. Additionally, a significant correlation between Yield suggestibility and dissociative experiences (DES) was documented. Finally, we found some indications for the predictive validity of the GSS Yield scale.

As to the internal consistency of the GSS, it should be noted that the alpha coefficients that we found come close to the alpha coefficients reported by Gudjonsson (1984a). Yet, the test-retest correlation for the Yield scale of the GSS 1 was remarkably lower than that obtained by Gudjonsson (1992). That is, the present study found a correlation of 0.55 versus 0.81 reported by Gudjonsson (1992). Meanwhile, it should be emphasized that our sample relied on undergraduate students whereas Gudjonsson used a sample from the general population that was probably characterized by a broader range of suggestibility. Thus, it may be well be the case that reliance on a homogeneous sample of undergraduate students who exhibit little variation in suggestibility and who are confronted with the GSS twice within a limited time frame, results in an underestimation of the test-retest stability.

Much the same is true for the findings of our experimental validation study. The correlation between GSS Yield scores and the tendency to give in to leading questions was low but significant. Again, bearing in mind that the sample of this study consisted of a homogeneous group of educated subjects, this result can be taken as support for the validity of the GSS Yield scale. Although Gudjonsson (1984b; 1992) collected field data underlining the validity of the GSS, few studies have attempted to cross-validate the GSS along experimental lines. The present findings suggest that it might be worthwhile to pursue the experimental approach using more heterogeneous samples (see also Register & Kihlstrom, 1988; Blagrove, 1996).

The finding that dissociative experiences (DES), but not self-reported cognitive failures (CFQ) were related to Yield suggestibility deserves some comment.

Apparently, while a high CFQ score indicates a high frequency of self-reported failures in attention, memory, and actions, such a negative evaluation of one's cognitive efficiency is not necessarily accompanied by heightened suggestibility. One reason for this might be that people with high CFQ scores have a more critical attitude towards their own cognitive efficiency. This could imply that they are also more aware of suggestive influences and in that case it would be difficult to detect a correlation between CFQ and suggestibility. The significant correlation between DES and Yield suggestibility is interesting for two reasons. First, this finding nicely fits with the results of recent experimental studies that found a strong relationship between subjects' DES scores and their tendency to develop pseudomemories on the basis of misleading memory cues given by the experimenter (e.g., Hyman & Billings, 1997). Second, patients with recovered memories of trauma are often described as people with a strong tendency to have dissociative experiences (e.g., Frankel, 1996). While these connections between suggestibility, dissociative experiences, and trauma reports seem to underline Brown's (1995) recommendation to use the GSS as a tool for identifying patients who are prone to develop pseudomemories of trauma, the present data are, of course, silent about the issue of whether the GSS is suitable for assessing the credibility of recovered trauma memories. In this context, one could raise a more fundamental point, namely whether interrogative suggestibility bears some resemblance to traumatic memories that surface in a clinical setting. Obviously, this point warrants further research.

It is worth noting that there were some procedural variations in the present studies that deviated from the standard procedure of the GSS as proposed by Gudjonsson (1992). More specifically, the studies described above did not include a free recall of the story paragraph. A free recall task was not included because it does not seem to be essential for measuring Yield and Shift suggestibility. Also, for practical reasons, the re-test part of study 1 as well as study 2 did not collect Shift suggestibility data. Despite these limitations, the current findings add to the evidence that the GSS is a promising instrument for measuring interrogative suggestibility.

## REFERENCES

- Barnier, A.J. & McConkey, K.M. (1992). Reports of real and false memories: The relevance of hypnosis, hypnotizability, and context of memory test. *Journal of Abnormal Psychology*, **101**, 521-527.
- Bernstein, E.M. & Putnam, F.W. (1986). Development, reliability and validity of a dissociation scale. *Journal of Nervous and Mental Disease*, **181**, 246-256.
- Blagrove, M. (1996). Effects of length of sleep deprivation on interrogative suggestibility. *Journal of Experimental Psychology: Applied*, **2**, 48-59.
- Broadbent, D.E., Cooper, P.F., Fitzgerald, P. & Parkes, L. R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology*, **21**, 1-16.

- Brown, D. (1995). Pseudomemories: The standard of science and the standard of care in trauma treatment. *American Journal of Clinical Hypnosis*, **37**, 1-24.
- Danielsdottir, G., Sigurgeirsdottir, S., Einarsdottir, H.R. & Haraldsson, E. (1993). Interrogative suggestibility and its relationship with memory and vocabulary. *Personality and Individual Differences*, **14**, 499-502.
- Frankel, F.H. (1996). Dissociation: The clinical realities. *American Journal of Psychiatry*, **153**, 64-70.
- Gudjonsson, G.H. (1984a). A new scale of interrogative suggestibility. *Personality and Individual Differences*, **5**, 303-314.
- Gudjonsson, G.H. (1984b). Interrogative suggestibility: Comparison between "false confessors" and "deniers" in criminal trials. *Medicine, Science, and the Law*, **24**, 56-60.
- Gudjonsson, G.H. (1988). Interrogative suggestibility: Its relationship with assertiveness, social-evaluative anxiety, state anxiety and method of coping. *British Journal of Clinical Psychology*, **27**, 159-166.
- Gudjonsson, G.H. (1992). *The psychology of interrogations, confessions, and testimony*. Chichester: Wiley.
- Gudjonsson, G.H. (1995). The effects of interrogative pressure on strategic coping. *Psychology, Crime & Law*, **1**, 309-318.
- Gudjonsson, G.H. (1997). *The Gudjonsson suggestibility scales manual*. Hove, East Sussex: Psychology Press.
- Gudjonsson, G.H., Rutter, S.C. & Clare, I.C.H. (1995). The relationship between suggestibility and anxiety among suspects detained at police stations. *Psychological Medicine*, **25**, 875-878.
- Gudjonsson, G.H. & Singh, K.K. (1984). Interrogative suggestibility and delinquent boys: An empirical validation study. *Personality and Individual Differences*, **5**, 425-430.
- Hyman, I.E. & Billings, F.J. (1998). Individual differences and the creation of false childhood memories. *Memory*, in press.
- Kassin, S.M. (1997). *The psychology of confession evidence*, *American Psychologist*, **52**, 321-233.
- Merckelbach, H., Muris, P., Nijman, H. & de Jong, P.J. (1996). Self-reported cognitive failures and neurotic symptomatology. *Personality and Individual Differences*, **20**, 715-724.
- Register, P.A. & Kihlstrom, J.F. (1988). Hypnosis and interrogative suggestibility. *Personality and Individual Differences*, **9**, 549-558.
- Singh, K.K. & Gudjonsson, G.H. (1992). The vulnerability of adolescent boys to interrogative pressure: An experimental study. *Journal of Forensic Psychiatry*, **3**, 167-170.
- Wessel, I. (1997). *Attentional narrowing as a model for partial amnesia*. Doctoral dissertation, Maastricht University. Maastricht: University Press.