International Journal of Clinical and Experimental Hypnosis

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/nhyp20

Test–Retest Reliability of the Stanford Hypnotic Susceptibility Scale, Form C and the Elkins Hypnotizability Scale

Zoltan Kekecs, Lynae Roberts, Hyeji Na, Ming Hwei Yek, Elizabeth E. Slonena, Ezrhiel Racelis, Tamara A. Voor, Robert Johansson, Pietro Rizzo, Endre Csikos, Vanda Vizkievicz & Gary Elkins

To cite this article: Zoltan Kekecs, Lynae Roberts, Hyeji Na, Ming Hwei Yek, Elizabeth E. Slonena, Ezrhiel Racelis, Tamara A. Voor, Robert Johansson, Pietro Rizzo, Endre Csikos, Vanda Vizkievicz & Gary Elkins (2021) Test–Retest Reliability of the Stanford Hypnotic Susceptibility Scale, Form C and the Elkins Hypnotizability Scale, International Journal of Clinical and Experimental Hypnosis, 69:1, 142-161, DOI: 10.1080/00207144.2021.1834858

To link to this article: https://doi.org/10.1080/00207144.2021.1834858

INTERNATIONAL JOURNAL OF CLINICAL AND EXPERIMENTAL HYPNOSIS
2021, VOL. 69, NO. 1, 142–161
https://doi.org/10.1080/00207144.2021.1834858

Test–Retest Reliability of the Stanford Hypnotic Susceptibility Scale, Form C and the Elkins Hypnotizability Scale

Zoltan Kekecs[a,b], Lynae Roberts[c], Hyeji Na[c], Ming Hwei Yek[c], Elizabeth E. Slonena[c], Ezrhiel Racelis[c], Tamara A. Voor[c], Robert Johansson[b], Pietro Rizzo[b], Endre Csikos[a], Vanda Vizkievicz[a], and Gary Elkins[c]

[a]Department of Affective Psychology, ELTE, Eötvös Loránd University, Budapest, Hungary; [b]Department of Psychology, Lund University, Skåne, Sweden; [c]Department of Psychology and Neuroscience, Baylor University, Waco, Texas, USA

ABSTRACT

This project aimed to assess the consistency of hypnotizability over repeated assessments when measured by the Stanford Hypnotic Susceptibility Scale: Form C (SHSS:C), and the Elkins Hypnotizability Scale (EHS) and to contrast score distribution and pleasantness of these scales. University students were administered either the SHSS:C or the EHS twice with a one-week delay by separate experimenters. Test–retest reliability of the EHS and the SHSS:C was r_s =.82 (.71-.92) and r_s =.66, 95% (.47-.86), respectively (Spearman's correlation). Hypnotizability was comparable at test and retest in the EHS group, SHSS:C scores decreased by the retest. We found that the SHSS:C produced higher scores than the EHS, and the pleasantness of the 2 scales was comparable. Overall, our results supported the reliability of the EHS, while SHSS:C scores were more inconsistent between the 2 assessments. More research is warranted.

ARTICLE HISTORY

Received 28 September 2019
Revised 28 February 2020
Accepted 18 March 2020

Introduction

People's responsiveness to suggestions during hypnosis have shown individual differences. This ability to respond to hypnotic suggestions is measured by hypnotizability scales (Elkins et al., 2015). Because hypnotic depth and spontaneous phenomena following a neutral hypnotic induction vary as a function of hypnotizability (Cardeña et al., 2013) and because of hypnotizability's consistent (albeit small) correlation with clinical effects of hypnotic interventions (Montgomery et al., 2011), we rely heavily on hypnotizability scales in hypnosis research. Due to hypnotizability being assessed commonly in hypnosis research, it is paramount to have highly reliable measures that are also practical and safe to administer. However, data are missing on the consistency over time of the Stanford Hypnotic Susceptibility Scale: Form C (SHSS:C; Weitzenhoffer & Hilgard, 1962) and the Elkins Hypnotizability Scale (EHS; Elkins, 2014). In the present study, we measured the test-retest reliability of these two commonly used hypnotizability scales, while contrasting their pleasantness, the average scores they produced, and whether they showed evidence of diminished responsiveness from first to second administration.

CONTACT Zoltan Kekecs kekecs.zoltan@gmail.com Department of Affective Psychology, Eötvös Loránd University International, Budapest 1064 Hungary.

Supplemental data for this article can be accessed on the publisher's website.

The SHSS:C has been considered the gold standard scale in the field, but over time this scale has been criticized on multiple accounts. Lack of safety and pleasantness, lengthy administration (60 minutes), and dichotomous ratings of responses were among these criticisms (see Kekecs et al., 2016, for a summary). The EHS, a recently developed alternative to the SHSS:C, was designed to fit the needs of modern hypnosis research, with safety, pleasantness, and speed (30 minutes administration time) in mind. The hypnotizability scores measured with these two scales are highly correlated (r = .86) when they are measured back-to-back by the same experimenter (Kekecs et al., 2016), and they are commonly used in hypnosis research. However, information on the consistency of these measures over time is limited. Information on the reliability of the scales is not only important to establish the consistency of their scores, but it also puts data on convergent validity into wider perspective, since reliability of measurement puts an upper limit on the possible correlation of two scales (Carmines & Zeller, 1979). The lack of understanding of the test–retest reliability of these two scales is a critical literature gap.

To our knowledge, there is only one published report on the test–retest reliability of the SHSS:C, which indicates a test–retest correlation of .75, although in that study the administration of the scale did not follow the standard protocol (Taslimbakhsh et al., 2017). For example, the SHSS:C was administered together with the Hypnotic Induction Profile (HIP, Spiegel et al., 1976), and the items in the SHSS:C were self-scored. The lack of data on test–retest reliability of the SHSS:C is surprising given its critical importance in the field. Nevertheless, one might expect to see a relatively good test–retest reliability for the SHSS:C based on other available data. We could infer the retest reliability based on data from other related scales. For example, the test–retest correlation of SHSS, Forms A and B, is .83 – .90 (Hilgard, 1965), and there are also data on the long-term retest correlation of hypnotizability measured with the SHSS:A over a 25-year follow-up (.64 – .82; Piccione et al., 1989). However, it is uncertain how much these results can be extended to the SHSS:C, due to the mismatch in the types of test suggestions these scales contain. Retest correlations of the shared items of SHSS:A and SHSS:C are only .60 – .77, which indicates that the retest reliability of the SHSS:C might be lower than that measured for Forms A and B (Hilgard, 1965).The internal consistency of the SHSS:C is also good, Chronbach's alpha reaching .85 (Hilgard, 1965), which is a prerequisite for high test–retest reliability.

Data on the test–retest reliability of the EHS are also lacking, so in terms of reliability we can only rely on reports of internal consistency. The two previous studies assessing reliability of the scale reported Cronbach's alpha to be .85 and .78 (Elkins, 2014; Kekecs et al., 2016). Past research on the consistency of hypnotizability scores over time and convergent validity of different hypnotizability scales is also limited by the fact that the scales are often administered by the same experimenter. Thus, knowledge about the results of the previous administration can leave room for experimenter biases, potentially artificially increasing the correspondence of the two scores. Furthermore, if the two scales are administered shortly after one another, this could lead to the enhancement of responsiveness on the second administration (Hilgard et al., 1961). These limitations increase our uncertainty about the consistency of SHSS:C and EHS scores over time. For example, the work of Fassler et al. (2008) indicated that responsiveness to suggestions diminished by the second administration of the Carleton University Responsiveness to Suggestion Scale (CURSS; Spanos et al., 1983) when the administrations were 1 week apart. However, so far this effect was only shown with the CURSS. It is important to assess whether scores would decrease by the second administration for the SHSS:C and EHS as well to understand the stability of these measures. There is also limited information about the comparability of certain properties of the two scales, such as the average hypnotizability scores produced by them and the subjective pleasantness of undergoing SHSS:C and EHS.

