ORIGINAL ARTICLES

Classical Test Theory and Item Response Theory for Harvard Group Scale of Hypnotic Susceptibility with African American College Students

Ling-Lun (Crystal) Chien, Ph.D., Marty Sapp, Ed.D.

Hypnotizability is not an all-or-nothing construct; rather, it is an individual's reactions to being hypnotized, and this may fall on a continuum (1). In addition, even thought hypnotizability is relatively stable, it is also influenced by the client's rapport with the hypnotherapist, the client's expectations and beliefs, and other social or psychological factors (2). The purpose of this project focused on assessing hypnotizability by using the HGSH:A with African American college students. By analyzing Harvard Group Scale of Hypnotic Susceptibility, Form A (HGSHS:A) by both Classical Test Theory and Item Response Theory, results from both measurement models were consistent. Both measurement models indicated that items 2 (eye closure) and 7 (moving hand together) were good items, and item 9 (experiencing of a fly) and 11 (post-hypnotic suggestion) were bad items for this African American College student sample. Based on the eleven variables of HGSHS:A, through principal components analysis, there were three factors: direct suggestion, motor challenge, and perceptual cognitive. However, the reliability of this study was low which may due to African American college students appear to have expectancies for hypnotic responding to was to occur without using techniques to assist responding such as imagining and thinking along with suggestions. **(Sleep and Hypnosis 2012;14(1-2):13-19)**

Key words: Hypnotizability, African American college students, cultural differences

INTRODUCTION

Hypnotic susceptibility is a measured by how easily a person can be hypnotized (2). There are several types of scales used; among all of them, the Harvard Group Scale of Hypnotic Susceptibility (HGSHS) (3,4,5) and the Stanford Hypnotic Susceptibility Scales (SHSS) (4,5) are the most common (6). Both of the Standard scales and Harvard scales are connected to the traditional perspective of hypnosis. The biggest difference between SHSS and HGSHS is the way they are administered, SHSS is administered to individuals while HGSHS is administered predominantly to large groups of people (6). Also, because the SHSS is administered individually, it is more accurate than group measures (6).

The Stanford Scale of Hypnotic Susceptibility (SHSS) Forms A, B, and C (4,5) are the benchmarks for other susceptibility scales (6). Among these three forms, the Stanford Scale of Hypnotic Susceptibility Form C (SHSS:C) has

Address reprint requests to: Dr. Marty Sapp,
Department of Educational Psychology, Milwaukee, Wisconsin, USA.
E–mail: Sapp@UWM.EDU

Accepted June 3, 2012

better sampling of fantasy and cognitive distortion items with reliability of 0.85 (2). Items on this scale include the following: hand lowering, moving hands apart, mosquito hallucination, taste hallucination, arm rigidity, dream, age regression, arm immobilization, anosmia to ammonia, hallucinated voice, negative visual hallucination, and posthypnotic amnesia (2). The SHSS are scored on an objective criteria ranging between 0 and 12. If a client did not respond to any items, he or she would receive a score of 0, and if a client responded to all items, he or she would obtain a score of 12 (7,8).

Around 1950s and early 1960s, close to the same time that the Stanford Scales were being developed, the Harvard Group Scale of Hypnotic Susceptibility (HGSHS) Form A was developed (3). The Harvard Group Scale of Hypnotic Susceptibility, Form A (HGSHS:A) is standard procedure for estimating group susceptibility to hypnosis (6). Sapp (2) indicated that an individual's susceptibility to hypnosis may change within different circumstances; in addition, an individual who appears relatively unsusceptible at this time might appear differently under different circumstances. The advantage of this scale is that it can be administered to groups via tape recording or by directly reading the instructions (2). The HGSHS:A has 12–item scale with 9 items adapted from the SHSS. Scores on the HGSHS:A range from 0–12, and it has a reliability measure of .83. The HGSHS:A requires about one hour to administer.

Items of HGSHS:A usually consist of motor tasks and cognitive tasks with the motor tasks being easier to complete. The 12 items on HGSHS:A are head falling, eye closure, hand lowering, arm immobilization, finger lock, arm rigidity, moving hands together, communication inhibition, experiencing of a fly, eye catalepsy, post–hypnotic suggestion, and amnesia.

