

1-1-2018

Assessing Instructed Feigner's Response to the Gudjonsson Suggestibility Scale

Stephen Scott Grabner
stephensgrabner@gmail.com

This document is a product of extensive research conducted at the Nova Southeastern University [College of Psychology](#). For more information on research and degree programs at the NSU College of Psychology, please [click here](#).

Follow this and additional works at: https://nsuworks.nova.edu/cps_stuetd

 Part of the [Psychology Commons](#)

All rights reserved. This publication is intended for use solely by faculty, students, and staff of Nova Southeastern University. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, now known or later developed, including but not limited to photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the author or the publisher.

NSUWorks Citation

Grabner, S. S. (2018). Assessing Instructed Feigner's Response to the Gudjonsson Suggestibility Scale. . Available at: https://nsuworks.nova.edu/cps_stuetd/121

This Dissertation is brought to you by the College of Psychology at NSUWorks. It has been accepted for inclusion in College of Psychology Theses and Dissertations by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.

**ASSESSING INSTRUCTED FEIGNER'S RESONSE TO THE GUDJONSSON
SUGGESTIBILITY SCALE**

by

Stephen S Grabner II

A Dissertation Presented to the College of Psychology
of Nova Southeastern University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

NOVA SOUTHEASTERN UNIVERSITY

2018

DISSERTATION APPROVAL SHEET

This dissertation was submitted by Stephen S. Grabner II under the supervision of the Chairperson of the dissertation committee listed below. It was submitted to the College of Psychology and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Clinical Psychology at Nova Southeastern University.

Approved:

6/14/18
Date of Defense

David L Shapiro Ph.D.
David Shapiro, Ph.D., ABPP

Lenore E Walker
Lenore E. Walker, Ed.D., ABPP

C. DeLucia
Christian DeLucia, Ph.D

8/14/18
Date of Final Approval

David L Shapiro Ph.D.
David Shapiro, Ph.D., ABPP

Statement of Original Work

I declare the following:

I have read the Code of Student Conduct and Academic Responsibility as described in the *Student Handbook* of Nova Southeastern University. This dissertation represents my original work, except where I have acknowledged the ideas, words, or material of other authors.

Where another author's ideas have been presented in this dissertation, I have acknowledged the author's ideas by citing them in the required style.

Where another author's words have been presented in this dissertation, I have acknowledged the author's words by using appropriate quotation devices and citations in the required style.

I have obtained permission from the author or publisher—in accordance with the required guidelines—to include any copyrighted material (e.g., tables, figures, survey instruments, large portions of text) in this dissertation manuscript.

Stephen Grabner _____

Name

14 AUG 2018

Date

ACKNOWLEDGMENTS

I am deeply indebted to the many wonderful individuals who have supported me in completing this research endeavor. I would first like to thank Dr. David Shapiro, Dr. Lenore Walker, and Dr. Christian DeLucia for their assistance as both my dissertation committee and my mentors. Their patience and flexibility allowed me to engage in this pursuit of knowledge in ways I did not fathom when I began the program five short years ago. In addition, this research would not have been possible without the support of my research assistants, namely Brittany, Lauren, and Rose. They went above and beyond over the course of two years to ensure the project ran as smoothly as possible, and I am grateful for their contributions.

I would be remiss if I did not thank my parents, Amy and Scott, as well as my brother, Nick, for their unconditional support. My family is the reason I entered the field of Psychology, and they instilled in me the values that guide me today. They continue to encourage my professional and personal growth, and I hope that I have made them proud.

Lastly and most importantly, I want to give thanks to my wife, Elizabeth, who inspires and motivates me every day. May this be but one accomplishment of many on our lifelong voyage together.

Table of Contents

LIST OF TABLES	ix
APPENDICIES	xi
ABSTRACT	xii
CHAPTER I.....	1
Statement of the Problem	1
CHAPTER II.....	4
Literature Review.....	4
False Confessions.....	5
False Confession Typology	8
Protecting and Defending Against False Confessions	10
Suggestibility, Compliance, and Acquiescence	14
Suggestibility.	15
Interrogative Suggestibility.	16
Suggestibility and Prevention of False Convictions	19
Rules of evidence and false confession testimony admissibility.....	20
The Gudjonsson Suggestibility Scale	22
Creation and Development.....	22
Administering the Gudjonsson Suggestibility Scale.....	23
Scoring the Gudjonsson Suggestibility Scale.....	25
Interpreting Gudjonsson Suggestibility Scores.....	26
Use of the Gudjonsson Suggestibility Scale	28
Manipulating Suggestibility Scores	30
Validity Indicators.....	32
Gudjonsson Suggestibility Scale and Validity Indicators.	33
Empirical Research Regarding Feigning Suggestibility	34
Methods of Feigning Research.....	35
Trends in GSS Scores.....	37
Yield and Shift score relation.	38
Total Suggestibility score trends	40
Trends regarding Free Recall scores	40

Comparison to Vulnerable Populations	42
Impact of Baseline Suggestibility	44
Summary of Findings and Literature Critique	45
Administration Discrepancies	45
Justification for shortening the administration length.	47
Manipulating Cognitive Load	49
Additional Administration Differences	51
Undergraduate Sample Comparisons	52
Intelligence	53
Additional Assessment of Valid Responding Pattern	55
Purpose of the Study	57
Research Hypotheses	58
Hypothesis 1: Gudjonsson Suggestibility Scale Score Differences	59
Hypothesis 1.1.	59
Hypothesis 1.2.	59
Hypothesis 1.3.	59
Hypothesis 1.4.	59
Hypothesis 1.5.	59
Hypothesis 1.6.	59
Hypothesis 2: Personality Assessment Inventory Validity Scale Differences	60
Hypothesis 2.1.	60
Hypothesis 2.2.	60
Hypothesis 2.3.	60
Hypothesis 2.4.	60
Hypothesis 3: Test of Memory Malingering Scale Differences.....	60
Hypothesis 3.1.	60
Hypothesis 3.2.	61
Hypothesis 3.3.	61
Hypothesis 4: Gudjonsson Compliance Scale Score Differences.....	61
Exploratory Analysis: Gudjonsson Suggestibility Scale Mean Comparisons	61
CHAPTER III	63
Methodology.....	63