The current study aims to assess the test–retest reliability of the SHSS:C and the EHS and the overall consistency of the scores produced by these scales after repeated administration. Furthermore, we compared the average hypnotizability scores produced by these scales and the reported pleasantness of these scales by participants. To avoid potential biases from demand characteristics, researcher expectancy, or other biases resulting from the same researcher administering the hypnotizability test multiple times with the same participants, the administrations of the scales were separated by multiple days, and the researchers administering the scales were blinded to participants' previous hypnotizability scores.

We expected to find that both the EHS and the SHSS:C will have a good (>8) test–retest reliability based on high internal consistency of the scales and prior data presented above. We also expected to find a small decrease in the hypnotizability scores due to the above-mentioned result by Fassler et al. (2008) on diminished suggestibility on the second administration of the CURSS. Since the two scales measure the same construct (hypnotizability) on the same scale (0–12), we expected to find comparable mean scores on the two scales. Finally, we also expected that the EHS will be rated as more pleasant compared to the SHSS:C, because one of the goals during the design of the EHS was to create a scale that is pleasant for the participant, avoiding items that could cause discomfort, to make it easier to integrate into clinical research and practice.

The six specific hypotheses derived from these expectations are listed below. These hypotheses were preregistered after the data collection was completed but before analysis (see https://osf.io/xa6ym in the files menu):

Hypothesis 1a. We expected that the EHS would have good (> .8) test-retest reliability.

Hypothesis 1b. We expected that the SHSS:C would have good (> .8) test-retest reliability.

Hypothesis 2a. We expected that EHS scores would slightly decrease by the second administration.

Hypothesis 2b. We expected that SHSS:C scores would slightly decrease by the second administration.

Hypothesis 3. We expected that the average score measured by the EHS and the SHSS:C would be similar.

Hypothesis 4. We expected that the EHS would be rated as more pleasant compared to the SHSS:C.

Method

The transparency checklist report for this study is included in the online supplement.

Sample Size Planning

A sample size target of 100 participants who completed both sessions was set for this study (*N* = 50 participants per each scale). This sample size target was set based on the staff capacity of the lab for running this (nonfunded) project, and the projected recruitment capacity.

Participants

Participants were Baylor University students recruited through IRB-approved flyers and the Baylor SONA system, which is an online research participant management system. Participants recruited through SONA were awarded course credit at the end of study participation.

Eligible participants were English-speaking individuals of at least 18 years of age with the ability to give his or her own consent for study participation. Individuals with a self-reported history of borderline personality disorder, schizophrenia, or other psychiatric conditions involving psychosis were excluded from the study due to contraindication with hypnosis (Kekecs et al., 2016; MacHovec, 1986; Walker, 2016). (Previous experience with hypnosis was not exclusionary.)

Measures

The following measures were used in this study.

Demographic Questionnaire

This is an 11-item self-report questionnaire that included questions regarding age, gender, race/ethnicity, undergraduate classification, marital status, and psychiatric diagnoses.

Elkins Hypnotizability Scale (EHS)

The EHS (Elkins, 2014) is a brief six-item measure of hypnotizability that takes approximately 25 minutes to administer by a trained assessor. The EHS starts with a standardized introduction of the scale followed by a standardized induction, which includes suggestions for relaxation and calmness. The six items include: (1) arm heaviness or immobilization, (2) arm levitation, (3) imagery involvement or dissociation of being in a garden, (4) positive hallucination of the smell of a rose, (5) positive hallucination of a blue block, and (6) posthypnotic amnesia. Responses are scored based on subjective experience of the participant and behavioral observation by the assessor. A total score is obtained from summing all item scores, producing a range of scores from 0 to 12. The total score can be used to classify participants into hypnotizability ranges: very low (0–1), low (2–3), middle (4–8), high (9–10) and very high (11–12) (Elkins, 2014). The EHS demonstrates good internal consistency (.78 – .85) (Elkins, 2014; Kekecs et al., 2016). Correlations between the EHS and SHSS:C (Weitzenhoffer & Hilgard, 1962) range from .82 to .91, indicating good convergent validity (Elkins, 2014; Kekecs et al., 2016).

Stanford Hypnotic Susceptibility Scale: Form C (SHSS:C)

The SHSS: C (Weitzenhoffer & Hilgard, 1962) is a 12-item measure of hypnotizability that takes approximately 45 to 60 minutes to administer by a trained assessor. Items range from simple (motor responses) to difficult (posthypnotic amnesia). The 12 items include: (1) hand lowering, (2) moving hands apart, (3) mosquito hallucination, (4) taste hallucination, (5) arm rigidity, (6) dream, (7) age regression, (8) arm immobilization, (9) anosmia to ammonia, (10) hallucinated voice, (11) negative visual hallucination, (12) posthypnotic amnesia. Responses are scored as a dichotomous pass (score of 1) or fail (score of 0) scoring based on assessor observation and summed to produce a total score ranging from 0 to 12. The total score can be used to classify participants into hypnotizability ranges. Importantly, these ranges are different from the ones used in the EHS: very low (0–1), low (2–3), middle (4–8), high (9–10) and very high (11–12) (Weitzenhoffer & Hilgard, 1962). Good internal consistency of the SHSS:C has been reported at .85 (Hilgard, 1965). The SHSS:C has been considered the gold standard hypnotizability scale and frequently serves as the standard against which all other scales are compared to (Kihlstrom, 1985; Woody & Barnier, 2008).

Pleasantness Scale

Participants rated the overall pleasantness of the hypnotizability scale and pleasantness of each item on the hypnotizability measure on a 10-point Likert scale, ranging from 1 (*not at all pleasant*) to 10 (*very pleasant*).

There were a number of other scales administered to participants. Measures about self-reported relaxation and absorption, hypnosis depth, perceived effectiveness of hypnosis, attitudes about hypnosis and motivation to get hypnotized, and trait dissociative ability. However, the research questions preregistered for this paper did not concern these measures, so results on these are not reported here. Information on these scales can be found in the online supplement, which can be found in the Materials section of the OSF page of the project (https://osf.io/kg8tx/).

Procedures

This study was approved by the Baylor University Institutional Review Board (IRB number: 698,395). Informed consent was completed prior to the start of the study. The study was executed according to the principles described by the Declaration of Helsinki.

In this study, participants were administered either the EHS or the SHSS:C twice during two lab visits that were 1 week apart. The period of 1 week was chosen because it was judged to be a short enough time period so that no substantial change in the underlying characteristic (hypnotizability) would take place, since hypnotizability is generally understood as a stable trait (Piccione et al., 1989). This time period was also considered to be a long enough time for any potential lingering effect of the first hypnosis administration to disappear (for example, experiencing two hypnosis inductions back-to-back may have a deepening effect on the second induction). A 1 week delay was also chosen for practical reasons: it is likely that the participants would be free at the same time slot 1 week after the first session, and doing repeated research sessions on the same weekday may also control for any potential day-of-the-week effects. The sequence of which participant was administered EHS or SHSS:C was determined randomly before data collection was started. This procedure was set up with the assumption that all experimenters were trained in administering both scales. However, during the course of the project several experimenters joined who were only trained in administering one of the scales. For this reason, the prerandomized sequence was broken, and group allocation was based on which scale the experimenter available at a given time slot was trained in.

