The Sussex-Waterloo Scale of Hypnotizability (SWASH): measuring capacity for altering conscious experience

P. Lush[1,2,*], G. Moga[2], N. McLatchie[3] and Z. Dienes[1,2]

[1]Sackler Centre for Consciousness Science, University of Sussex, Falmer BN1 9RH, UK; [2]School of Psychology, Pevensey Building, University of Sussex, Falmer BN1 9RH, UK; and [3]Department of Psychology, Fylde College, Lancaster University, Lancaster LA1 4YF, UK

*Correspondence address. Sackler Centre for Consciousness Science, University of Sussex, Falmer BN1 9RH, UK. E-mail: p.lush@sussex.ac.uk

Abstract

The ability to respond to hypnotic suggestibility (hypnotizability) is a stable trait which can be measured in a standardized procedure consisting of a hypnotic induction and a series of hypnotic suggestions. The SWASH is a 10-item adaptation of an established scale, the Waterloo-Stanford Group C Scale of Hypnotic Suggestibility (WSGC). Development of the SWASH was motivated by three distinct aims: to reduce required screening time, to provide an induction which more accurately reflects current theoretical understanding and to supplement the objective scoring with experiential scoring. Screening time was reduced by shortening the induction, removing two suggestions which may cause distress (dream and age regression) and by modifications which allow administration in lecture theatres, so that more participants can be screened simultaneously. Theoretical issues were addressed by removing references to sleep, absorption and eye fixation and closure. Data from 418 participants at the University of Sussex and the Lancaster University are presented, along with data from 66 participants who completed a retest screening. The subjective and objective scales were highly correlated. The subjective scale showed good reliability and objective scale reliability was comparable to the WSGC. The addition of subjective scale responses to the post-hypnotic suggestion (PHS) item suggested a high probability that responses to PHS are inflated in WSGC screening. The SWASH is an effective measure of hypnotizability, which reflects changes in conscious experience and presents practical and theoretical advantages over existing scales.

Key words: agency; intention; volition; hypnosis; contents of consciousness

Introduction

Hypnosis involves reliable changes in experience which present a unique opportunity for experimentally investigating consciousness. In particular, the experience of involuntariness is central to hypnotic responding (Weitzenhoffer 1980). Hypnosis is an effective tool for experimentally investigating alterations in the sense of agency or the experience of voluntary action (Haggard *et al.* 2004; Lush *et al.* 2017); thus, it creates illusions in agentic consciousness. Additionally, many highly hypnotizable people can experience vivid hallucinations or other altered sensory experiences; thus, hypnosis creates illusions in perceptual consciousness. That is, hypnosis can be used instrumentally for investigating a wide range of conscious experiences (Cardeña 2014; Terhune *et al.* 2017). However, hypnosis is under-used in comparison to established experimental methods of altering, for example, visual consciousness (e.g. continuous flash suppression; Tsuchiya and Koch 2005) or bodily self-consciousness (e.g. the rubber hand illusion; Botvinick and Cohen 1998).

This relative lack of attention may be at least partly attributable to the barrier presented by the time-consuming process of identifying samples of varying hypnotizability for hypnosis research, and therefore there is a need for time-efficient procedures to increase the viability of hypnosis as a research tool.

Hypnotizability can be considered a stable trait (Piccione *et al.* 1989), and empirical investigation of hypnosis commonly employs standardized scales in order to identify potential participants. While scale administration is a straightforward process which requires little training and requires no more than the ability to read a script (Kihlstrom 2008), established inductions are unnecessarily long and fail to reflect contemporary theoretical understanding (Woody and Barnier 2008; Terhune and Cardeña 2016). Here we present a revised version of an established scale, with the aim of creating a practical and theoretically relevant screening procedure which we hope will make hypnosis research more widely accessible to the consciousness research community.

The development of the Sussex-Waterloo Scale of Hypnotizability (SWASH) has been guided by several distinct aims. First, we aimed to construct a time-efficient screening procedure. Second, we wanted to remove some allusions to theories which are no longer considered to be true. Third, we wanted to include an integrated experiential scale, as although hypnosis is characterized by changes in subjective experience, hypnotizability is often measured only according to objective criteria.

The Waterloo-Stanford Group Scale of Hypnotic Susceptibility, Form C (WSGC; Bowers 1993) is a 12-item scale adapted from the earlier Stanford Hypnotic Susceptibility Scale: Form C (Weitzenhoffer and Hilgard 1962), in which participants are screened individually. The WSGC was developed to have items more difficult than the Harvard Group Scale of Hypnotic Susceptibility, Form A (HGSHS:A) (Shor and Orne 1962), which contains a relatively high proportion of easier ideomotor suggestions and therefore can fail to sufficiently identify high hypnotizables (Bowers 1993; Laurence *et al.* 2008).