Participants	63
Protection of Human Participants	63
Measures.....	64
Gudjonsson Suggestibility Scale-1 (GSS-1).....	64
Test of Memory Malingering.	66
Personality Assessment Inventory (PAI).....	67
Gudjonsson Compliance Scale.	68
Demographic Questionnaire.	69
Procedure.....	69
Statistical Analysis	72
CHAPTER IV	74
Results.....	74
Hypothesis 1	74
Hypothesis 1.1.	76
Hypothesis 1.2.	76
Hypothesis 1.3.	76
Hypothesis 1.4.	77
Hypothesis 1.5	77
Hypothesis 1.6.	77
Hypothesis 2.....	77
Hypothesis 2.1.	79
Hypothesis 2.2	80
Hypothesis 2.3	80
Hypothesis 2.4	81
Hypothesis 3.....	81
Hypothesis 4.....	81
Exploratory Analyses	82
Comparison to previous literature.	82
Comparison to GSS-1 Normative sample.	83
Comparison to intellectually disabled sample.	84
Comparison to Frumkin, Lally, and Sexton (2012) US Forensic norms.	85
CHAPTER V	87

Discussion	87
Interpretation of Hypothesis 1: Gudjonsson Suggestibility Scale Findings.....	87
Free Recall Scores.	88
GSS-1 Primary Scales.	89
Comparison to Previous Literature	90
Proposed Interpretations.	91
Hypotheses 2-4: Personality Assessment Inventory, Gudjonsson Compliance Scale, and Test of Memory Malingering Findings	93
Integration of Findings and the Statement of the Problem	96
Limitations of the Study.....	98
Implication for Future Research	100
Clinical Implications	101
Conclusions.....	103
References.....	104
APPENDICES	113

LIST OF TABLES

Table 1: Descriptive Statistics for Adults in the General Population on the Gudjonsson Suggestibility Scale-1, Gudjonsson 1997

Table 2: Yield Score Comparison

Table 3: Shift Score Comparison

Table 4: Total Suggestibility Score Comparison

Table 5: Free Recall Score Comparison

Table 6: Comparisons to Vulnerable Populations' GSS-1 Scores, Boon, Gozna, & Hall, 2008

Table 7: Comparison to Vulnerable Populations' GSS-1 Scores, Hansen, Smeets, and Jelcic (2010)

Table 8: Cognitive Load Comparisons, Drake et al., 2013

Table 9: Correlation of WASI IQ with GSS and GCS, Sondanaa, Rasmussen, Palmstierna, and Nottestad, 2010

Table 10: GSS Normative Sample, Gudjonsson, 1997

Table 11: GSS Norms for Individuals with Intellectual Disabilities, Gudjonsson, 1997

Table 12: GSS Test-Retest Correlation, Gudjonsson, 1987

Table 13: Central Tendency and Variability for the GSS-1

Table 14: Levene's Test for the GSS-1

Table 15: Shapiro-Wilk Test of Normality for the GSS-1

Table 16: ANOVA Results for the GSS-1

Table 17: Central Tendency and Variability for the PAI

Table 18: Levene's Test for the PAI

Table 19: Shapiro-Wilk Test of Normality for the PAI

Table 20: ANOVA Results for the PAI

Table 21: Welsh's ANOVA Results for the PAI

Table 22: Central Tendency and Variability for the TOMM

Table 23: One-Sample t-test comparison to Boon, Gozna, & Hall (2010) "Faking Bad" group

Table 24: One-Sample t-test comparison to the GSS-1 normative sample (Gudjonsson, 1997)

Table 25: One-Sample t-test comparison to the intellectually disabled normative sample for the GSS-2 (Gudjonsson, 1997)

Table 26: One-Sample t-test comparison between control group means

Table 27: One-Sample t-test comparison between manipulation group means

APPENDICIES

Appendix A: Previous Research Group Instructions

Appendix B: Recruitment Flyer

Appendix C: Informed Consent

Appendix D: Demographics Questionnaire

ABSTRACT

ASSESSING INSTRUCTED FEIGNER'S RESPONSE TO THE GUDJONSSON SUGGESTIBILITY SCALE

by:

Stephen S. Grabner II

Nova Southeastern University

Research has suggested that the interrogative suggestibility levels play an important role in the elicitation of a false confession within a police interrogation. The Gudjonsson Suggestibility Scale (GSS-1) is currently the only validated assessment tool that measures interrogative suggestibility levels, and it is frequently used in forensic evaluations to help support or refute false confession claims. While it is imperative that the GSS-1 readily differentiate between those who are genuinely suggestible and those who are feigning higher suggestibility levels, past researchers have raised concerns regarding the layperson's ability to feign higher suggestibility levels as measured by the GSS-1. This paper examines the ability to feign higher interrogative suggestibility levels as measured by the GSS-1 following instructions and incentive to engage in the manipulation. Undergraduate students from a local university (n=32) were randomly assigned to instructed feigners and a control group, and administered the GSS-1, as well as additional assessment tools with embedded validity indicators. Statistical Analyses, including one-way analysis of variances (ANOVA) and one-sample t-tests were employed. Results indicate that while instructed feigners can successfully decrease their free recall scores when compared to controls, they cannot successfully manipulate principle measures on the primary GSS-1 scales. These findings support claims that the GSS-1 is robust to feigning efforts when administered as outlined in the instruction manual, as well as in conjunction with sufficient distractor tasks. These findings are explored in the context of the current feigning literature, and recommendations for future use of the GSS-1 in false confession evaluations are discussed.

CHAPTER I

Statement of the Problem

Over the past 30 years, researchers have begun to explore a phenomenon in which individuals falsely confess to crimes that they did not commit (Kassin & Wrightsman, 1985). The increased attention to these false confession cases has led legal authorities and psychologists to critically examine factors that could lead to these false confessions occurring, primarily focusing on tactics utilized by police within an interrogation (Kassin et al., 2010). These efforts have increased awareness of the false confession phenomenon and have led to changes in the interrogation tactics utilized by interrogators, but has not eliminated the possibility of a false confession from being elicited throughout the course of an interrogation (Kassin et al., 2010). Further, these interrogative recommendations do not directly assist those who are currently incarcerated due, at least in part, to a false confession made in the period where these changes were not implemented during interrogative practice.

For those who have falsely confessed to a crime, there are limited courses of action to establish their innocence after an admission of guilt (Kassin, 2012). If the evidence against an individual is limited to their retracted confession, the confessor has the burden of proving that the confession is false. Often a forensic psychologist serving in the role of an expert witness is utilized to support the veracity of the claim that their confession is untrue (Frumkin, Lally, & Sexton, 2012). This expert testimony could either discuss general factors that could lead a person to falsely confess to a crime they did not commit, or they could speak specifically to the case in question. In the latter, experts base their testimony on an examination of the interrogation that brought about the retracted

confession, as well as a psychological evaluation of the confessor.