Lab visits were conducted one-on-one. During the first lab visit, participants provided informed consent, followed by a demographic questionnaire and a series of numbered analog scales with ratings about cognitive expectancy, dissociation, trust, and resistance. Participants were then given information about hypnosis, and any myths or misconceptions about hypnosis were addressed by the researcher. The APA Division 30 Brochure about hypnosis was used in this process (see Materials on Open Science Framework [OSF]). Prior to hypnotizability scale administration, all participants rated their relaxation and absorption levels (see online supplement). Participants were then administered either the EHS or the SHSS:C (in a one-on-one setting). Participants who were randomly assigned to the EHS group were asked to rate their relaxation and absorption following the item on positive hallucination of a rose scent. After completing the hypnotizability scale, participants rated their posthypnotic relaxation and absorption levels, rated pleasantness of the items on the hypnotizability scale, and completed the Dissociative Ability Scale. Posthypnotic debriefing was provided as needed.

One week later, participants returned for their second lab visit. During this visit, participants underwent the same hypnotizability scale as their first visit, administered by a different researcher. Similar to the first visit, all participants provided relaxation and absorption ratings before and after the hypnotizability scale, and participants in the EHS group provided additional relaxation and absorption ratings after the item on positive hallucination of a rose scent. All participants provided pleasantness ratings for the items on the hypnotizability scale, and completed the Dissociative Ability Scale a second time. Posthypnotic debriefing was again provided as needed.

Data Management

Data were recorded by participants and experimenters on preprinted charts. Data from these charts were entered in duplicate and later compared to ensure accuracy of data entry. The few mismatches (0.3%) between the two data entry sheets were reconciled based on the original charts.

The analysis plan was devised and registered by researchers who had no access to the data until the preregistration was complete.

Data of participants who had missing data on any item of the hypnotizability scale were discarded from all confirmatory analysis involving hypnotizability. Data of participants who had missing data on the overall pleasantness rating of the hypnosis session on either of the two sessions were discarded from all confirmatory analysis involving this variable. Missing data were not replaced in this study. The cases that were excluded due to missing data on any of the relevant outcome measures (i.e., ratings of hypnotizability and pleasantness from session one and two) were excluded on the basis of not showing up for reassessment. This included 11 cases excluded from the EHS data analysis, and 3 cases excluded from the SHSS: C data analysis.

Data availability statement: Data collected in this study cannot be shared because the participants enrolled in this study did not consent to such data sharing.

Data Analysis

We hypothesized that the EHS and the SHSS:C would demonstrate good (> .80) test–retest reliability. Assumptions of normality were tested with the Shapiro-Wilks test, with the criterion for nonnormality set to $p < .05$. Because the normality assumption of parametric regression was violated in the case of total hypnotizability scores, the test–retest reliability of both scales was assessed by means of Spearman's rank-order correlation. As a complement to these point estimates, confidence intervals for these test–retest correlations were bootstrap estimated with the help from the *boot* package (Canty & Ripley, 2017) in R utilizing 10 000 resamples per parameter.

Equivalence tests were conducted to assess whether participants exhibited diminished responsiveness upon reassessment. For each participant, the difference in total score between the first and second hypnotizability assessment was calculated. A Smallest Effect Size of Interest (SESOI) was set by establishing a symmetric equivalence bound of −1 to +1 raw difference score for both the EHS and the SHSS:C. It was specified that if the 95% confidence interval of the difference between administrations was completely contained within the equivalence bound, the effect would be deemed as too small to be interesting. The same rationale was used for the equivalence tests described below. Because all distributions were nonnormal, basic type confidence intervals for all equivalence tests were bootstrapped using 10,000 resamples per parameter. Complementary Wilcoxon signed-rank tests were then computed for the two scales to assess whether there is a difference in the scores between the first and second administration of the scales.

The hypnotizability scores measured at first and second administrations were then aggregated by averaging the score within participants for both the EHS and the SHSS:C. The difference between the mean aggregated hypnotizability scores evoked by the two scales was used to determine whether the population mean of hypnotizability is equivalent when measured with the EHS and SHSS:C. Because these aggregated responsiveness scores were not normally distributed, a nonparametric Wilcoxon signed-rank test was used to test this hypothesis. A SESOI was specified by establishing a symmetric equivalence bound of −1 to +1 raw difference scores between the means.

Finally, to assess whether the EHS was rated as generally more enjoyable than the SHSS: C, we compared the pleasantness ratings given after the administrations of the two scales.

Overall pleasantness ratings were aggregated by averaging the scores within participants for both scales. We will refer to overall pleasantness ratings as pleasantness rating from hereon. In this paper we do not analyze item-by-item pleasantness ratings, since this was not preregistered, and we expect similar results to that of the overall pleasantness ratings. Because the mean pleasantness ratings violated the assumption of normality, a Mann-Whitney *U* test was conducted to test the hypothesis that the EHS was rated as more overall pleasant than the SHSS:C. An SESOI was specified as a ±1 mean difference in pleasantness between the scales.

All statistical procedures were carried out in version 3.5.1 of the R software for statistical computing (Team, 2019) with the alpha thresholds of statistical significance set to $p < .05$ for all confirmatory tests. Because all data distributions were nonnormal, standardized effect sizes with bias-corrected-and-accelerated (BCa) confidence intervals were estimated with bootstrapping methods as advised by Kelley (2005) with help from the *bootES* package (Gerlanc & Kirby, 2013) utilizing 10,000 resamples per parameter. All basic type confidence intervals for the unstandardized differences between group means were likewise bootstrapped in this manner. Other R packages utilized for data management, visual inspection and assumptions testing were, in alphabetical order, *car* (Fox & Weisburg, 2011), *dplyr* (Wickham et al., 2018), *ggplot* (Wickham, 2016), *lmtest* (Zeileis & Hothorn, 2002), *psych* (Revelle, 2017) and *rio* (Chan et al., 2018). The confirmatory analysis plan and the associated R-code was preregistered through OSF before viewing the data. The preregistered and final analysis codes are accessible as digital supplementary materials through https://osf.io/wcegj/.

Results

A total of 112 undergraduate students at Baylor University participated in the experiment in exchange for course credits. Fourteen of them were not included in this analysis, 1 participant due to the researcher's scoring being unclear, and 13 on account of not showing up for the retest. Those who did not show up for the second session in the EHS group had a lower mean hypnotizability score on their first assessment ($M = 4.82$, $SD = 2.27$) compared to those who attended both sessions ($M = 5.49$, $SD = 2.86$), but this difference was not statistically significant ($W = 322.5$, $p = .44$). (There were only 3 dropouts in the SHSS:C group, so no statistical analysis could be carried out on the difference between dropouts and nondropouts).