Woody and Barnier (2008) proposed a four-facet model for hypnotic suggestions in standardized scales, with items categorized as either direct or challenge suggestions which require either a motor or perceptual-cognitive response, and the WSGC contains a representative mix of these suggestion types. However, the WSGC can be time consuming to run, as sessions are administered to group sizes of a dozen or fewer and can take up to 90 min to complete. Commonly, the highest and lowest scoring 10–15% of participants are identified as highly and low hypnotizable. Therefore, a large number of participants must be screened in order to obtain an acceptable sample sizes for studies which require these groups. At Sussex, screening with the WSGC typically involves approximately 900 min of screening to identify just 10–12 highly hypnotizable participants. In constructing the SWASH, our primary aim was to reduce the time necessary to establish a participant pool, while retaining the difficult items used by the WSGC and the SHSS: C, so that people at the high end of hypnotizability can be still identified.

We reduced screening time in two ways. First, session length was reduced by editing the WSGC induction and by reducing the number of suggestions. Second, the WSGC was modified for increased group size. We will first address the steps taken to decrease screening session time and then the modification necessary for large group presentation.

The common theme in hypnotic inductions is that they establish a hypnotic context for the period of time in which suggestions are delivered (Sheehan and Perry 1976; Lynn *et al.* 2017). There is evidence that the increase in response to suggestion attributable to hypnotic induction is small (Braffman and Kirsch 1999; see Connors *et al.* 2012, for an exception), and that any increase in responding over non-hypnotic suggestibility may be attributable to the use of the word 'hypnosis' (Gandhi and Oakley 2005). Thus, a minimal condition for an induction to enhance response may be simply defining the context as one appropriate for hypnotic response. It is unclear exactly what else, if anything, may be needed to constitute a minimal context, but as there is evidence for very brief inductions being less effective (Klinger 1970), we aimed to cut the induction to only around half its original length.

While evidence for effects on responding is mixed (for a review see Terhune and Cardeña 2016), the prevalence of relaxation instructions in hypnotic inductions suggest that a requirement for relaxation might be expected by many and therefore be useful for establishing the hypnotic context (Lynn *et al.* 2017). Therefore, and although relaxation is not a necessary component of hypnotic inductions (Banyai and Hilgard 1976; Cardeña 2005), we retained elements of the WSGC relating to relaxation and counting down in order to ensure the induction was long enough to meet participants' possible expectations.

All direct references to sleep were removed from the induction script, as hypnosis is distinct from sleep (Hull 1933). However, some references to tiredness were retained as part of the relaxation procedure. An analogy between hypnosis and inattention to the environment whilst driving was also removed, as being distracted or absorbed is not the same as hypnotic responding (the correlation of hypnotizability with absorption is about 0.3 and usually is not found if tested out of the hypnotic context; Laurence *et al.* 2008). Finally, eye fixation and closure (the Braid effect; Weitzenhoffer *et al.* 1959) is a feature of most hypnotic inductions, but comparison of inductions with and without eye fixation provides no evidence for an increase in suggestibility related to the effect (Weitzenhoffer and Sakata 1970). As a substantial proportion of the WSGC induction is related to eye closure, the removal of this material considerably shortened the induction. In total, the pre-suggestion induction script (including preliminary instructions) was cut from 1636 words to 873 words. In summary, the SWASH induction script retains some elements of the WSGC script intended to establish rapport and to motivate or reassure participants, suggestions of relaxation and a counting down procedure followed by a counting up de-induction.

Screening time was further reduced by the removal of two perceptual-cognitive WSGC suggestions: dream and age regression. There have been reports of negative responses to the age regression suggestion (Cardeña and Terhune 2009), and the dream suggestion also involves highly personalized experiences that may be negative (Hilgard 1974). The WSGC contains a disproportionately high number of perceptual-cognitive suggestions (Woody and Barnier 2008), so these items could be dropped without leaving this facet underrepresented. Further, the average score on these two items matches that of the WSGC overall (age regression 6.1, dream 4.4 = 5.3, mean WSGC = 5.8; Bowers 1993), so the removal of these items does not change the level of difficulty of the scale as a whole. The SWASH therefore contains ten suggestion items: two motor (hand lowering; moving hands together), two motor challenge (arm rigidity; arm immobilization), three perceptual-cognitive (mosquito hallucination; music hallucination; taste hallucination), two

perceptual-cognitive challenge (amnesia; negative visual hallucination) and a post-hypnotic suggestion (PHS). For the PHS, the suggestion is given before the de-induction begins. Participants are told that they will draw a tree in the corner of their booklet when they are instructed to open their booklets and also that they will forget this suggestion. The PHS item therefore combines a direct motor suggestion to draw with a perceptual-cognitive challenge of amnesia (Woody and Sadler 2008).

With these adjustments, the total time to administer SWASH is around 40 min rather than approximately 90. Because it can be administered to more people simultaneously (it has been tested with up to 50), establishing a participant pool with the WSGC should take almost 10% of the time required by the WSGC.