In these cases, psychologists are tasked to testify about the existence of false confessions, factors that can lead to the false confession being elicited, and sub case-specific factors that led for the confession to be elicited in that specific situation (Frumkin, 2010). Factors relied upon when preparing for such testimony include, but are not limited to, interrogation tactics utilized when the confession was elicited, the length of the interrogation, and the personal characteristics of the confessor. Explorations of the factors within the interrogation are then juxtaposed with individual assessments of the alleged confessor, often gathered through clinical assessment tools, that together help substantiate qualitative claims of a false confession being elicited (Frumkin, Lally, & Sexton, 2012).

To better support a false confession claim, an expert witness is likely to include the Gudjonsson Suggestibility Scale-1 (GSS-1) as part of the psychological battery administered to the alleged confessor. The GSS-1 is a validated clinical assessment tool that can help highlight vulnerabilities that an individual could exhibit when faced with the unique pressures present within police interrogations (Kassin et al., 2010). The GSS-1 is often cited as an important factor in these evaluations of false confession claims to assess an individual's vulnerabilities as they relate to the context of a police interrogation (Frumkin, Lally, & Sexton, 2012). Of note, the GSS-1 does not contain validity indicators, which raises concerns regarding how the assessment reacts to individuals who may attempt to manipulate their verbal behaviors in response to the assessment. Previous research has explored the layperson's ability to manipulate their suggestibility scores as measured by the GSS-1, showing mixed ability to manipulate at least some of

the subscales following instruction. This raises concerns regarding the ability for the GSS-1 to discern credible test-takers from ingenuine responders.

While there is an existing literature base that examines the layperson's ability to manipulate GSS-1 scores following written instructions, this study addresses methodological concerns within this research base and subsequently discusses how these may have impacted previous findings. Three major concerns will be addressed by the present study: first, the GSS-1 will be administered with its designated time delay to assess if this impacts the layperson's ability to manipulate their responses. Second, instructed feigners will be presented with an additional incentive to engage in their manipulation instructions in an effort to mimic incentives in the forensic setting. Finally, participants will complete additional measures alongside the GSS-1 to better understand the overall impact of manipulation instructions. Results will be compared to existing datasets in order to understand ways in which laypersons respond to the GSS-1 following instructions to manipulate their response patterns. This will assist psychologists serving in the role of an expert witness to more astutely identify individuals who may attempt to manipulate their suggestibility scores and discuss assessments that may be administered alongside the GSS in a psychological battery in evaluations of false confession claims.

CHAPTER II

Literature Review

Over the past three decades, DNA evidence has become the gold-standard to help bolster accusations of guilt, as well as attestations of innocence; this is particularly true in more serious cases involving murder, arson, and rape (Saks & Koehler, 2005). DNA evidence is widely believed to be an unequivocal liberator in all criminal cases, to such a degree that it's presence is "presumed to be the ultimate safeguard preventing wrongful convictions" (p. 127; Appleby & Kassin, 2016). Despite this prevailing belief, DNA is estimated to be present in only 20% of cases of violent crimes (Gould & Leo, 2010) and even less prevalent in non-violent cases. Further, while DNA flourishes as incriminating evidence, research assessing the utility of DNA as exculpatory evidence with lay jury members shows much narrower utility (Appleby & Kassin, 2016).

With such limited availability of DNA evidence in violent and serious crimes, and even sparser presence in non-violent crimes, police and prosecutors must rely on the historical crown jewel of evidence: a confession elicited from an alleged perpetrator. Confessions, ranging from brief admissions of guilt to full narrative of the crime in question (Kassin, 2012), remain one of the strongest pieces of evidence establishing guilt within in the criminal justice system (Drizin & Leo, 2004; Kassin & Neumann, 1997). Confession evidence is considered to be so strong that, following a confession being elicited, the investigation is often hastily concluded so that trial procedures can begin (Leo & Ofshe, 1998). Within cases in which confession evidence is present, a guilty verdict is almost inevitable due to the weight it is given, even compared to eyewitness testimony (Kassin & Neumann, 1997).

To claim that confession evidence is weighed heavily in the jurisprudence process is an understatement. To many players within the criminal justice system, a guilty verdict may appear intuitive in cases where the defendant confesses to serious crime such as murder, arson, or rape (Kassin, 2012). Research has shown that a confession, even when retracted, is given undue influence by both lay jurors (Redlich, Gheiti, & Quas, 2008) and judges alike (Wallace & Kassin, 2012). This is likely because, in the majority of criminal cases, a confession *is* a true indicator of guilt. Often confession narratives offer additional information that corroborate the individuals guilt, such as containing facts about the crime not known to the public, or leading investigators to undiscovered evidence (Leo & Ofshe, 1998). Based on its history, “because a confession is universally treated as damning and compelling evidence of guilt, it is likely to dominate all other case evidence and lead a trier of fact to convict the defendant” (p.429, Leo & Ofshe, 1998).

False Confessions

Despite the inherent trust that has been placed in confession evidence, history has demonstrated that confessions are not always accurate. Rather, researchers suggest that confession evidence falls on a continuum between a true and false confession, that varies based on the specific factors contained within the confession (Frumkin, 2008). A false confession is defined as a false statement, or a series of false statements, made in which an individual indicates an admission of guilt to a crime of which they are factually innocent (Kassin, 2017). This phenomenon is counter to the laypersons’ common sense and may be hard to conceptualize (Henkel, Coffman, & Dailey, 2008). Yet, there are volumes of evidence documenting individuals who have falsely confessed to crimes that they have not committed.

Examples are readily available throughout history, with cases seen in early United States history in cases where many women were condemned to death based on beliefs that they were practicing witchcraft; this is better known as the Salem Witch Trials of the 1690s. Based on suspicion alone, women were confined and interrogated regarding alleged practices of witchcraft. While some withstood the questioning and confinement that accompanied these accusations, others confessed to escape torturous interrogation techniques. Because of these confessions, many women were sentenced to death and subsequently hung (See: Kassin, 2008). These were not the first documented cases of false confessions within history, and they are not the last false confessions to come about using techniques akin to torture.

Inflicting pain or torturing suspects in the context of police interrogations was common practice prior to the 20th century, to such a degree that the United States Supreme Court needed to intervene. In *Brown v. Mississippi* (1936), it became law that physical interrogation techniques were unconstitutional. With torture removed from the repertoire of interrogators, police interrogation tactics within the United States shifted to using psychological approaches to bolster efficiency in garnering confessions (Kassin et al., 2018). Manuals such as *Criminal Interrogations and Confessions* (Inbau & Reid, 1962) became widely disseminated to help train interrogators in these psychological approaches. Techniques found in this manual are collectively known today as the 'Reid Method' (Leo, 2008). Currently on its fifth edition (Inbau, Reid, Buckley, & Jayne, 2013), the Reid Method has become the centerpiece of criminal investigations for the past half-decade and is utilized almost unanimously in police interrogations throughout the nation (Kassin et al., 2010).