The final sample for this analysis thus consisted of 98 participants (M = 28, F = 70) with a mean age of 18.97 (*SD* = 1, Range =. 18–25), who completed the entire experimental procedure by participating in both hypnotizability assessment sessions and providing complementary pleasantness ratings. Most of them were freshman ($n = 55$), but sophomores ($n = 36$) and juniors ($n = 7$) were also represented. Seventy-one percent identified themselves as Caucasian, 16% as African American, and 7% as Asian. Fifteen percent categorized their ethnicity as "other." None of them were currently married. The groups did not differ significantly in any of the demographic variables. Table 1 displays the distributions of participant demographics across the two experimental groups. Nineteen experimenters conducted the hypnotizability assessments, the number of sessions conducted by one experimenter ranging between 1 to 13. Summary statistics about the most

Table 1. Distribution of Participant Demographics Across the Two Experimental Groups

	EHS (*n* = 51)	SHSS:C (*n* = 47)	Test of group difference
Age	Mean = 19, *SD* = 0.78	*Mean* = 18.97, SD = 1	*t*(87) =.29, *p* =.77
Gender	Male = 13, Female = 38	Male = 15, Female = 32	χ2(1) = 0.23, *p* =.63
Attrition	11/62 (16%)	3/50(6%)	χ2(1) = 3.49, *p* =.06
Marital status	All unmarried	All unmarried	NA
Employed	14/51 (27%)	15/47 (32%)	χ2(1) = 0.23, *p* =.62
Level of education	Freshman = 31 (61%) Junior = 2 (4%) Sophomore = 18 (35%)	Freshman = 24 (51%) Junior = 5 (11%) Sophomore = 18 (38%)	χ2(2) = 2.01, *p* =.36
Ethnicity	Asian = 3 (6%) Black or African American = 9 (18%) Indian, Chinese, Middle Eastern = 1 (2%) Mestizo = 1 (2%) Middle Eastern = 1 (2%) White = 36 (71%)	Asian = 4 (9%) Black or African American = 7 (15%) Italian, Japanese = 1 (2%) Hispanic/Latino = 1 (2%) White = 34 (72%)	NA

important outcome measures broken down by experimenters are presented in the online supplement Table S1.

The test–retest reliability was good for the EHS, r_s = .82, 95% CI [0.71, 0.92]. As shown in Figure 1, 16% of those who were allocated to the EHS scored identically on both administrations. In total, the test and retest scores fell within one point of each other in 63% of the cases and within two points in 88% of the cases. When the EHS scores were assigned into their corresponding hypnotizability range, it was found that 55% of all participants were placed in the matching category upon reassessment. For the sake of a clearer comparison between the two scales, the scores were also assigned into their corresponding SHSS:C hypnotizability ranges (see below); in this case 67% of all participants were placed in the matching category upon reassessment.

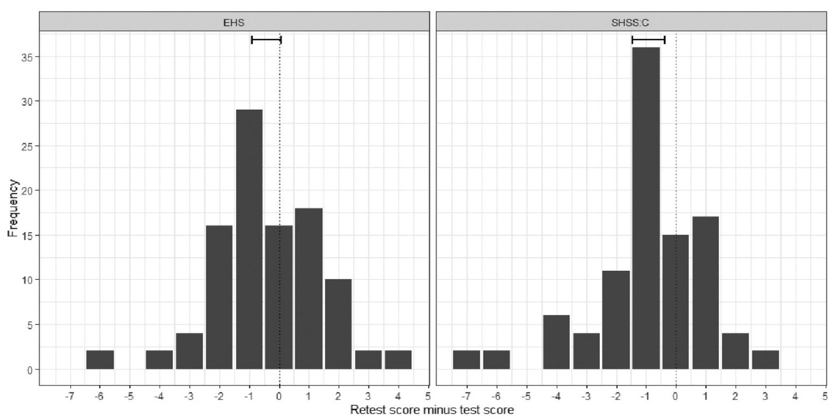

Figure 1. Distribution of EHS and SHSS:C Test–Retest Score Difference
The error bars represent the unstandardized confidence interval of the difference of first-to-second administration of the hypnotizability scale. The vertical dotted line marks zero (No difference between first and second administration).

The test–retest reliability of the SHSS:C was r_s =.66, 95% CI [0.47, 0.86], so it did not reach our preregistered criterion for good reliability. As shown in Figure 1, 15% of all participants who underwent the SHSS:C received identical scores upon reassessment. In total, the test and retest scores fell within one point of each other in 68% of the cases and within two points in 83% of the cases. When the SHSS:C scores were assigned into their corresponding hypnotizability range, it was found that 47% of all participants were placed in the matching category upon reassessment. For the sake of a clearer comparison between the two scales, the scores were also assigned into their corresponding EHS hypnotizability ranges (see above); in this case 64% of all participants were placed in the matching category upon reassessment.

The low reliability of the SHSS:C was a surprising result, so we conducted some post hoc data exploration in an attempt to find the source or sources of the low test–retest correlation. During this exploration, we noted that 2 participants in the SHSS:C group received hypnotizability scores at the second administration that were 6 and 7 points lower than their score at the first assessment. The removal of these extreme cases would increase SHSS:C's reliability from .66 to .78 (still below the expected reliability, but closer). There was 1 participant with similarly great difference between test and retest in the EHS group. Removing this participant from the dataset raised the scale's reliability but did not affect the reliability of the EHS substantially (reliability changed from .82 to .85). We did not preregister outlier exclusion in our analysis plan, and there were no indications in the experiments' notes that would point to irregularities in these particular cases that would warrant the exclusion of these cases, so the analyses presented below are performed on the full dataset.

Participants who were allocated to the EHS scored an average of 0.41 raw units lower upon reassessment, but this difference was not statistically significant, $Z = 598$, $p = .12$, $g_{rm} = -.13$, BCa 95% CI [−0.29, 0.03]. The confidence interval of the unstandardized difference was completely contained within the ±1 equivalence bound, 95% CI [−0.90, 0.07]. The research hypothesis specifying diminished responsiveness upon reassessment for the EHS group was thus rejected, and the difference between the first and second administration scores was deemed as negligible.

Changes in individual item scores for EHS between the two sessions are explored in Table 2. The greatest difference in pass frequency from test to retest is seen in the elbow-lift task and in hallucinating a rose smell.

There was, on the other hand, a statistically significant difference in the mean of the total scores of the first and second administration of the SHSS:C, so that participants scored an average of 0.91 units raw score lower upon reassessment, $Z = 624$, $p = .003$, $g_{rm} = -.36$, BCa 95% CI [−0.59, −0.16]. Moreover, the unstandardized confidence interval of this difference spanned outside of the ±1 equivalence bound, 95% CI [−1.46, −0.36]. The research hypothesis that repeated administration of the SHSS:C would result in diminished responsiveness was thus retained, and the effect size was deemed as nonnegligible.

Changes in individual item scores for SHSS:C between the two sessions are explored in Table 3. The table reveals that the mosquito hallucination, taste hallucination, dream, anosmia, and posthypnotic amnesia items all showed at least 10 percentage points or more decrease on average by the second administration. Note that these are characterized as "cognitive" or "perceptual" suggestions (Acunzo & Terhune, 2021).