The negative visual hallucination suggestion was modified for large group presentation. The WSGC negative hallucination suggestion involves placing three coloured balls in the centre of the room. In the SWASH, a picture of three coloured balls are presented on a slide. Some minor modifications to other suggestions were made to improve universality (in particular for non-native English speakers). Baseball and billiard ball were replaced by bowling ball for the arm heaviness suggestion and Jingle Bells was replaced by Happy Birthday for the music hallucination, as it is perhaps the most widely recognized song worldwide (Brauneis 2008).

The WSGC relies on behavioural scoring to generate a pass or fail score for each item. However, it is not the visible physical responses to suggestion but the experience which accompanies the suggested behaviour which is of particular interest in hypnotic responding. A subjective scale has been developed for the WSGC (Kirsch *et al.* 1998), but unfortunately has received little attention from researchers. The SWASH subjective scale is similar to this existing scale, with responses to questions regarding the veridicality of the suggested experience recorded on a scale between 0 and 5.

The SWASH differs from the earlier scale in requiring two subjective responses to the PHS item. The PHS in the Harvard Group Scale of Hypnotic Susceptibility, Form A (HGSHS: A; Shor and Orne 1962) requires participants to touch their ankles following a signal but not to remember doing so. Sadler and Woody (2004) have criticized objective scoring of this item on the basis that it cannot exclude 'spurious passes' that include actions experienced as voluntary and with full memory of the suggestion. Similarly, the WSGC PHS suggestion is passed if participants draw a tree in their booklet. However, the suggestion states that participants will draw a tree but forget that they were told to do so. For the SWASH, therefore, a PHS response only receives a subjective score if participants report both an urge to draw a tree and some amnesia about hearing the suggestion. We anticipate that this approach is likely to identify a substantial number of spurious passes for this item.

Participants often report spontaneous experiences which occur following a hypnotic induction, and such effects have previously been measured by subjective ratings of hypnotic 'depth' (e.g. Tart 1970). As such depth ratings can correlate with response to suggestion, it has been suggested that induction depth could act as a proxy measure of hypnotizability (Wagstaff *et al.* 2008). We included such a rating of depth in order to investigate this possibility and as a check that the edited induction produced hypnotic depth experiences.

The purpose of this study was to produce a more efficient version of the WSGC, measuring hypnotizability across the range of ability and potentially opening hypnosis research up to a greater number of researchers who might otherwise be put off by the impracticalities of screening.

Materials and Methods

Participants

Four-hundred and twenty-nine participants were recruited to undergo a hypnotic screening procedure at the University of Sussex or at Lancaster University. Eleven participants were excluded for incomplete data, so data from 418 participants (331 female, 87 male) were analysed, of which 331 participated at the University of Sussex. The mean age of participants was 19.9 years [standard deviation (SD) = 4.0]. Participants at the University of Sussex were invited to return for a retest screening approximately 2 months after their initial screening and retest data was recorded for 66 participants. Psychology student participants received course credits, and no other compensation was offered.

Materials

The materials required for screening (induction script, response booklet, slide and scoring procedure) are available for download at https://osf.io/wujk8/. An induction and suggestion script was adapted from the WSGC (Bowers 1998). Participants recorded their responses in a booklet adapted from the WSGC. This contained subjective responses on a scale from 0 to 5. There were two versions of the booklet used. The second booklet differed from the first only in the addition of anchoring labels at each end of the scale used to record each subjective response. Anchoring labels were added to the second version of the booklet to reduce the possibility of erroneous responses to the subjective scale items. These restate the instructions provided in the text. For example, for Item 2, 'Moving hands together', participants are provided with the following instruction:

> 'On a scale from 0 to 5, how strongly did you feel a force between your hands, where 0 means you felt no force at all and 5 means you felt a force so strong it was as if your hands were real magnets'?

The label 'No force' is provided alongside 0 on the scale and the label 'Strong force' alongside 5 on the scale.

Approximately half of the participants completed each booklet, with 206 participants completing booklet 1 and 212 completing booklet 2. For the repeat screening, 59 participants completed booklet 2 on both occasions and 6 completed booklet 1. The second version of the booklet (containing anchoring labels) is recommended for future screenings and is available at https://osf.io/wujk8/.

A slide containing a picture of three coloured balls (green, blue and red) presented in a triangular formation on a black background was projected onto a screen at the front of the lecture theatre during the screening.

Procedure

Participants were screened in a lecture theatre in groups of up to 50 and were instructed to leave a seat free between them and the next participant in order to allow freedom for arms to move (e.g. during the magnetic hands suggestion). A slide instructing participants to turn off their mobile phones was displayed, along with information about the length of the procedure. Before the session began, participants were instructed to fill in the front of the booklet with their personal information and

then to sit back in their chair. The experimenter then thanked participants for their attendance and introduced his or her self and informed participants how long the procedure would take before reading from the script. The script contained a brief introductory passage and an induction.

Analyses

Mean objective and subjective scores were calculated. The scores for each version of the booklet were compared for objective and subjective scales. Data from the two booklets were then combined for subsequent analyses.