The Harvard Group Scale of Hypnotic Susceptibility, Form A (HGSHS:A) developed by Shor and Orne (8) is probably the most widely used group test for initial screening of hypnotic susceptibility (2). Although the limitations of the instrument as an estimate of the full range hypnotic susceptibility has been criticized (9,10), it is a good tool for preliminary screening of participants for research purposes. The purpose of this project focused on assessing hypnotizability by using the HGSH:A with African American college students. In addition, Item Response Theory was used to test the quality of HGSH:A items with African American college students.

This study examined the psychometric properties of HGSHS:A. Classical Test Theory (CTT) and Item Response Theory (IRT) were applied to study this scale's reliability, and quality of test items. Classical Test Theory (CTT) is one of the most significant issues with the concept of correlation (11). The classic true score model is one of the most significant issues from British psychologist Charles Spearman's fascination with the concept of correlation (11). True score theory maintains that every measurement is an additive composite of two components: true score of the respondent on that measure and random error (11). In other words, the classical true score model is represented as $X = T + E$, where X represents the observed score, T represents the individual's true score, and E represents a random error component (11). This equation means that the observed score (raw score) equals the true score plus the error score. Even though we do not observe the true score and the error score, we assume that there are two components. Classical Test Theory (CTT) was assessed because it is a simple and powerful model for measurement; it reminds us that most measurement has an error component. Moreover, Classical Test Theory (CTT) is the foundation of reliability theory (11).

CTT is a group approach to assessment, since it compares a person's score to a standardization sample. This group assessment model is based on the number of individuals who succeeded on items and item difficulty, and these two components are dependent on the appropriateness of the standardization

sample used. The CTT item analysis is based on three assumptions. First, true and error scores are unrelated. Second, the mean of error scores in the population of examinees equal zero. Third, error scores on parallel tests are uncorrelated.

For Item Response Theory (IRT), an individual's response to a specific test item is determined by an unobserved mental attribute of the individual (11). This function specifies that as the level of the trait increases, the probability of a correct response to an item increases (11). IRT focuses on item level and is able to tell you how a person will react to a particular item.

IRT assumes unidimensionality of test items. If this assumption holds, examinees' ability levels can be obtained that are independent of the instrument used. Actually, IRT is a measurement model and item analyses or scaling model that is not based upon group norms. Finally, IRT states that when an examinee encounters a test item, his or her outcome is determined by the product of his or her ability and the item difficulty.

A criticism of Classical Test Theory, when examining scale issues, is that the scale will not distinguish if an item is much harder or easier (11). Item Response Theory (IRT) will give you a scale that will be able to discriminate between harder or easier items. Furthermore, you can take item difficulty into account when constructing the scale (11).

METHOD

Participants

Participants recruited for this study were self–identified African Americans from a four–year college that was predominately African American. There were 103 participants in this study with 61 being female, and 42 being male. The mean age was 19.44 years old, and the standard deviation was 2.82 years. Participants ranged in age from 17 to 41 years old. These participants were part of the study conducted by Sapp (6).

Procedure

Participants completed the experimental procedures in groups, and they received the tape–recorded HGSHS:A. The session began with the presentation of preliminary information, after which the subjects were asked to fill in the first page of the response booklet. After the entire procedure was complete, these participants had the opportunity to discuss their experiences with the researcher. Finally, participants received extra credit for their participants.

Measure

As previously stated, the Harvard Group Scale of Hypnotic Susceptibility, Form A (HGSHS:A) is one of the most widely used hypnosis measurements in estimating ones susceptibility to hypnosis (2). The advantage of this scale is that it can be administered to groups via tape recording or by directly reading the instructions (2). Again, items of HGSHS:A has two part: motor tasks and cognitive. The 12 items on HGSHS:A are head falling, eye closure, hand lowering, arm immobilization, finger lock, arm rigidity, moving hand together, communication inhibition, experiencing of a fly, eye catalepsy, post–hypnotic suggestion amnesia.

Scoring

The scoring of the response booklets followed the procedure described by Shor and Orne (8). It is a behavioral method in which the subjects, after receiving a hypnotic induction, evaluate their overt responses with a self–rate scale. Participants received a score of 1 if they had marked item A (indicating an experienced behavioral change for a given suggestion) and a 0 if item B was marked (indicating that the behavioral change was not experienced). The average score is 0.5 out of 12. Amnesia was

Table 1. Mean of Item Difficulty of HGSHS: A Items

Item	1	2	3	4	5	6	7	8	9	10	11
Mean of Item Difficulty(CTT)	0.485	0.466	0.563	0.291	0.490	0.379	0.456	0.340	0.757	0.456	0.835
Mean of Item Difficulty (IRT)	-.183	-.100	.787	-3.59	-.151	-1.55	-.221	-1.88	67.04	-.230	53.96

scored as 1 if the subject recalled fewer than four out of twelve items before the amnesia was filled, however, in this study, these researchers did not score the amnesia sessions. Therefore, there were only 11 items analyzed for this project.