Table 2. Changes in EHS Items Scores in Time

Item	Test pass frequency (%)	Retest pass frequency (%)	Retest pass frequency minus test pass frequency (%)	Same score across test and retest frequency (%)
1• Subjective heaviness	92	86	−6	82
2• Arm immobilization	47	49	2	78
3• Subjective lightness	75	75	0	84
4• Arm levitation	41	43	2	86
5• Elbow lift	24	12	−12	84
6• Imagery	78	75	−4	76
7• Dissociation	45	43	−2	78
8• Faint rose smell	59	47	−12	69
9• Distinct rose smell	18	10	−8	84
10• Vague hallucination	41	43	2	78
11• Clear hallucination	0	4	4	96
12• Recalls one or fewer items	24	24	0	84

Table 3. Changes in SHSS:C Items Scores in Time

Item	Test pass frequency (%)	Retest pass frequency (%)	Retest pass frequency minus test pass frequency (%)	Same score across test and retest frequency (%)
1• Hand lowering	100	98	−2	98
2• Moving hands apart	100	94	−6	94
3• Mosquito hallucination	87	77	−11	85
4• Taste hallucination	74	57	−17	83
5• Arm rigidity	66	62	−4	70
6• Dream	66	55	−11	72
7• Age regression	21	17	−4	79
8• Arm immobilization	47	49	2	77
9• Anosmia to ammonia	64	45	−19	60
10• Hallucinated voice	2	6	4	91
11• Negative visual hallucination	21	17	−4	87
12• Posthypnotic amnesia	43	23	−19	77

On average, the total hypnotizability scores of those participants who were allocated to the SHSS:C (M = 6.45) was 1.17 units raw score higher than for those who were allocated to the EHS (M = 5.28). The confidence interval of the unstandardized difference between the mean responsiveness elicited by the two scales spanned outside of the ±1 equivalence bound, 95% CI [0.40, 1.95]. This difference was also statistically significant with a two-tailed criterion, U = 1499.5, p = .032, g = .45, BCa 95% CI [0.05, 0.84]. Therefore, we interpreted this finding as evidence against the null hypothesis that the population mean of hypnotizability would be equivalent when measured with the two scales. To help to understand the source of this difference, the distribution of the hypnotizability scores on the two scales averaged over the two administrations is displayed on Figure 2. The main difference is that there are more participants scoring low on the EHS while very low scores are absent from the SHSS:C group.

The difference between the means of the aggregated pleasantness rating from both scales was 0.18 units raw score higher for the EHS (M = 8.4) than for the SHSS:C (M = 8.2), 95% CI [−0.21, 0.60], meaning that the responses on average were very close to the end of the scale

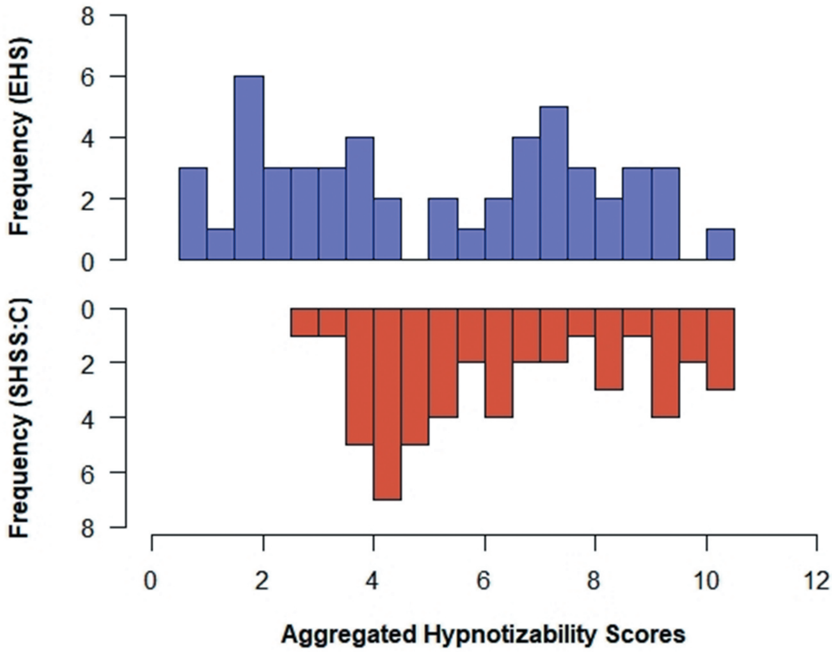

Figure 2. Frequency Histograms of the Aggregated Total Hypnotizability Scores (The Mean of the Assessment and Reassessment Total Hypnotizability Score for Each Participant) in Purple for the EHS (Top) and in Red for the SHSS:C (Below) Groups

anchored at "very pleasant" in both groups. The 95% confidence interval of this difference was fully contained within the equivalence bound and the rank sum test was not statistically significant, $W = 1359$, $p = .25$, $g = .13$, BCa 95% CI [−0.26, 0.55]. The research hypothesis that the EHS would be rated as generally more pleasant than the SHSS:C was thus rejected, instead, we found evidence that the pleasantness of the two assessment methods was equivalent. Since this result was contrary to our hypothesis, we wanted to explore pleasantness ratings further. We expected the SHSS:C to be less pleasant due to some of its items containing unpleasant stimuli (anosmia to ammonia, mosquito hallucination), and previous reports of negative reactions to age regression items in hypnotizability scales (see, eg., Cardeña & Terhune, 2009). Thus, we explored the pleasantness ratings of each individual item. When looking at the item-by-item pleasantness ratings, we found that the mean of the item ratings was numerically greater for the EHS ($M = 6.59$, $SD = 0.47$) compared to the SHSS ($M = 5.43$, $SD = 0.88$). We did not conduct a statistical significance testing here, since this was not a preregistered analysis. When looking at the individual items, we found that the mosquito hallucination and anosmia to ammonia items were indeed among the items rated as least pleasant (mean pleasantness rating 4.46 and 4.70, respectively), although the age regression item was among the most pleasant SHSS:C items with a mean pleasantness rating of 6.71. The least pleasant item in the EHS was the amnesia item, but the mean pleasantness rating of this item (5.89) was still greater numerically than the mean pleasantness rating of 8 out of the 12 SHSS:C items. The mean pleasantness ratings for each item are displayed in Table S2 and Table S3 in the online supplement.

Discussion

This project assessed the consistency of hypnotizability scores over two assessments when measured by the EHS and the SHSS:C and compared the distribution of hypnotizability scores and pleasantness ratings between the two scales. Test–retest reliability between two scale administrations was good for the EHS, which is in line with our original hypothesis. However, reliability was questionable and lower than expected for the SHSS:C. We found that hypnotizability scores decreased significantly, by about 1 point on average, from the first to the second administration when measured by the SHSS:C, while scores did not decrease significantly from the first to the second administration when measured by the EHS. Furthermore, the observed difference between the first and second administration was significantly lower than 1 point, making the two measurements using the EHS equivalent according to our preregistered criteria. The mean hypnotizability score was about 1 point higher in the group where the SHSS:C was used compared to the EHS. We also found that the two scales were comparable in pleasantness.

Surprisingly, the test–retest consistency of the SHSS:C scores did not reach 0.8, or prespecified criteria for good reliability, whereas the EHS did surpass this threshold. Previous test–retest data for the SHSS:C and EHS are limited. The only study we could find directly measuring the test–retest reliability of the SHSS:C indicated a pre-to-posttest correlation of .75, which is also below 0.8 (Taslimbakhsh et al., 2017). However, the test-retest correlation of SHSS:A and B, the precursor scales of the SHSS:C, was found to be good (.83 – .90, Hilgard, 1965). The low reliability may be an indication that previous studies overestimated the test–retest reliability of hypnotizability scales, which might be due to the lack of blinding. In prior studies, the same researchers may have delivered the scale at both sessions, potentially leading to bias because of the knowledge about the previous score of the participant. In the present study, we avoided such experimenter effects with subsequent administrations being performed by different researchers who were blinded about the previous scores of the participants. The difference in retest reliability between the earlier forms of the SHSS might also be due to the cognitive and perceptual test suggestions that were added in Form C. Exploration of outliers revealed that the low consistency of SHSS:C in our study may be partially due to a few extreme cases who scored much lower on the second administration than the first one. Nevertheless, reliability still remained below 0.8 even without these extreme cases.