Objective scores were scored according to the WSGC booklet (Bowers 1998). Each item has a dichotomous response which is recorded as a pass or fail (scored as 1 or 0). For example, for Item 2, moving hands together, participants report whether or not their hands were less than 6 inch apart after 10 s. Subjective scores were taken on a 0–5 scale for each item. For example, for Item 2, moving hands together the following instruction was given for subjective response: 'On a scale from 0 to 5, how strongly did you feel a force between your hands, where 0 means you felt no force at all and 5 means you felt a force so strong it was as if your hands were real magnets'?

For two suggestions there were two subjective responses requested. For 'taste' these were about the experience of 'sweet' and 'sour' suggestions and for the PHS item there were questions relating to experienced urge and to amnesia.

Objective scale scores between 0 and 10 for each participant were calculated by summing the successful objective responses for that participant. Subjective scale scores between 0 and 5 were calculated from the average of subjective scale responses. The final subjective response score for taste is the mean of the sweet and sour responses. For the PHS, the geometric mean of the urge and amnesia responses for the item was calculated, so that a subjective response for this item would be zero if either of the components of the suggestion did not generate a subjective response.

An additional measure of induction depth was taken: 'On a scale from 0 to 5, to what degree did you enter a hypnotic state, where 0 means your general state of consciousness was just the same as normal, 1 means you were slightly hypnotized and 5 means you entered very deep hypnosis'?

Scale validity was investigated by correlation analysis of subjective and objective scales, point biserial correlations between objective and subjective responses for each item and by comparison with data for the SWASH item responses from a 2014 WSGC screening of 202 participants. In all but four cases, *n* was greater than 100. However, there were 66 participants in the return screening sample and 3 point-biserial correlations in which the sample which passed the suggestion by objective criteria was below 100: music ($n = 21$), negative hallucination ($n = 41$) and amnesia ($n = 61$). The ratio of variances in point-biserial correlations was more than 3 for 2 (out of ten) suggestions: music hallucination and negative visual hallucination. For comparison, non-parametric correlations (Spearman's rank coefficient) are reported for point-biserial and test–retest correlations in the Supplementary material.

Reliability of objective and subjective scales was checked with coefficient omega, an alternative to Cronbach's alpha which overcomes some of alpha's known deficiencies (Dunn *et al.* 2014). Reliability was further examined by calculating the omega coefficient when each SWASH suggestion was dropped on each scale, the corrected same scale item-total correlations and test–retest correlations.

Table 1. Mean subjective score and point biserial correlations between behavioural and experiential scoring of suggestions

Suggestion	M	SD	r_{pb}
1. Hand lowering	3.4	1.3	0.46 [0.38, 0.53]
2. Moving hands together	2.9	1.4	0.33 [0.24, 0.41]
3. Mosquito hallucination	1.0	1.4	0.65 [0.59, 0.70]
4. Taste hallucination	1.4	1.3	0.65 [0.59, 0.70]
5. Arm rigidity	2.7	1.5	0.57 [0.50, 0.63]
6. Arm immobilization	2.3	1.5	0.44 [0.36, 0.51]
7. Music hallucination	0.23	0.7	0.56 [0.49, 0.62]
8. Negative visual hallucination	0.43	1.1	0.49 [0.41, 0.56]
9. Amnesia	1.4	1.3	0.34 [0.25, 0.42]
10. Post-hypnotic suggestion	0.93	1.4	0.14 [0.04, 0.23]

To investigate how well the induction depth rating reflects the subjective and objective scales, correlations between strength of induction and each scale/item were run.

Finally, we compared the classification of participants into high, medium and low hypnotizable groups across the WSGC (with the dream suggestion and age regression suggestion responses removed) and the objective SWASH responses. We identified a cut-off for low and high hypnotizables that was close to 15% of the sample (a percentage commonly used to define high and low hypnotizables; Barnier and McConkey 2004) for the lowest and high scorers and then applied these cut-offs to the WSGC data.

95% confidence intervals are reported throughout, which can be interpreted as 95% credibility intervals with uniform priors.

Results

Scores

Objective scores on booklet 1 were very similar to scores on booklet 2, with a mean difference in score of 3.70 vs 3.61 = 0.09 [standard error (SE) = 0.18], 95% CI [−0.26, 0.44]. Subjective scores across the two booklets were also similar, with a mean difference in score of 3.44 vs 3.28 = 0.17 (SE = 0.16), 95% CI [−0.15, 0.48]. There was a difference in correlations between objective and subjective scores across the two booklets of just 0.71 vs 0.68 = 0.03, 95% CI [−0.07, 0.13]. Therefore, results from the two booklets were combined. Mean score out of ten on the objective scale was 3.7 (SD = 1.8) and mean subjective score out of five was 1.7 (SD = 0.8).