While it could be easy to write off false confessions as byproducts of early American frontier justice that disappeared after the movement from physical to psychological interrogation tactics, this is unfortunately not the case. Over the past thirty years, psychologists and legal authorities have continued to identify false confessions elicited in interrogations that utilize psychological tactics within the Reid Method (Kassin et al., 2010). The false confessions cases identified led to a variety of consequences, ranging from brief false imprisonments to individuals currently incarcerated due, at least in part, to a confession to a crime they did not factually commit (see: Leo & Ofshe, 1998; Perske, 2008; Perske, 2011). To better identify and assist individuals negatively impacted by false confessions, organizations such as the Innocence Project have been established (www.innocenceproject.org).

Over the past 16 years, the Innocence Project has spearheaded the exoneration of at least 350 people in the United States through advocating for examining DNA evidence. Within these initial 350 cases, false confessions were found to be a contributing factor in at least 28% of cases (Kassin, Redlich, Alceste, & Luke, 2018). This does not accurately represent the true number of false confession cases throughout history, which is likely much higher. However, much of the work completed by these organizations has been with male exonerees, likely due to the increased readiness of DNA evidence to assist their exoneration. This does not consider cases in which women falsely confessed, which indicates that these estimations may be lower than the true base rate of false confessions in today's criminal justice system (Walker, Conte, & Grabner, 2014).

It is difficult to determine the base rate for the frequency of wrongful convictions or the frequency in which false confessions are elicited because these numbers are not

collected and very difficult to measure. However, it is possible to extrapolate from similar data. For example, within capital cases the conviction error rate is estimated to be between 3.3 and 5% (Risinger, 2007). Such prevalence shows the scope of the issue in today's criminal justice system, and indicates that the problem deserves continued exploration in both the research and judicial realms.

False Confession Typology

False confessions are considered a significant contributing factor in wrongful conviction cases throughout history (Leo & Ofshe, 1998), yet it has only been within the last several decades that researchers have attempted to systematically identify cases and examine this phenomenon. Kassin and Wrightsman (1985) were the first to introduce a classification system to better understand and identify the mechanisms that influence the elicitation of a false confession in the context of police interrogations. Their research identified three different categories of false confessions: voluntary confessions, coerced-compliance confessions, and coerced-internalized confessions.

Voluntary confessions are classified as false assertions of guilt without an identifiable external source of pressure on the individual to confess to the crime. Factors that could lead an individual to a voluntary false confession include mental illness (e.g. schizophrenia or other delusions) or seeking notoriety from subsuming the identity of the factually guilty party. Coerced-compliant false confessions cases are identified when a person admits culpability verbally in the context of an interrogation, while internally maintaining their innocence. These cases result from an individual succumbing to pressures to escape the interrogation, and are marked by swift recantations when the individual is removed from the stressors of the interrogation environment. Often this

would account for acquiescence and compliance to interrogative questioning rather than truly believing what is being said. Third, coerced-internalized confessions represent admissions of guilt which are accompanied by the individual internally accepting responsibility for the crime. In these cases, the person leaves the interrogation with the belief that they, at least in part, were guilty of the crime in question. This is often maintained following removal from the pressures of the interrogative setting for a period of time. Coerced confessions highlight the role of the interrogation on the elicitation of a false confession, and have garnered the most literature (see Kassin et al., 2010).

However, it is understood that not all pressures come from the interrogative environment. McCann (1998) proposed that external influences could also impact the verbalization of false confessions, leading to a fourth category: coerced-reactive. This could include pressures to confess to protect a true perpetrator, or threats made outside of the interrogation that influence verbal behavior during the interrogation. These have become increasingly noted in female false confession cases, where confessions are elicited to protect a significant other (Grabner, Conte, Walker, Nagle, & Shaprio, 2015). More recently, Frumkin (2010) proposed that in some cases individuals have committed a plethora of prior offences and subsequently confess to a crime they did not commit due to the confession; these cases would be classified as coerced-substituted false confessions.

Although these five false confession types may not be detailed enough to encompass all the nuances contained within each situation in which a false confession is elicited, Kassin and Wrightsman's false confession typology is most commonly utilized by researchers to discuss broader trends within the false confession cases of the 20th and 21st century. This allows for more detailed discussion of individual pressures within the

interrogative setting, individual factors such as level of concern regarding social approval, as well as levels of suggestibility. It is from this framework that researchers have broadly explored factors that lead to these confession types, as well as explored means by which these vulnerabilities within the legal process can be neutralized.

Protecting and Defending Against False Confessions

Given the weight which confession evidence is given in the court of law, researchers and legal authorities have increasingly focused on how to prevent false confessions from being elicited within an interrogation. While this assists in increasing the effectiveness of interrogation tactics and decreasing the base rate of false confessions to some degree, additional efforts are necessary to help defend those who have already falsely confessed and must battle their own confession (Leo & Ofshe, 1998).

Challenging the veracity of confession is relatively new concept within the courts, with a lot of development within the 1980s. *Colorado v. Connelly* (1986) upheld that statements, even those given while suffering from symptoms of a mental illness such as delusions, were admissible and collecting incriminating statements as evidence did not on its own violate the fifth or fourteenth amendment. This sets the precedent for specific evidence of coercion being required to violate the Due Process Clause, but made no mention of what could occur if coercion was present.

Prior to *Crane v. Kentucky* (1986), the reliability of the confession was not able to be challenged within the trial. In this case, a 16-year-old boy falsely confessed to a murder following a lengthy interrogation. During his trial, Crane's defense team was not permitted to introduce testimony discussing facts about the interrogation because the trial court would not admit testimony that only served to discuss the issue of the voluntariness

of the confession. This was upheld by the Kentucky Supreme court; however, the U.S. Supreme Court ruled unanimously that this right was protected by the Confrontation Clause of the 6th Amendment and the Due Process clause of the 14th Amendment (*Crane v. Kentucky*, 1986). Following this ruling, it became possible for defendants to introduce evidence that helped give credence to their claims of false confessions.

Over the past thirty years, *Crane v. Kentucky* (1986) has set the legal precedent for examining confession veracity. In the instances where this avenue is pursued, forensic psychologists are called to testify in order to support the trier of fact understand false confessions as a psychological construct and the process in which they can be elicited (Frumkin, Lally, & Sexton, 2012). However, the true utility in expert witness testimony occurs when the psychologist can provide both general information regarding the elicitation of false confession as well as case-specific information.