We also assessed the stability of the hypnotizability categories assigned by the two scales. The two scales use different categorization schemes. The EHS uses very low (0–1), low (2–3), middle (4–8), high (9–10), and very high (11–12), whereas the SHSS:C uses low (0–4), medium (5–7), high (8–10), and very high (11–12). When assessed in their own respective categorization schemes, only about 50% of the participants were categorized into the same hypnotizability range in the second administration by both scales. Looking at the relatively high test–retest reliability of the EHS, its poor categorization stability might come as a surprise to some readers. The reason is that the EHS's original categorization scheme uses 5 categories instead of the SHSS:C's 4 categories, splitting the low range into two very narrow categories "very low" and "low," both of which only include two scores. Because hypnotizability scores from test to retest varied ±2 points on average and because the EHS group had a lot of participants with low scores in it, it was common for people to move between these two categories from test to retest. This also could explain why the categorization stability was numerically higher in the SHSS:C group when using the EHS's categorization scheme. Since the SHSS:C group had only seven participants with a score below 4 at any administration, the narrowness of the two low hypnotizability ranges did not affect categorization stability that much in this group. All in all, our results indicate that the hypnotizability categories are quite unstable and that researchers should use the numerical scores instead whenever possible, as to not overly rely on the categories.

A larger sample size study is necessary to get a better estimate for the test–retest reliability of SHSS:C. This is especially important as the SHSS:C is often used in the validation of other hypnotizability scales, and the reliability of a scale puts an upper limit on the achievable convergent validity with other scales. For example, if the test–retest correlation of measure A of a construct is .7, the highest possible correlation between measure A and measure B of the same construct is also .7 in the population, assuming that the conditions are the same for the two measurements (for example, the same delay is used between the administration of the two measures as in the test–retest study). If later studies confirm the poor reliability of SHSS:C, it may be necessary to consider other, more reliable measures as gold standards for future validation studies. The reliability of the EHS surpassed the preset criterion for good reliability, but it is important to realize that the current study was not powered to directly contrast the reliability of the two scales in this study, and this study does not provide decisive evidence for the superiority of the test–retest reliability of the EHS over the SHSS:C. A larger scale study is necessary to make such a direct comparison, and the present report may provide valuable data to properly power such a study.

The variation in the hypnotizability scores from the first to the second administration seemed to be random for the EHS, while we found a systematic negative effect of time for the SHSS:C in addition to the random variation. That is, according to our statistical tests first and second administration scores measured by the EHS were equivalent (significantly within the preregistered SESOI range), while scores on the second administration were significantly lower compared to the first with the SHSS:C (and not significantly within the SESOI range). Fassler et al. (2008) found a similar decreased responsiveness over time when using the CURSS and attributed the diminished scores to boredom and disengagement from the procedure. We did not measure boredom and engagement with the procedure in our study, but it is possible that the novelty of the SHSS:C engaged participants during the first administration, but the interest dropped in the repeated exposure because of the lengthy procedure. Likewise, due to its relative brevity, the EHS was perhaps less prone to boredom effects, but more studies are necessary to confirm this. The comparison of item-by-item pass frequencies at test and retest indicated that the cognitive and perceptual suggestions in SHSS:C may yield lower responses at the second administration. There were five items that showed a more than 10-percentage-point decrease in pass frequency in the SHSS:C group, all of which were cognitive items. As a comparison, there were only two items with a 10-percentage-point decrease in the EHS group, one of which is a perceptual task (positive hallucination of the smell of a rose), the other a motor task (arm levitation). The EHS contains both cognitive and perceptual tasks just like the SHSS:C, and most of these tasks did not show a decreased pass frequency in the EHS from test to retest. Thus, the presence of such tasks alone is unlikely to provide sufficient explanation to why we see diminished responses in SHSS:C from test to retest. Nevertheless, the mosquito hallucination, taste hallucination, dream, anosmia, and posthypnotic amnesia suggestions in SHSS:C should be investigated further in test–retest studies to assess their consistency over time. It should be noted, however, that the actual observed mean decrease in hypnotizability scores from the first to the second administration was not substantially different between the two scales (0.41 mean decrease for EHS and 0.91 mean decrease for SHSS:C). A study with a larger sample size is necessary to directly contrast the two scales on the test–retest stability of the scores. It should also be kept in mind that even if there is a diminished responsiveness in suggestibility from first to second administration in either of these scales, it is likely a small effect and probably does not have a great clinical relevance.

Average scores of participants receiving the SHSS:C were approximately one unit higher than that of EHS participants, which is contrary to the hypothesis that the population mean would be equivalent for EHS and SHSS:C scores. It is unclear why average scores would differ between the two scales. A previous study has found high correlation for hypnotizability scores on the two scales and no asymmetry was reported in the difference scores between the two scales (Kekecs et al., 2016). However, in that study the two scales were administered by the same experimenter and within a single session. Thus, the similarity of the scores in that study might have been increased by expectancy and experimenter effects. The difference in score averages is mainly a result of the more frequent occurrence of low hypnotizability scores in the EHS group compared to the SHSS:C group. This may indicate that the SHSS:C has more test suggestion items that are easy to pass relative to the EHS.

At the end of each study visit, participants were asked to rate how pleasant their experiences were with the hypnotizability scales overall. Though we hypothesized that the EHS would be rated as more pleasant, the ratings for the two scales were statistically equivalent with an average answer close to the end marker of "very pleasant" for both. It was thought that the EHS would be rated more pleasant, because during the design of the EHS specific attention was taken to avoid items that may be unpleasant or stressful, such as the anosmia to ammonia, mosquito hallucination, and age regression, items in the SHSS:C, which could be upsetting for some individuals (Elkins, 2014). It is possible, however, that these items do not have a major impact on the overall pleasantness rating of SHSS:C itself or that the population of university students do not find these items so unpleasant. Even though overall pleasantness was found to be equivalent between the two scales, we did find that the pleasantness ratings of individual items were higher for the EHS in this sample. However, we did not test the statistical significance of this difference, since this analysis was not preregistered. The hallucinated voice, the anosmia to ammonia, and the mosquito hallucination items were rated the least pleasant in the SHSS:C, while the age regression item was rated among the more pleasant items on average. It is also possible that social desirability and a tendency to please the experimenter have biased the results, masking differences between the two scales.

Limitations

Since the EHS is used more frequently in the experiments in our laboratory, experimenters in our lab are more familiar with administering the EHS than the SHSS:C. This may have led to mistakes or deviations from the SHSS:C protocol that the experimenters were not aware of or did not report and could have contributed to the finding of poorer reliability for the SHSS:C. Closely monitoring protocol fidelity and assessing the effect of familiarity with each scale may be necessary in a replication attempt of the current study. Future studies should attempt to contrast test–retest reliability of the two scales directly to confirm that the EHS produces more consistent scores over time. Also, as described above, the random sequence for scale administration was broken and group allocation was determined by the availability of experimenters who were trained in administering either EHS or SHSS:C. This, together with the fact that some experimenters only administered the EHS and not the SHSS:C, may have resulted in experimenter effects or some effects of time and schedule. Although unlikely, if such effects were present, they may partially explain the difference in the distribution of the hypnotizability scores between the two scales. Thus, we encourage the use of randomized group allocation in similar future studies. There was also a slight deviation in the protocol in the EHS and the SHSS:C groups in that absorption was assessed intra-hypnosis as well in the EHS group. This is part of the standard procedure of doing the EHS administration at the laboratory where the data were collected and included in the EHS assessment forms in the lab, so this was left in the protocol of the EHS assessments by mistake. We find it unlikely that this single item question in hypnosis would have a serious effect on the stability of any of the scores or the other outcomes assessed in this study, but it would be preferable in a future study to leave this assessment out. Another thing to consider is that the analyzed sample might be slightly higher in hypnotizability than the population average, because there was some indication those who did not show up for the second session had lower hypnotizability in the first session than those who attended both sessions. This was not a statistically significant difference, but this may be because of the low number of dropouts that decreases the power of the test. Future studies could employ different retention strategies to further decrease attrition.