Validity

There was a high correlation between objective and subjective scales, $r(418) = 0.70$, 95% CI [0.65, 0.75], providing support for the validity of the subjective scale. Table 1 shows mean subjective score and point biserial correlations between objective and subjective responses for each item. Objective and subjective responses were all at least moderately correlated, (with a mean coefficient of 0.46) except for the PHS to draw a tree. The subjective response for this item was calculated as the geometric mean of a participant's responses to two questions: the first about their urge to draw a tree and the second about amnesia for the suggestion. While urge to draw correlated reasonably with objective response, $r(418) = 0.54$, 95% CI [0.47, 0.60], the plausible range of for correlations between objective response and amnesia for the suggestion was small, $r(418) = -0.09$, 95% CI

Table 2. Comparison of Sussex 2014 WSGC and SWASH percentage of participants passing each suggestion on the objective criterion

Suggestion	SWASH	WSGC
1. Hand lowering	71.8	77.2
2. Moving hands together	76.8	76.7
3. Mosquito hallucination	26.1	27.7
4. Taste hallucination	30.9	29.7
5. Arm rigidity	54.9	68.3
6. Arm immobilization	36.4	46.5
7. Music hallucination	5.0	5.9
8. Negative visual hallucination	9.3	21.8
9. Amnesia	14.6	5.9
10. Post-hypnotic suggestion	39.5	24.3

Table 3. OMEGA (If item dropped) (95% CI in brackets)

	Objective omega	Subjective omega
1. Hand lowering	0.52 [0.44, 0.60]	0.81 [0.78, 0.84]
2. Moving hands together	0.52 [0.44, 0.53]	0.82 [0.79, 0.85]
3. Mosquito hallucination	0.49 [0.41, 0.57]	0.82 [0.79, 0.84]
4. Taste hallucination	0.46 [0.37, 0.54]	0.81 [0.77, 0.83]
5. Arm rigidity	0.44 [0.33, 0.53]	0.80 [0.77, 0.83]
6. Arm immobilization	0.50 [0.41, 0.57]	0.80 [0.77, 0.83]
7. Music hallucination	0.52 [0.44, 0.59]	0.82 [0.80, 0.85]
8. Negative visual hallucination	0.50 [0.42, 0.57]	0.83 [0.80, 0.85]
9. Amnesia	0.51 [0.48, 0.61]	0.81 [0.78, 0.84]
10. Post-hypnotic suggestion	0.55 [0.48, 0.61]	0.82 [0.79, 0.84]

Table 4. Corrected same scale item-total correlations for objective and subjective scores (95% CI in brackets)

	Objective r	Subjective r
1. Hand lowering	0.19 [0.10, 0.28]	0.53 [0.46, 0.60]
2. Moving hands together	0.20 [0.11, 0.29]	0.48 [0.40, 0.55]
3. Mosquito hallucination	0.24 [0.15, 0.33]	0.50 [0.42, 0.57]
4. Taste hallucination	0.32 [0.23, 0.40]	0.59 [0.52, 0.65]
5. Arm rigidity	0.36 [0.27, 0.44]	0.61 [0.55, 0.67]
6. Arm immobilization	0.23 [0.14, 0.32]	0.62 [0.56, 0.68]
7. Music hallucination	0.12 [0.02, 0.21]	0.34 [0.25, 0.42]
8. Negative visual hallucination	0.25 [0.16, 0.34]	0.35 [0.26, 0.43]
9. Amnesia	0.18 [0.09, 0.27]	0.55 [0.48, 0.61]
10. Post-hypnotic suggestion	0.11 [0.01, 0.20]	0.57 [0.50, 0.63]

Table 5. Correlations between induction depth score and individual items on each scale

	Objective	Subjective
1. Hand lowering	0.24 [0.15, 0.32]	0.47 [0.39, 0.54]
2. Moving hands together	0.12 [0.02, 0.21]	0.43 [0.35, 0.51]
3. Mosquito hallucination	0.16 [0.07, 0.25]	0.34 [0.25, 0.42]
4. Taste hallucination	0.34 [0.25, 0.42]	0.43 [0.35, 0.51]
5. Arm rigidity	0.29 [0.20, 0.38]	0.47 [0.39, 0.54]
6. Arm immobilization	0.17 [0.08, 0.26]	0.42 [0.34, 0.50]
7. Music hallucination	0.06 [−0.04, 0.16]	0.21 [0.12, 0.30]
8. Negative visual hallucination	0.19 [0.10, 0.28]	0.25 [0.16, 0.34]
9. Amnesia	0.22 [0.13, 0.31]	0.45 [0.37, 0.52]
10. Post-hypnotic suggestion	0.07 [−0.03, 0.16]	0.32 [0.23, 0.40]

[−0.18, 0.01]. The plausible range for the correlation between urge and amnesia questions was also small, $r(418) = 0.01$, 95% CI [−0.09, 0.11].

Table 2 shows percentage of objective suggestions passed on the 2014 WSGC run at Sussex and the SWASH from 2014 to 2016. Scores are comparable between SWASH and Sussex except for negative hallucination, PHS and amnesia.