For experts who are called to provide case-specific false confession testimony, Frumkin (2010) offers helpful guidelines that help identify constructs that should be considered. This focuses on understanding and examining individual and interrogation factors that could have influenced the individual during the interrogation, which can be gathered through a psychological interview and a comprehensive psychological assessment battery. This should be followed by intensive examination of the available data from the interrogation such as a transcript, as well as audio or video recordings, when available (Frumkin, 2010). Conclusions drawn from this information should be connected to empirical research and then presented in a digestible manner to the judge and jury to help support their understanding of the case.

To bolster the empirical support for these individual and interrogative factors

discussed in this testimony, the culmination of the individual risk factors must be examined to understand the system in which these confessions can occur. This process can be understood best by applying accident causation models such as the Cumulative Act Effects model (Reason, 2000) to help conceptualize the individual vulnerabilities and how those lead to adverse outcomes. The Cumulative Act Effects model was initially proposed to assess lapses and weaknesses in defensive systems that intend to prevent patient harm in hospital settings. As such, the model accounts for the many barriers that are already in place in a system as protective layers; individually, these layers have weaknesses but cumulatively decrease the chance of an adverse outcome occurring. Although the system is insufficient in stopping all adverse outcomes, by looking at the weakness in the individual layers it is possible to decrease the likelihood of future adverse outcomes that follow a similar trajectory.

Examining the false confession process within a Cumulative Act Effects model clarifies the complex interaction that leads to a confession occurring, and highlights individual areas of risk that increased the opportunity for false confession to be elicited. An expert utilizing this model would examine the trajectory between the crime occurring and a false confession being elicited, which would help highlight unique vulnerabilities exhibited that may have led to the false confession being elicited. These include individual vulnerabilities, as well as police and interrogation factors that together impacted the elicitation of a false confessions. As Kassin (2017) states, “in the service of social justice, psychologists-armed not only with a cache of forensically focused studies and wrongful convictions but with core principles of psychology- are uniquely positioned” (p.959) to discuss the vulnerabilities present within the interrogation. By

looking at the interrogative process through this lens, experts can identify the areas that lead to the most adverse outcomes and identify additional safeguards that could prevent future negative outcomes following a similar trajectory.

It should be noted that the majority of clinical and research attention within the false confession literature has been on police-centric interrogation factors. Examining these factors have highlighted many detrimental means by which false confessions have arisen based on police actions through the various stages of confinement and interrogation. For example, data suggests that police-centric factors such as mistaken eyewitness identification, utilizing interrogation tactics that result in false confessions, or investigators experiencing “tunnel vision” (p. 481, Gould & Leo, 2010) in which law enforcement becomes convinced of a suspect’s guilt to such a degree that they discount alternatives hypotheses with more credence. Further research has identified key interrogation techniques that are staples of the Reid Technique also significantly increase the likelihood of a false confession being elicited. Such tactics include maximizing the evidence police imply they have on a suspect (known as “Maximization”), minimizing the offense and the subsequent consequences of confessing (known as “Minimization”) and the presentation of outright false evidence. These tactics psychologically manipulate individual’s responses, leading to an increase in both false and true confessions (See Kassin et al., 2010 for full review on the status of this interrogation research).

While researchers call for changes to the interrogation process such as mandatory video-recording of all police interrogation based on these findings (see Kassin et al., 2010), it becomes imperative to also understand the individual factors that predispose a person to a higher likelihood of falsely confessing when exposed to these police-centric

factors. This line of research has identified unique variables that are present in the custodial and interrogation setting, and integrates an individual's personal vulnerability into how they are conceptualized. In the forefront of the confession literature is examining the motivations for verbal behavior in an interrogation, as they represent how the police-centric factors are moderated by individual factors to lead to these false confessions.

Suggestibility, Compliance, and Acquiescence

To best understand why someone would falsely confess to a crime, researchers focus on the motivations of verbal behavior and intent of the verbalization. Gudjonsson (2003) highlighted three degrees by which false admissions could come about. The first is verbalizations that represent acquiescence to pressure, which have been widely understood within the literature as experiencing motivation to respond to questioning, often in the affirmative, to appease others without any internal processing regarding what is being verbalized or agreed to (Cronbach, 1946). However, in many cases there is a degree of internal consideration to what is being verbalized. The inclusion of such contemplation of the ramifications of the verbal behavior would be considered as either compliance or suggestibility. Compliance focuses on verbalizations that are not internally consistent with the person's beliefs; rather, the person is simply complying with forces outside of themselves but internally maintaining their innocence. This is demonstrated in the coerced-compliance subtype of false confessions, and often marked by hasty recanting of confessions when free from the pressures.

However, in some cases these statements are, at least for a short time, believed to be true. The construct of suggestibility helps conceptualize this internal acceptance of

what is being verbalized, representing an internal shift to accept it as the truth. This is conceptualized as a coerced-internalized false confession. This internalization process has been the primary focus of a complete branch of false confession research, as the shift in verbal behavior represents a different underlying mechanism than both acquiescence and compliance.

Suggestibility. The definition of suggestibility has evolved throughout its lifetime as the understanding of what it means to be “suggestible” has been challenged by researchers over the past two centuries. Prior to the 20th century, the term suggestibility was primarily used when discussing participant’s response posthypnotic suggestions from a hypnotist. Suggestibility was thus viewed as an unconscious mechanism with little relevance in conscious decisions; therefore, there were no early attempts to measure it empirically. Suggestibility remained in the shadows of the unconscious until 1886, when researchers such as James Cattell began studying witness memory and testimony to understand the level of subjectivity of memories.

Cattell’s research began by exposing Columbia students to a staged event and subsequently asking questions about what the students had seen. In addition, students were asked to report their perceived confidence in their answers. Cattell found the confidence level in the responses did not correlate to accuracy; some individuals were confident without regard to the accuracy of their responses, while others were strikingly insecure about their responses despite being correct. His research was one of the first to help understand the volatile nature of witness memory, which at the time was considered to have high credibility.

Following Cattell’s work, psychologists began to become more interested in

memory and its malleability. Alfred Binet (1900) performed a study measuring participant's reaction to leading questions. He first showed participants various pictures, then asking participants to recall what they had seen. Notably, the wording of Binet's questions conveyed a desired response from the participants. By using leading questions, Binet elicited distorted recollections of the image that the participants believed to be true. These findings cemented the foundation for research regarding the volatile nature of witness memory, and became a pivotal moment in suggestibility research. This was the first time that research had indicated that being suggestible could be a conscious and active process that could be studied and measured. To understand the human experience as malleable rather than concrete allowed for lines of research questioning what exactly could change one's memory, how likely that was to happen, and explore protective and risk factors for these memory distortions.