RESULTS

Reliability

The reliability for items of the HGSHS–A was low (Cronbach's Alpha= .181), which may be due to the specific target population in this study. Since African American college students tend to anticipate that hypnotic experiences will happen automatically (2), they may respond negatively even though they have been hypnotized. Also, this is the same finding Sapp (6) found.

Classical Test Theory (CTT) and Item Response Theory (IRT) analysis

Due to the scoring systems, 0 means non–hypnotized while 1 means hypnotized. The larger means suggest more hypnotizability. From CTT's mean of each item difficulty indicate that the cognitive items such as experiencing of fly and post–hypnotic suggestion (item 9, and 11) were the easy items. In contrast, motor items, arm immobilization, arm rigidity, and communication inhibition (item 4, 6, and 8) were harder items.

The item difficulty histogram of IRT shows that most of the items fall in the negative area. It indicates that that most of the items are easier to measure the hypnosis ability. Among those 11 items, item 9&11 are positive and have the large value. It shows that those two items are hard items for these students to feel that they are hypnotized.

Principal Component Analysis

The histogram of HGSHS: A for total scores showed that the scores were positively skewed. This indicated that most of these African American college students were not hypnotized. The mean score of the research was 16.62, in the histogram also showed that most of students fell within this area. The standard deviation of this data is 2.07. From the mean and standard deviation, we can state that even though not all of these scores fall into a normal distribution; they are within the one standard deviation.

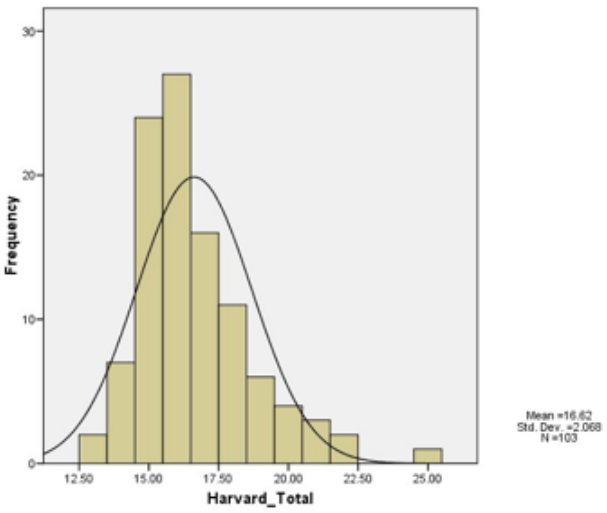

Figure 1. Histogram of HGSHS:A Total Score

Based on the scoring systems, the larger means suggest less hypnotizability. The mean of each item difficulty indicated that the cognitive items such as experiencing of fly and post–hypnotic suggestion (item 9 and 11) were the most difficult items. The correlation matrix contains the Pearson correlation coefficient between all pairs of variables. This correlation matrix was used to check for the patterns of

Table 2. Correlation Matrix of HGSHS: A Items

Correlation Matrix

		Head Falling	Eye Closure	Hand Lowering (left hand)	Arm Immobolilization (right arm)	Finger Lock	Arm Rigidity (left)	Moving Hands Together	Communication Inhibition
Correlation	Head Falling	1.000	.105	.190	-.024	-.095	-.117	.046	.000
	Eye Closure	.105	1.000	.156	.087	-.092	.073	.277	.069
	Hand Lowering (left hand)	.190	.156	1.000	-.168	-.160	.002	.060	-.153
	Arm Immobolilization (right arm)	-.024	.087	-.168	1.000	-.070	.028	-.072	.307
	Finger Lock	-.095	-.092	-.160	-.070	1.000	.156	-.031	-.031
	Arm Rigidity (left)	-.117	.073	.002	.028	.156	1.000	.089	.032
	Moving Hands Together	.046	.277	.060	-.072	-.031	.089	1.000	.084
	Communication Inhibition	.000	.069	-.153	.307	-.031	.032	.084	1.000
	Experiencing of Fly	.051	-.016	.141	-.185	.135	-.025	.155	-.120
	Eye Catalepsy	.085	.238	.021	.099	-.014	.129	.100	.125
	Post-Hypnotic Suggestion (touching left ankle)	.118	-.004	.241	-.175	.002	-.030	.092	-.233

these relationships. However, from the correlation matrix, these correlations between most of these variables are low.