Reliability

Omega for the objective scale was 0.53, 95% CI [0.44, 0.60], suggesting that internal consistency was not high for this scale. For the subjective scale, omega was considerably higher, 0.83 (SE = 0.013) 95% CI [0.80, 0.85], indicating good internal consistency for this scale (for comparison, Cronbach's alpha point estimates were 0.52 for objective and 0.82 for subjective scales). 2014 WSGC data at Sussex also had low reliability, with omega 0.56, 95% CI [0.46, 0.66] for all 12 WSGC suggestions and omega = 0.47, 95% CI [0.31, 0.57] for the WSGC with just the ten items shared with the SWASH included.

Table 3 shows omega coefficient if the item is dropped for each SWASH suggestion on each scale. Point estimates of the coefficient were lower in all cases on the subjective scale. However, omega was slightly higher for the objective scale (but not for the subjective scale) PHS item.

Table 4 shows item-total correlations. Each item was correlated with the corrected total scale score, in which for each correlation the total is recalculated without that item. However, while item-total correlations for the subjective scale were substantial (mean $r = 0.51$), the objective scale item-total correlations were on average small (mean $r = 0.22$).

Test/retest reliability

Objective score on retest ($M = 4.0$, SD = 2.0) was not strongly correlated with the original objective score of those taking part in the retest ($M = 3.7$, SD = 2.0), $r(66) = 0.56$, 95% CI [0.37, 0.71]. However, there was a strong correlation between mean retest subjective score, $M = 3.1$ (SD = 1.9) and the original subjective score of those taking part in the retest ($M = 3.4$, SD = 1.6), $r(66) = 0.77$, 95% CI [0.65, 0.85].

Objective return score correlated highly with subjective return score, $r(66) = 0.81$, [0.71, 0.88].

Induction correlations

Mean induction rating was 2.3 (SD = 1.1). The induction rating correlated well with the objective score, $r(418) = 0.44$, 95% CI [0.36, 0.51] and with the subjective score $r(418) = 0.62$, 95% CI [0.56, 0.68].

Table 5 shows correlations between induction depth score and individual items on each scale. All subjective items correlated moderately with the induction depth. For the objective scale, plausible ranges of correlations were small to moderate.

Concordance between classification into high and low hypnotizable groups between SWASH and WSGC scales

Thirteen percentage of SWASH participants scored 1 or below and 15.1% scored 5 or above. Applying these cut-offs to the WSGC data revealed that 5% of WSGC participants scored 1 or below and 13.9% scored 5 and above. The highest score on the WSGC data was 8, while 1% of SWASH participants scored 9 or 10.

Discussion

We tested participants on the SWASH, a modified version of the Waterloo-Stanford Group Scale of Hypnotic Susceptibility. The subjective scale of the SWASH showed good reliability and objective scale reliability was comparable to that of the WSGC. Although the study design does not allow us to disentangle any particular element of the induction, we report no evidence for a substantial decline in scores over the WSGC. We suggest, therefore, that a simple relaxation procedure with counting down and repeated use of the word hypnosis is sufficient to generate response to hypnotic suggestions comparable to the WSGC.

We found good reliability for the subjective scale, but not for the objective scale. It is not surprising that a scale based on ratings outperforms one based on dichotomous items. We therefore suggest recruiting for experiments using either the subjective score or a combined objective and subjective score. For example, one could take the simple mean of the two scores (with the subjective score multiplied by two to be on a 0–10 scale), as we report in Lush *et al.* (2016).

The strength of the correlation between objective and subjective SWASH scales suggests that the subjective scale is a valid measure of hypnotizability. However, the PHS item showed only a weak correlation between objective and subjective responses. This issue arose as a result of breaking the subjective response to this item down into the two components of the suggestion: urge to perform the action and amnesia for the suggestion. While there was a relationship between successfully responding to the suggestion on the objective criterion (by drawing a tree) and reporting an urge to respond, relatively few participants who drew a tree reported amnesia for the suggestion. This suggests that PHSs scored solely on objective criteria produce an unacceptably high level of false positives. In our sample, more than half of the objective passes to the PHS suggestion did not pass the subjective criteria. If only those reporting a high level of amnesia (4–5 on the subjective scale) are included, the pass rate is less than 5%, (or 2.1% for full amnesia). It is therefore likely that the rate of successful post-hypnotic responding has been over-estimated in previous WSGC samples (and indeed screens from other scales). As researchers routinely recruit for hypnosis experiments based on overall scores on a scale, this systematic confound is likely to have resulted in a substantial number of falsely inflated hypnotizability scores. As Sadler and Woody (2004, p. 151) have argued, the continued use of unmodified PHS suggestions 'represents a triumph of tradition over science'.

Objective scale SWASH pass rates were numerically low compared to the WSGC for negative visual hallucination and for motor challenge suggestions, and numerically high for PHS and amnesia; overall similar means were reported for both scales, the population difference likely being no larger than 0.46. It is possible that this is a result of different response rates across the cohorts. However, it is also possible that some or all of these numerical differences are attributable to methodological changes, e.g. increased group size or modifications to the induction. In particular, the reduction in successful response to the negative hallucination item may result from the changes necessary to present the visual stimulus in a group setting. However, further studies would be required to establish whether the observed response rates to particular items represent real differences between the scales, if so, why they occur, and whether or not they have any implications for participant recruitment.