Research then began differentiating "Primary" and "Secondary" suggestibility; primary suggestibility referred to the unconscious processes underlying the utility of hypnosis, while secondary suggestibility was outlined as the conscious process similar to gullibility (Eysenck & Furneaux, 1945). Here research split; the present line of research developed from the further exploration on secondary suggestibility, which was explored as the level of susceptibility to the influences of others on a conscious level. This secondary suggestibility allows for more specific and active research to isolate important influences in memory formation, as well as how individuals acted upon those memories.

Interrogative Suggestibility. Although researchers have worked to explore the concept of suggestibility over the course of the eighty years following Cattell and Binet's work, suggestibility and memory malleability seemed to be ignored in the context of

police interrogations. It is within this conscious process that the research delved to better understand the impact this plays within interrogations. Today, when considering suggestibility, which is defined as the likelihood “to believe that what someone says is true or may be true” (Suggestible, n.d.), indicating suggestibility is geared towards the more active and conscious process that Binet and Cattell offered. Although researchers were identifying the impact of interrogation stressors on false confessions (Kassin & Wrightsman, 1985) during this time, the underlying constructs that made the individuals more susceptible to these pressures were not yet clarified.

It wasn't until 1986 that Gudjonsson and Clark proposed a groundbreaking model of suggestibility specifically addressing interrogations, arguing that previous research on suggestibility did not accurately map onto the unique pressures present when individuals are interrogated in a custodial setting. Gudjonsson and Clark (1986) stated that previous understandings of suggestibility did not consider the nature of the power differential in custodial settings or the potential for false confessions within such interrogation. Further, there had been no consideration of how the implications of statements made in these situations impacted response patterns, as the consequences of this verbal behavior is unique to custodial and judicial settings. By not exploring these areas, Gudjonsson and Clarke argued that suggestibility studies conducted prior to that date were not relevant when attempting to understand any police interrogation, and that new paradigms were needed to explore these phenomena.

Gudjonsson and Clark (1986) highlight several areas within interrogative situation that differ from other social interactions where suggestibility could impact the outcome of responses. When being interrogated in a custodial setting, they argue individuals

experience: 1) feelings of uncertainty about the correct answers to questions they are being asked; 2) feelings of being expected to know the correct answers; 3) negative feedback when responses are “incorrect;” and 4) feelings of rapport and trust (Gudjonsson & Clark, 1986). These factors intertwine as the interrogation continues, and over time place unique pressures on an individual being interrogated. Gudjonsson and Clark determined that a more clearly defined construct was necessary to adequately encompass these interrogative pressures. This was labeled as Interrogative Suggestibility (IS), and was defined as “the extent to which, within a closed social interaction, people come to accept messages communicated during formal questioning, as a result of which their subsequent behavioural response is affected” (Gudjonsson & Clark, 1986, p. 84). The authors note that this does not necessitate being interviewed by a police officer, but rather in any custodial setting that contains a similar power dynamic.

Based on this understanding of IS, Gudjonsson and Clark proposed that two important forces at play within an interrogation, which they coined as “Yield” and “Shift.” Yield is an individual’s reaction to “give in” to a question that is leading in nature, such that there is a perceived expectation to answer in a certain way. Thus, an individual could yield to this pressure to give the interviewer the response that they convey as desired, with a belief that doing so would be beneficial for the individual in some way. In tandem with this process, shift refers to changing previous responses following receiving negative feedback from a person of authority regarding. Shift results in changing initial responses to please the authority figure. The pressure to yield when initially giving responses, as well as shifting previous responses, were proposed to work together during an interrogation to elicit both true and false confessions at higher rates.

Notably, some individuals are more likely to yield to this pressure than others, some are more likely to shift their responses, and others may not be impacted at all by the interrogative pressures. This highlights the importance of understanding not only whether these pressures are present, but also measuring the degree to which the person in the situation would be impacted by them.

Together Yield and Shift helped define what Gudjonsson and Clark saw as the key factors that had yet to be explored when discussing the interrogation. The literature has begun to explore Yield, Shift, and Interrogative Suggestibility in various contexts to better understand how they could be manipulated (Roos & Gow, 2007), how they are modulated by levels of intelligence (Beail, 2002), and more specifically how they impact the elicitation of false confessions (Kassin et al., 2010). This conceptualization of interrogative factors paved a new road to understanding suggestibility in the interrogation context, but initially lacked a reliable way to measure and assess Yield and Shift levels in practice.

Suggestibility and Prevention of False Convictions

Having established the impact of Yield and Shift within the interrogation, it becomes imperative to consider how this information could be applied to the current difficulties posed by false confessions. While many researchers have utilized fruit from this research to inform calls for change within the interrogative process (See Kassin et al., 2010 for a complete review), fewer have focused on how this information can help support those who have already confessed. As noted, many players in the legal system appear to believe false confessions are counterintuitive at best, and therefore once a false confession has been elicited there are limited courses of action the individual can

take to establish their innocence (Leo & Ofshe, 1998).

In the absence of other exonerating data such as DNA evidence, the defendant must attempt to challenge the veracity of their own confession in the court system. Kassin, Redlich, Alceste, and Luke (2018) examined the problem faced by defendants who must overcome their own confession. Often this is done using expert witness testimony, which helps either frame the existence of false confessions in general or addresses factors that led to the false confession within the specific case. The authors surveyed 131 experts who had published on interrogations and confessions to understand the acceptability and the influence of their testimony. Their results indicate that experts in the field estimated receiving 3,889 requests to testify relating to false confessions, of which they testified in approximately 1,647 cases, which almost exclusively was in service of the defense rather than the prosecution. With a such a large number of testimonies regarding false confessions as a whole, information regarding individual factors influencing false confessions become even more significant are to explore.

Rules of evidence and false confession testimony admissibility. While confession evidence is generally accepted without significant questioning, for expert witness testimony to be accepted as evidence in the United States legal system it must rise to certain legal precedents; these are generally set by legal precedent such as U.S. Supreme Court rulings. Historically, evidence submitted to the court must meet what is known as the *Frye* test, which is based on *Frye v. United States* (1923) D.C. Court of Appeals ruling. The *Frye* test established the need for expert witness testimony to be based on research that was generally accepted by the relevant scientific community. While helpful in establishing the need for reliability of the information being accepting

into the courtroom, the general acceptability requirement is vague and its standards were not applied uniformly. Presently, only eight states exclusively rely on the *Frye* test as their standard for accepting expert witness testimony.