The KMO statistic for the sample data set is 0.609. KMO is an overall index telling us that these data are likely to factor analyze well based on correlation and partial correlation. Based on this KMO value of 0.609, these data supported the use of factor analysis and suggested that the data may be grouped into a smaller set of underlying factors. The Bartlett's measure tests the null hypothesis that the original correlation matrix is an identity matrix (11). The Bartlett test shows that non–zero correlations and exists at the 0.015 significance level. Therefore, it is appropriate to continue the analysis.

Table 3. KMO and Bartlett's Test of HGSHS:A

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.609
Bartlett's Test of Sphericity	Approx. Chi-Square	80.004
	df	55
	Sig.	.015

There were 11 linear components within these data. These eigenvalues associated with each factor represent the variance explained by that particular linear component and the percentage of variance explained (11. Within this analysis, there were four components that had the eigenvalues larger than 1. However, due to the forth component only slightly larger than 1 (Eigenvalues= 1.020). The scree plot showed the elbow that indicated the point of inflexion on the curve (11). There are three factors above the elbow, therefore, we justify retain three factors. Based on the scree plot, we pick three factors above the elbow for this measurement. Three factors were picked for this measurement model.

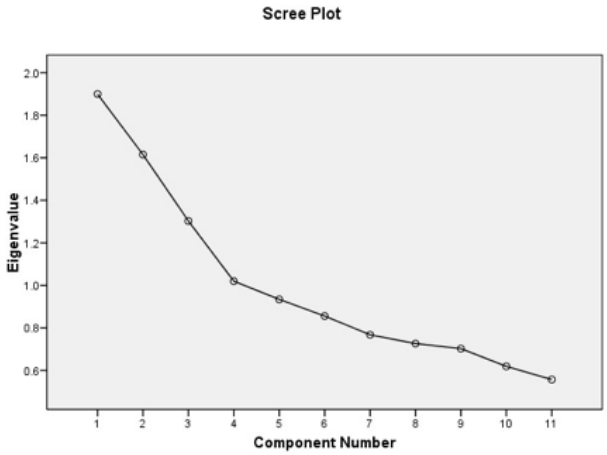

Figure 2. Scree Plot of HGSHS:A

These eigenvalues and the percentage of variance associated with these factors are again displayed in the columns labeled Extraction Sums of Squared Loadings. The value in this part of the table is the same as the values before extraction, except that the values for the discarded factors, which are those factors 4th to 11th, are ignored. The final part of the table, rotation sums of squared loadings, indicated the

Table 4. Total Variance Explained of HGSHS:A

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.900	17.271	17.271	1.900	17.271	17.271	1.869	16.986	16.986
2	1.615	14.683	31.954	1.615	14.683	31.954	1.562	14.201	31.188
3	1.303	11.845	43.799	1.303	11.845	43.799	1.387	12.611	43.799
4	1.020	9.270	53.069						
5	.934	8.491	61.560						
6	.856	7.780	69.339						
7	.768	6.980	76.319						
8	.726	6.602	82.921						
9	.703	6.387	89.308						
10	.619	5.627	94.935						
11	.557	5.065	100.000						

Extraction Method: Principal Component Analysis.

eigen values of the factors after rotation are displayed. Rotation has the effect of optimizing the factor structure and one consequence for these hypnosis data is that the relative important of the three factors is equalized. For example, factor 1, before rotation, factor 1 accounted for more variance than the remaining two (17.27% compared to 14.68 and 11.85). After rotation factor 1 accounted for only 16.99% of variance (compared to 14.20 and 12.61). Moreover, these three factors together explained 43.79% of the variance.

The table shows the communalities before and after extraction. Principal component analysis works on the initial assumption that all variance is common. Therefore, before extraction the communalities are all 1. The communalities in the column labeled extraction reflect the common variance in the data structure. We can say that 35.3% of the variance associated with the variable, Head Falling, is shared variance.