Woody and Barnier (2008) point out that while factor analyses over many suggestions indicates a single underlying factor of hypnotizability, there remain sub-factors consistent with specific capacities (namely for response to motor and perceptual-cognitive suggestions in direct or challenge forms.) Thus, researchers may wish to select not only or even on overall score, but on specific suggestions most relevant to the suggestions that will be used in specific further studies.

According to Weitzenhoffer (1980), the experience of involuntariness is what distinguishes a hypnotic response from a voluntary action. Here, we employed a subjective scale of veridicality as this may be an indirect index of the experience of involuntariness (people experience e.g. hallucinations as real because they don't experience their intention in generating the experience; Dienes, 2012). However, given the centrality of the experience of involuntariness to hypnotic responding, future scales might benefit from the inclusion of a scale which directly taps this experience. Using scales developed for the SHSS: C, Bowers (1981) reported that 20% of responses considered successful by behavioural criteria were not accompanied by the classical suggestion effect and Bowers *et al.* (1988) reported that 20% of failed responses were accompanied by reports of the experience of involuntariness. However, directly asking questions about involuntariness may not be straightforward as for certain suggestions they can be confusing for participants (Kirsch and Braffman 2001) and reported changes in sense of agency may reflect different underlying factors (Polito *et al.* 2013).

In summary, the SWASH is an effective instrument for measuring hypnotizability. The scale offers considerable practical and theoretical advantages over existing scales which tap a similarly wide range of hypnotic experience. The procedure can be administered to large groups of participants in a lecture hall setting and completed well within the time period of a typical lecture. It therefore can be employed to rapidly establish a participant pool for hypnosis-related research. We hope that it makes hypnosis studies practical for researchers interested in investigating reliable experimental manipulation of conscious experience.

Data available at https://osf.io/wujk8/.

Supplementary data

Supplementary data is available at *NCONSC Journal* online.

Funding

References

Barnier AJ, McConkey KM. Defining and identifying the highly hypnotizable person. In: Heap MJ, Brown RJ, Oakley DA (eds), *The Highly Hypnotizable Person: Theoretical, Experimental and Clinical Issues*. London: Routledge, 2004, 30–60.

Banyai EI, Hilgard ER. A comparison of active-alert hypnotic induction with traditional relaxation induction. *J Abnorm Psychol* 1976;**85**:218–24.

Botvinick M, Cohen J. Rubber hands 'feel' touch that eyes see. *Nature* 1998;**391**:756.

Bowers K. Do the Stanford scales tap the "classic suggestion effect"? *Int J Clin Exp Hypn* 1981;**29**:42–53.

Bowers KS. The Waterloo-Stanford Group C (WSGC) scale of hypnotic susceptibility: normative and comparative data. *Int J Clin Exp Hypn* 1993;**41**:35–46.

Bowers KS. Waterloo-Stanford group scale of hypnotic susceptibility, form C: manual and response booklet. *Int J Clin Exp Hypn* 1998;**46**:250–68.

Bowers P, Laurence JR, Hart D. The experience of hypnotic suggestions. *Int J Clin Exp Hypn* 1988;**36**:336–49.

Braffman W, Kirsch I. Imaginative suggestibility and hypnotizability: an empirical analysis. *J Pers Soc Psychol* 1999;**77**:578–87.

Brauneis R. Copyright and the world's most popular song. *J Copyr Soc USA* 2008;**56**:335.

Cardeña E. The phenomenology of deep hypnosis: quiescent and physically active. *Int J Clin Exp Hypn* 2005;**53**:37–59.

Cardeña E. Hypnos and psyche: how hypnosis has contributed to the study of consciousness. *Psychol Conscious* 2014;**1**:123–38.

Cardeña E, Terhune DB. A note of caution on the Waterloo-Stanford Group Scale of hypnotic susceptibility: a brief communication. *Int J Clin Exp Hypn* 2009;**57**:222–6.

Connors MH, Barnier AJ, Coltheart M, *et al*. Mirrored-self misidentification in the hypnosis laboratory: recreating the delusion from its component factors. *Cogn Neuropsychiatry* 2012;**17**:151–76.

Dienes Z. Is hypnotic responding the strategic relinquishment of metacognition? In: Beran M, Brandl J, Perner J, Proust J (eds), *The Foundations of Metacognition*. Oxford: Oxford University Press, 2012, 267–77.

Dienes Z, Perner J. Executive control without conscious awareness: the cold control theory of hypnosis. In: Jamieson G (ed.), *Hypnosis and Conscious states: The Cognitive Neuroscience Perspective*. Oxford: Oxford University Press, 2007, 293–314.

Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol* 2014;**105**:399–412.

Gandhi B, Oakley DA. Does 'hypnosis' by any other name smell as sweet? The efficacy of 'hypnotic' inductions depends on the label 'hypnosis'. *Conscious Cogn* 2005;**14**:304–15.