The remaining thirty-nine states follow a more recent ruling following *Daubert v. Merrell Dow Pharmaceuticals Inc.* (1993; Kassin et al., 2018). The *Daubert* rule, based on Rule 702 of the Federal Rules of Evidence (1975), a U.S. Supreme court raised the need for research supporting expert witness testimony to not only be accepted in the scientific community, but to also be considered reliable and valid. The decision whether expert testimony is based on the current science resides with the judge, and they must determine whether this information assist the triers of fact in the current case (*Daubert v. Merrell Dow Pharmaceuticals Inc.*, 1993). Together, these standards set the bar for psychological expert witness testimony; more importantly, it establishes the barriers that need to be overcome if trying to defend against a false confession.

Kassin et al. (2018) indicated that for psychological experts testifying on false confession cases, the *Daubert* standards have been met based on both the large literature base as well as the large number of accepted testimonies they were able to identify in their sample. While an expert can testify regarding false confessions, there is no set standard on what kinds of information should be present in these testimonies or what areas should be addressed by the testimony. Suggestions are presented in Frumkin (2010) that focus on the use of reliable and valid measurement tools to examine the individual case in question, without which researchers would be left with presenting only facts about false confessions in general. To increase the precision of these testimonies, and to further support the *Daubert* requirements of being both reliable and relevant to the current case,

experts utilize specific assessment tools to help bolster their findings.

The Gudjonsson Suggestibility Scale

Although no standardized battery of assessments having been established when assessing false confession claims, specific tools have been developed that shed light on constructs specific to the enigma that is a false confession (Frumkin, Lally, & Sexton, 2012). This paper will turn towards understanding the tools used to measure suggestibility as indicated by verbal behavior, namely the Gudjonsson Suggestibility Scale.

Creation and Development

During the late 1980s while involved as an expert witness in the United Kingdom, Gisli Gudjonsson was asked to assess the suggestibility level of multiple individuals with intellectual disabilities who were believed to have falsely confessed during a police interrogation. However, at this time there were no validated assessment tools in use that could assess interrogative suggestibility, namely the Shift and Yield constructs proposed in his newly proposed IS model (Gudjonsson, 1997). With no assessment tool viable for his needs, Gudjonsson found it necessary to “develop an instrument that could help identify people who were particularly susceptible to giving erroneous accounts of events when subjected to questioning” (p.3, Gudjonsson, 1997). He aimed to objectively measure the extent to which individuals gave into leading questions and how they responded to negative feedback in a constructed behavioral assessment. More importantly, for Gudjonsson it was paramount that the tool be empirically validated so that it could be admitted as a part of an evaluation completed within the criminal justice system. With this data, Gudjonsson intended to help identify those who may be

vulnerable in interrogations, or utilized in cases where a confession is retracted.

Through his subsequent research, the Gudjonsson Suggestibility Scale (GSS-1; Gudjonsson 1984) was developed. An alternative form (GSS-2, Gudjonsson 1987) was later created to help understand test-retest reliability and broaden the range of applicability outside of the forensic context. Although the GSS-1 and GSS-2 are identical in administration format, they differ in narratives read out to the participant and the subsequent questions regarding that narrative: the GSS-1 contains a narrative about a fictitious robbery, whereas the GSS-2 contains a narrative about a bicycle accident without forensic implications. In the manual for the Gudjonsson Suggestibility Scale, Gudjonsson notes the GSS was developed to “measure objectively the vulnerabilities or proneness of people to give erroneous accounts when interviewed. The GSS-1 is particularly applicable to police interviewing contexts, but can be applied to any interview situation, including clinical practice.” (p.ix, Gudjonsson, 1997). By understanding how the individual responds to pressures in a simulated environment, it is possible to extrapolate how they would respond in an interrogative setting. Through careful administration, the GSS-1 provides valuable insight

Administering the Gudjonsson Suggestibility Scale

To administer the GSS-1, it is important to first consider the environment and presentation of the tool. First, the individual should be placed in an environment that is similar to those found in an interrogation. The room is suggested to be neutrally decorated and only those administering or being administered the assessment should be present. The person who administers the GSS-1 is encouraged build rapport with the individual prior to beginning the test. To begin the assessments, the participants are first

informed that they are taking a memory test to conceal the true nature of the suggestibility measure; participants are not told about the true nature of the assessment. Participants are first asked non-standardized questions regarding their memory to solidify the presentation of the assessment as a memory test.

The participants are then told the following verbatim: “I want you to listen to a short story. Listen carefully because when I am finished I want you to tell me everything you remember” (p.11; Gudjonsson, 1997). Participants are presented with a short narrative that contains 40 pieces of information about an incident; the GSS-1 narrative discusses a robbery, while the GSS-2 narrative discusses a bicycle accident. After the narrative is played once, participants are asked to recall as many aspects of the narrative as they can. All answers are intended to be recorded or written down verbatim by the researcher for scoring. Participants are asked to state as many facts from the story they can remember, but are given no feedback about the accuracy of their responses or overt encouragements. After discontinuation of spontaneous recall, participants are given unrelated distractor tasks for up to 50 minutes.

Following this time delay, participants are again asked to recall as much of the initial story as they can and their responses are recorded; again, no feedback as to the accuracy of these statements are provided. Next, participants are asked twenty scripted questions about the narrative, which promote forced-choice responses (i.e., yes/no questions). Fifteen questions contain information not reported in the narrative, and thus are leading questions due to the manner in which they convey to the participant there is an answer in the responses presented. The responses to these questions is believed to be representative of the individual’s initial level of yield to leading questions. If the

participant does not give a clear response, they are asked to solidify their statement. Upon answering all twenty questions initially, all participants are told the following: "You have made a number of errors. It is therefore necessary to go through the questions once more, and this time try to be more accurate" (p. 4-5, Gudjonsson, 1997). This negative feedback is presented in a firm and clear manner, regardless of the participant's actual responses to these questions. The participants are then asked the same questions that were presented prior to the negative feedback. Following the second administration of the questions, the assessment is complete.

Scoring the Gudjonsson Suggestibility Scale

The GSS contains six main subscales that can be generated from the participant's responses during the free recall and question response phases. The first two subscales, Immediate Recall and Delayed Recall, measures the number of aspects of the narrative that are correctly recalled during their respective recall phases. A maximum score of forty correct aspects of the narrative could be recalled by participants during each recall period. Scoring information to help understand how to code these free response answers are outlined in the GSS manual (Gudjonsson, 1997), with partial remembering of facts also garnering some credit.

The third and fourth subscales are the "Yield 1" and "Yield 2" scores, which indicates the number of leading questions the subject yielded to during the first and second questioning period, respectively. The significant difference is that Yield 2 scores follow the administration negative feedback. Both Yield 1 and Yield 2 have a maximum score of 15, as the five non-leading questions are not incorporated into this score. Participants who respond to leading questions with an affirmative are scored as having

yielded to the leading nature of the question. Additionally, participants who agree with either of the false alternative questions are scored as having yielded to the question. Examples of statements that qualify to be scored as a yield are found in the administration manual (Gudjonsson, 1997).