Table 5. Communalities of HGSHS:A

Communalities		
	Initial	Extraction
Head Falling	1.000	.353
Eye Closure	1.000	.542
Hand Lowering (left hand)	1.000	.462
Arm Immobolilization (right arm)	1.000	.400
Finger Lock	1.000	.520
Arm Rigidity (left)	1.000	.447
Moving Hands Together	1.000	.446
Communication Inhibition	1.000	.390
Experiencing of Fly	1.000	.425
Eye Catalepsy	1.000	.381
Post-Hypnotic Suggestion (touching left ankle)	1.000	.452

Extraction Method: Principal Component Analysis

Another way to look at these communities is in terms of the proportion of variance explained by the underlying factors. After extraction some of the factors are discarded and so some information is lost. The amount of variance in each variable that can be explained by the retained factor is represented by the communalities after extraction.

DISCUSSION

These results from IRT measurement were consistent with those from CTT. In comparison to these item difficulties and item discrimination indices of CTT and IRT, both measurement models indicated that item 2 (eye closure) and 7 (moving hand together) were good items, and item 9 (experiencing of a fly) and 11 (post–hypnotic suggestion) were bad items for African American College student population.

Based on the eleven variables of HGSHS: A, a principal components analysis found there were three factors underlying the HGSHS:A : direct suggestion: participants are told they may have a particular experience (e.g., hand lowering, hands together), motor challenge: participants are told they may not be able to do something in particular, but to try to do it anyway (e.g., arm immobilization, arm rigidity), and perceptual cognitive: participants are told that they may experience something (see/hear/remember) that is different than what is actually there (e.g., fly hallucination).

As Sapp (6) noted, the standard or behavioral

scoring of the HGSHS: A did not produce reliable items with African American college students. Even though the Harvard Group Scale of Hypnotic Susceptibility, Form A (HGSHS:A) developed by Shor and Orne (8) is the most widely used group test for initial screening of hypnotic susceptibility, these items of this scale were not reliable with African American college students. The low reliability may due to African American college students' expectancies. They appear to have expectancies for hypnotic responding to occur by itself (2). Since the African American college students expect that hypnotic experiences will happen automatically, this is a cultural dimension for these students. Again, these results were consistent with Sapp (6), who found that the inner subjective experiences method for scoring the HGSHS:A produced more reliable results than did the behavioral scoring method.

Kallio and Ihamuotila (12) stated that the normative studies indicate that the properties of HGSHS: A are generally comparable over different cultural and linguistic context, but this study showed the opposite. Based on all these results, the standard or behavioral scoring system of the HGSHS:A did not produce reliable items with African American college students. Additional research is needed to find out why African American college students scored lowered on this standard measure. Finally, additional research is needed that includes larger sample sizes and more diverse educational levels.

REFERENCES

1. *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, 289-295.*

2. *Sapp, M. (1997). Counseling and psychotherapy: Theory, associated research, and issues. Lanham, MD: University Press of America, 1997.*

3. *Kihlstrom, J. F. (1985). Hypnosis. Annual Review in Psychology, 36, 385-418.*

4. *Weitzenhoffer, A. M., & Hilgard, E. R. (1962). Stanford Hypnotic Susceptibility Scale, Form C. Palo Alto, CA: Consulting Psychologists Press.*

5. *Weitzenhoffer, A. M., & Hilgard, E. R. (1967). Revised Stanford Profile Scales of Hypnotic Susceptibility: Forms I and II. Palo Alto, CA: Consulting Psychologists Press.*

6. *Sapp, M. (2004). Hypnotizability: Harvard and Stanford Scales with African American College Students. Sleep and Hypnosis, 6, 14-18.*

7. *Sapp, M. (2000). Hypnosis, dissociation, and description: Theories, assessment, and treatment. Springfield, IL. Charles C. Thomas Publishers.*

8. *Shor, R. E., & Orne, E. C. (1962). Harvard Group Scale of Hypnotic Susceptibility, Form A. Palo Alto, CA: Consulting Psychologists Press.*

9. *Kirsch, I. (1997). Suggestibility or hypnosis: What do our scales really measure? International Journal of Clinical and Experimental Hypnosis, 45, 212-225.*

10. *Woody, E. Z. (1997). Have the Hypnotic Susceptibility Scales Outlived their Usefulness? International Journal of Clinical and Experimental Hypnosis, 45, 226-238.*

11. *Crocker, L., & Algina, J. (2006). Introduction to Classical and Modern Test Theory. Fort Worth, TX: Harcourt.*

12. *Kallio, S. P. I., & Ihamuotila, M. J. (1999). Finish Norms for the Harvard Group Scale of Hypnotic Susceptibility, Form A. International Journal of Clinical and Experimental.*