Haggard P, Cartledge P, Dafydd M, *et al*. Anomalous control: when 'free-will' is not conscious. *Conscious Cogn* 2004;**13**:646–54.

Hilgard JR. Sequelae to hypnosis. *Int J Clin Exp Hypn* 1974;**22**:281–98.

Hull CL. *Hypnosis and Suggestibility*. New York: Appleton Century, 1933.

Kihlstrom JF. The domain of hypnosis, revisited. In: Nash M, Barnier A (eds), *The Oxford Handbook of Hypnosis: Theory, Research, and Practice*. Oxford: Oxford University Press, 2008, 21–52.

Kirsch I, Braffman W. Imaginative suggestibility and hypnotizability. *Curr Direction Psychol Sci* 2001;**10**:57–61.

Kirsch I, Milling LS, Burgess C. Experiential scoring for the Waterloo-Stanford Group C scale. *Int J Clin Exp Hypn* 1998;**46**:269–79.

Klinger BI. Effect of peer model responsiveness and length of induction procedure on hypnotic responsiveness. *J Abnorm Psychol* 1970;**75**:15.

Laurence J-R, Beaulieu-Prevost D, Du Chene T. Measuring and understanding individual differences in hypnotisability. In: Nash M and Barnier A (eds.), *The Oxford Handbook of Hypnosis: Theory, Research, and Practice*. New York: Oxford University Press, 2008, 225–54.

Lush P, Naish P, Dienes Z. Metacognition of intentions in mindfulness and hypnosis. *Neurosci Conscious* 2016;**2016**:niw007.

Lush P, Caspar EA, Cleeremans A, *et al*. The power of suggestion: posthypnotically induced changes in the temporal binding of intentional action outcomes. *Psychol Sci* 2017;**28**:661–9.

Lynn SJ, Maxwell R, Green JP. The hypnotic induction in the broad scheme of hypnosis: a sociocognitive perspective. *Am J Clin Hypn* 2017;**59**:363–84.

Piccione C, Hilgard ER, Zimbardo PG. On the degree of stability of measured hypnotizability over a 25-year period. *J Pers Soc Psychol* 1989;**56**:289–95.

Polito V, Barnier AJ, Woody EZ. Developing the sense of agency rating scale (SOARS): an empirical measure of agency disruption in hypnosis. *Conscious Cogn* 2013;**22**:684–96.

Sadler P, Woody EZ. Four decades of group hypnosis scales: what does item-response theory tell us about what we've been measuring? *Int J Clin Exp Hypn* 2004;**52**:132–58.

Sheehan PW, Perry CW. *Methodologies of Hypnosis: A Critical Appraisal of Contemporary Paradigms of Hypnosis*. Lawrence Erlbaum, Oxford. 1976.

Shor RE, Orne EC. *Manual: Harvard Group Scale of Hypnotic Susceptibility*. Palo Alto: Consulting Psychologists Press, 1962.

Tart CT. Self-report scales of hypnotic depth. *Int J Clin Exp Hypn* 1970;**18**:105–25.

Terhune DB, Cardeña E. Nuances and uncertainties regarding hypnotic inductions: toward a theoretically informed praxis. *Am J Clin Hypn* 2016;**59**:155–74.

Terhune DB, Cleeremans A, Raz A, *et al*. Hypnosis and top-down regulation of consciousness. *Neurosci Biobehav Rev* 2017;**81**:59–74.

Tsuchiya N, Koch C. Continuous flash suppression reduces negative afterimages. *Nat Neurosci* 2005;**8**:1096.

Wagstaff GF, Cole JC, Brunas-Wagstaff J. Measuring hypnotizability: the case for self-report depth scales and normative data for the Long Stanford Scale. *Int J Clin Exp Hypn* 2008;**56**:119–42.

Weitzenhoffer AM. Hypnotic susceptibility revisited. *Am J Clin Hypn* 1980;**22**:130–46.

Weitzenhoffer AM, Hilgard ER. *Stanford Hypnotic Susceptibility Scale, Form C*. Palo Alto: Consulting Psychologists Press, 1962.

Weitzenhoffer AM, Sakata K. Role of visual fixation in the production of hypnotic behavior. *Am J Clin Hypn* 1970;**13**:104–7.

Weitzenhoffer AM, Gough PB, Landes J. A study of the Braid effect: (hypnosis by visual fixation). *J Psychol* 1959;**47**:67–80.

Woody EZ, Barnier AJ. Hypnosis scales for the 21st Century: what we need and how we should use them? In: Nash M, Barnier A (eds), *The Oxford Handbook of Hypnosis: Theory, Research, and Practice*. Oxford: Oxford University Press, 2008, 255–82.

Woody EZ, Sadler P. Dissociation theories of hypnosis. In: Nash M, Barnier A (eds), *The Oxford Handbook of Hypnosis: Theory, Research, and Practice*. Oxford: Oxford University Press, 2008, 81–110.