Further research should also assess factors involved in diminished responsiveness from the first to the second administration of the SHSS:C. For instance, changes in presession expectancies or perceptions of novelty versus boredom may be affecting some participants, especially as the delivery of the scale takes about an hour. Additionally, in the current study, all participants returned for the second visit 1 week after their first visit. Studies wherein the retest is administered closer or further apart in time from the original test would also shed light on factors involved in score differences over time. Future studies could also assess the consistency of subjective hypnotic experiences over time, for example, by looking at phenomenological scales such as the Phenomenology of Consciousness Inventory (Pekala, 1991) or by asking the participants their level of dissociation and automaticity.

Conclusions

Our results indicate that the EHS produces reliable results over repeated administrations, while the consistency of hypnotizability scores measured by the SHSS:C were less convincing. Specifically, the test–retest reliability of the SHSS:C was below expectations, and scores measured with SHSS:C were lower at the second administration of the scale. Further research is necessary to establish that the findings presented here are reproducible and to better understand the source of inconsistency in the SHSS:C scores. Nevertheless, if the findings turn out to be reproducible it would indicate that the EHS is better suited for scale validation purposes and for research studies assuming the consistency of hypnotizability over time. The hypnotizability categories were found to be unstable for both scales, thus, using these should be avoided in research, and the numerical scores should be preferred. We also found that the SHSS:C produced higher average scores compared to the EHS, indicating a difference in distribution between the scores produced by the two scales, possibly due to more easy-to-pass test suggestions in the SHSS:C. Our study suggests that the overall pleasantness of the EHS and the SHSS:C are comparable, at least in the university student population. Overall, our findings support the reliability and temporal stability of the EHS. Taken together with its other advantages discussed elsewhere (e.g., Kekecs et al., 2016), such as short administration time, this further underlines the usefulness of this scale in hypnosis research.

Author Contributions

Study idea (SI), Planning and design (includes ethics submission) (PD), Data collection (DC), Data entry (DE), Team coordination (CO), Data analysis (DA), Openness, registration, and data sharing (DS), Report write-up (RW), Funding and premises (FP).

ZK contributed in SI, PD, DC, CO, DA, and RW; RLR contributed in CO, DE, and RW; EES and TAV contributed in DC, CO, DE, and RW; MHY contributed in DC, CO, and RW; ER contributed in DC, DE, and RW; HN contributed in DC, and RW; RJ and PR contributed in DA and RW; ECs contributed in DS and RW; VV contributed in RW; GE contributed in SI, PD, CO, RW, and FP.

Disclosure Statement

No potential conflict of interest was reported by the authors.

ORCID

Ming Hwei Yek http://orcid.org/0000-0002-5587-9337

References

Acunzo, D., & Terhune, D. (2021). A critical review of standardized measures of hypnotizability. *International Journal of Clinical and Experimental Hypnosis*, *69*(1), 50–71.

Canty, A., & Ripley, B. (2017). *Boot: Bootstrap R (S-Plus) functions. R package version 1.3-18.2016*. [computer software].

Cardeña, E., Jönsson, P., Terhune, D. B., & Marcusson-Clavertz, D. (2013). The neurophenomenology of neutral hypnosis. *Cortex*, *49*(2), 375–385. https://doi.org/10.1016/j.cortex.2012.04.001

Cardeña, E., & Terhune, D. B. (2009). A note of caution on the Waterloo-Stanford Group Scale of Hypnotic Susceptibility: A brief communication. *International Journal of Clinical and Experimental Hypnosis*, *57*(2), 222–226. https://doi.org/10.1080/00207140802665484

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Sage Publications.

Chan, C., Chan, G. C., Leeper, T. J., & Becker, J. (2018). Rio: A Swiss-army knife for data file I/O (Version R package version 0.5.16). [computer software]

Elkins, G. R. (2014). *Hypnotic relaxation therapy: Principles and applications*. Springer.

Elkins, G. R., Barabasz, A. F., Council, J. R., & Spiegel, D. (2015). Advancing research and practice: The revised APA Division 30 definition of hypnosis. *International Journal of Clinical and Experimental Hypnosis*, *63*(1), 1–9. https://doi.org/10.1080/00207144.2014.961870

Fassler, O., Lynn, S., & Knox, J. (2008). Is hypnotic suggestibility a stable trait? *Consciousness and Cognition*, *17*(1), 240–253. https://doi.org/10.1016/j.concog.2007.05.004

Fox, J., & Weisburg, S. (2011). *Car, an {R} companion to applied regression* (3rd ed.). https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Gerlanc, D., & Kirby, K. (2013). *bootES: Bootstrap effect sizes. (Version 1.2)*. [computer software]. https://CRAN.R-project.org/package=bootES

Hilgard, E. R. (1965). *Hypnotic susceptibility*. Harcourt, Brace, and World.

Hilgard, E. R., Weitzenhoffer, A. M., Landes, J., & Moore, R. K. (1961). The distribution of susceptibility to hypnosis in a student population: A study using the Stanford Hypnotic Susceptibility Scale. *Psychological Monographs: General and Applied*, *75*(8), 1–22.https://doi.org/http://dx.doi.10.1037/h0093802

Kekecs, Z., Bowers, J., Johnson, A., Kendrick, C., & Elkins, G. (2016). The Elkins Hypnotizability Scale: Assessment of reliability and validity. *International Journal of Clinical and Experimental Hypnosis*, *64*(3), 285–304. https://doi.org/10.1080/00207144.2016.1171089

Kelley, K. (2005). The effects of nonnormal distributions on confidence intervals around the standardized mean difference: Bootstrap and parametric confidence intervals. *Educational and Psychological Measurement*, *65*(1), 51–69. https://doi.org/10.1177/0013164404264850

Kihlstrom, J. F. (1985). Hypnosis. *Annual Review of Psychology*, *36*(1), 385–418. https://doi.org/10.1146/annurev.ps.36.020185.002125

MacHovec, F. J. (1986). *Hypnosis complications: Prevention and risk management*. Charles C. Thomas.

Montgomery, G. H., Schnur, J. B., & David, D. (2011). The impact of hypnotic suggestibility in clinical care settings. *International Journal of Clinical and Experimental Hypnosis*, *59*(3), 294–309. https://doi.org/10.1080/00207144.2011.570656

Pekala, R. J. (1991). The Phenomenology of Consciousness Inventory. In R. J. Pekala (Ed.), *Quantifying consciousness* (pp. 127–143). Springer.

Piccione, C., Hilgard, E. R., & Zimbardo, P. G. (1989). On the degree of stability of measured hypnotizability over a 25-year period. *Journal of Personality and Social Psychology*, *56*(2), 289–295. https://doi.org/10.1037/0022-3514.56.2.289

Revelle, W. (2017). *Psych: Procedures for personality and psychological research*, Northwestern University. https://CRAN.R-project.org/package=psych Version = 1.8.10

Spanos, N. P., Radtke, H. L., Hodgins, D. C., Bertrand, L. D., Stam, H. J., & Dubreuil, D. L. (1983). The Carleton University Responsiveness to Suggestion Scale: Stability, reliability, and relationships with expectancy and "hypnotic experiences". *Psychological Reports*, *53*(2), 555–563. https://doi.org/10.2466/pr0.1983.53.2.555

Spiegel, H., Aronson, M., Fleiss, J. L., & Haber, J. (1976). Psychometric analysis of the Hypnotic Induction Profile. *International Journal of Clinical and Experimental Hypnosis*, *24*(3–4), 300–315. https://doi.org/10.1080/00207147608416210

Taslimbakhsh, Z., Sadeghi, K., Sadeghi, K., & Ahmadi, S. M. (2017). Investigating factor structure, validity, and reliability of the persian version of the Stanford Hypnotic Susceptibility Scale: Form C (SHSS: C). *Iranian Journal of Psychiatry and Clinical Psychology*, *23*(3), 362–379. https://doi.org/10.29252/nirp.ijpcp.23.3.362

Team, R. C. (2019). *R: A language and environment for statistical computing (Version 3.3.1)*. [computer software]. http://www.R-project.org/

Walker, W. L. (2016). Guidelines for the use of hypnosis: When to use hypnosis and when not to use. *Australian Journal of Clinical and Experimental Hypnosis*, *41*(1), 41–53. https://www.hypnosisaustralia.org.au/wp-content/uploads/10330_AJCEH_2016_FINAL.pdf#page=45.
Weitzenhoffer, A. M., & Hilgard, E. R. (1962). *Stanford Hypnotic Susceptibility Scale, Form C*. Consulting Psychologists Press.
Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis (Version 3.2.0)*. Springer.
Wickham, H., François, R., Henry, L., & Müller, K. (2018). *dplyr: A grammar of data manipulation (version 0.7.8)*. [computer software]. https://CRAN.R-project.org/package=dplyr
Woody, E. Z., & Barnier, A. J. (2008). Hypnosis scales for the twenty-first century: What do we need and how should we use them? In M. R. Nash & A. J. Barnier (Eds.), *The Oxford handbook of hypnosis: Theory, research, and practice* (pp. 255–282). Oxford University Press.
Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, *2*(3), 7–10. https://CRAN.R-project.org/doc/Rnews/

Test–Retest Reliabilität der Stanford Hypnotic Susceptibility Scale, Form C und der Elkins Hypnotizability Scale

ZOLTAN KEKECS, R. LYNAE ROBERTS, HYEJI NA, MING HWEI YEK, ELIZABETH E. SLONENA, EZRHIEL RACELIS, TAMARA A VOOR, ROBERT JOHANSSON, PIETRO RIZZO, ENDRE CSIKOS, VANDA VIZKIEVICZ, UND GARY ELKINS

Zusammenfassung: Dieses Vorhaben zielte darauf ab, die Konsistenz der Hypnotisierbarkeit über wiederholte Untersuchungen zu messen und zwar mittels der Stanford Hypnotizability Scale: Form C (SHSS:C) und der Elkins Hypnotizability Scale (EHS) sowie die Testwerte-Verteilung und die Annehmlichkeit dieser Skalen gegeneinander abzuwägen. Universitätsstudenten wurden entweder mit der SHSS:C oder der EHS getestet und zwar zweimal mit einer Woche Abstand von zwei unterschiedlichen Untersuchenden. Die Test–Retest Reliabilität der EHS und der SHSS:C lag bei r = .82(.71-.92) bzw. r = .66.95(.47-.86) (Spearman's Korrelation). Die Hypnotisierbarkeit war in der ESH-Gruppe bei Test und Retest vergleichbar, die SHSS:C Testwerte nahmen im Retest ab. Wir fanden heraus, dass die SHSS:S höhere Werte ergab als die EHS und dass beide Skalen in ihrer Annehmlichkeit vergleichbar waren. Insgesamt unterstützten unsere Ergebnisse die Reliabilität der EHS, während die SHSS:C Testwerte zwischen zwei Testdurchgängen inkonsistenter waren. Mehr Forschung ist nötig.

ALIDA IOST-PETER, DIPL.-PSYCH.

Fiabilité test–retest de l'échelle de sensibilité hypnotique de Stanford, de la forme C et de l'échelle d'hypnotisabilité Elkins

ZOLTAN KEKECS, R. LYNAE ROBERTS, HYEJI NA, MING HWEI YEK, ELIZABETH E. SLONENA, EZRHIEL RACELIS, TAMARA A VOOR, ROBERT JOHANSSON, PIETRO RIZZO, ENDRE CSIKOS, VANDA VIZKIEVICZ, ET GARY ELKINS

Résumé: Ce projet visait à évaluer la cohérence de l'hypnotisabilité sur des évaluations répétées mesurées par l'échelle de sensibilité hypnotique de Stanford: forme C (SHSS:C) et l'échelle d'hypnotisabilité Elkins (EHS) et à comparer la distribution des scores et l'agrément de ces échelles. Les étudiants universitaires ont reçu deux fois le SHSS:C ou l'EHS avec un délai d'une semaine par des expérimentateurs distincts. Fiabilité test–retest de l'EHS et du SHSS:C était respectivement r_s = 0,82 (0,71 -,92) et r_s = 0,66, 95% (0,47 -86) (corrélation de Spearman). L'hypnotisabilité était comparable au test et au nouveau test dans le groupe EHS, les scores SHSS:C ont diminué par le nouveau test. Nous avons constaté que le SHSS:C produisait des scores plus élevés que l'EHS, et l'agrément des 2 échelles était comparable. Dans l'ensemble, nos résultats

ont confirmé la fiabilité de l'EHS, tandis que les scores SHSS:C étaient plus incohérents entre les 2 évaluations. Plus de recherche est justifiée.

GERARD FITOUSSI, M.D.
Président of the European Society of Hypnosis

Fiabilidad prueba–posprueba de la Escala Stanford de Susceptibilidad Hipnótica, Forma C y la Escala Elkins de Hipnotizabilidad

ZOLTAN KEKECS, R. LYNAE ROBERTS, HYEJI NA, MING HWEI YEK, ELIZABETH E. SLONENA, EZRHIEL RACELIS, TAMARA A VOOR, ROBERT JOHANSSON, PIETRO RIZZO, ENDRE CSIKOS, VANDA VIZKIEVICZ, Y GARY ELKINS

Resumen: Este proyecto se enfocó en evaluar la consistencia de la hipnotizabilidad medida repetidamente mediante la Escala Stanford de Susceptibilidad Hipnótica, Forma C (SHSS:C) y la Escala Elkins de Hipnotizabilidad (EHS) y a contrastar la distribución de puntuaciones y qué tan placenteras son estas escalas. Se administró la SHSS:C o la EHS a estudiantes universitarios en dos ocasiones con una semana de diferencia por distintos evaluadores. La fiabilidad prueba–posprueba de la EHS y la SHSS:C fue de r_s = .82 (.71 - .92) y r_s = .66 (.47 - .86) respectivamente (correlación de Spearman). La hipnotizabilidad resultó comparable entre la primera y segunda aplicación para el grupo de la EHS, mientras que las puntuaciones de la SHSS:C decrecieron en la segunda aplicación. Encontramos que la SHSS:C produce puntaciones más elevadas que la EHS y que ambas son igualmente placenteras. En general, los resultados sustentan la fiabilidad de la EHS, mientras que las puntuaciones de la SHSS:C resultaron más inconsistentes entre ambas aplicaciones. Se requiere más investigación.

OMAR SÁNCHEZ-ARMÁSS CAPPELLO
Autonomous University of San Luis Potosi, Mexico