The fifth main subscale, “Shift,” notes the number of significant changes in responses when comparing responses before and after the negative feedback administration. For example, participants who change an affirmative answer to a negative one are scored as shifting to the negative feedback. Shifted responses are measured for all 20 questions, making it possible to measure changes in responses to non-leading and leading questions. The final subscale score is “Total Suggestibility,” which is the sum of Yield 1 and Shift scores, showing the level of IS as proposed by Gudjonsson and Clark (1986). Total Suggestibility scores ranges from 0 to 35, and accounts for both the reaction to leading questions and responses to negative feedback together.

Supplemental scores can also be generated from examining the recalled aspects of the GSS-1 narrative. These scores represent the number of confabulations made by the participant; this can be broken down into distortions and fabrications of facts related to the narrative distinguish types of memory errors (Gudjonsson, Rutter & Clare, 1995). However, confabulation scores were not a part of the original form of the test and are infrequently used in forensic applications of the GSS-1 (Gudjonsson, 1997).

Interpreting Gudjonsson Suggestibility Scores

The GSS-1 scores are measures of “actual behavior in a brief interrogative situation (p.29, Gudjonsson, 1997), and give insight into how an individual may react in a similar custodial or interrogative situation in which they are presented leading questions

or negative feedback. Thus, behavioral comparisons could be made based on scores that vary when compared to appropriate normative samples. Gudjonsson (1997) identified means for multiple UK normative samples for use in score comparison for the GSS-1. These comparison samples include, among others, healthy adults in the general population (See Table 1), a court referral sample, an Icelandic prisoner sample, prison inmates, and juvenile offenders. Scores obtained should be compared to their appropriate normative group and use that comparison to “establish how unusual or abnormal the scores are” in comparison (p.29, Gudjonsson, 1997). Research has suggested that Interrogative Suggestibility is relatively stable across nationalities, and likely these norms are applicable cross-culturally for clinical interpretation (Frumkin, Lally, & Sexton, 2012).

Table 1

Descriptive Statistics for Adults in the General Population on the Gudjonsson Suggestibility Scale-1

	<i>Mean</i>	<i>SD</i>	<i>Range</i>	<i>Std. Error</i>
Immediate Recall	21.3	7.1	4-36	2.78
Delayed Recall	19.5	7.5	4-34.5	3.48
Yield 1	4.6	3.0	0-13	1.36
Yield 2	5.6	3.8	0-15	1.61
Shift	2.9	2.5	0-12	1.00
Total Suggestibility	7.5	4.6	0-21	1.64

n=157; Gudjonsson, 1997

When compared to the appropriate normative group, Yield 1 or Yield 2 scores can help psychologists understand how a participant responds to leading questions in similar settings. Higher scores as compared to the appropriate normative group would indicate that an individual is more likely in an interrogative setting to yield to pressures. Similarly, higher shift scores could be used to indicate a greater susceptibility to negative feedback. While interpretations of these scores do not directly determine whether a person falsely

confessed, they give a better understanding to a forensic psychologist as to the level in which the individual may be vulnerable to these pressures if they were present in the interrogation. For example, a finding of an individual with a high yield would be ineffective in supporting false confession claims in cases where an examination of the interrogation transcripts indicate the absence of leading questions. Similarly, while high shift scores compared to the appropriate norms could suggest a higher likelihood to change questions in response to negative feedback, it is imperative to evaluate the whether any negative feedback was administered during the interrogation in which an alleged false confession was elicited.

Use of the Gudjonsson Suggestibility Scale

Due to the inherent trust that is offered to confession evidence (Kassin & Wrightsman, 1985), forensic psychologists must fight an uphill battle to help demonstrate that not only can confessions can be false, but how they are elicited in the context of a police interrogation. False confession cases that otherwise would fall to the whims of the criminal justice system now has support in their last lines of defense against a false conviction through an empirically supported evaluation conducted by a forensic psychologist (Frumkin, Lally, & Sexton, 2012). The GSS-1 is the only assessment tool for measuring interrogative suggestibility currently in use (Kassin et al., 2010), and as such it is most relevant for use in forensic evaluations where there is a question regarding the impact of interrogative pressures on the elicitation of a confession. While not the only relevant tool to these evaluations, it provides unique information to support claims that otherwise could not be empirically supported.

Although initially developed for use in the United Kingdom, the Gudjonsson

Suggestibility Scale have been adapted for use international use. Researchers interested in utilizing the GSS-1 have used various translations, including Dutch (Merckelbach, Muris, Wessel, & van koppen, 1998), Polish (Polczyk, 2005), and English translations (Polczyk, 2005), to name a few. Gudjonsson, Rutter and Clare (1995) attest that ethnicity has only minor influences on GSS-1 scores when controlling more influential factors such as memory, intelligence and anxiety. They suggest that their recommendations to conservatively interpret the results as compared to the normative sample adequately controls for any cross-cultural differences in normative comparison. Therefore with appropriate translations, the GSS-1 is available for use internationally without reports of validity interference.

To use the GSS-1 in the context of United States forensic evaluations, Frumkin, Lally and Sexton (2012) collected 334 individual GSS-1 protocols used in United States criminal evaluations as part of confession-related proceedings. For the US administration, the narrative and questions are edited such that that “holiday” is replaced with “vacation” and “pounds” is replaced with “dollars” to better reflect American nomenclature. The authors reported that when comparing their sample with samples from the United Kingdom and Iceland, differences were minimal: Yield 1 scores were similar, while Yield 2, Shift, and total GSS-1 scores were one-quarter to one-half a standard deviation greater in the United States sample (Frumkin, Lally, & Sexton 2012). Although these reflect statistically significant differences, the authors submit that this does not reflect a clinically significant difference in the individual scale scores due to the conservative way they suggest the GSS-1 be interpreted. Therefore, the authors propose that “intuitively, there is little reason to believe that those from the United States are any more suggestible

that those residing in the Great Britain or in Iceland” (p. 760-761, Frumkin, Lally, & Sexton, 2012). When using the GSS-1 within the United States for forensic cases, clinicians should consult both Gudjonsson’s (1997) sample for court referrals and offender groups as well as Frumkin, Lally, and Sexton’s (2012) U.S. Forensic sample for additional comparisons.

While Frumkin, Lally and Sexton (2012) attest that the GSS can be used in the context of criminal justice evaluations within the United States for confession-related assessments using the proper adaptations for local nomenclature, they note that “GSS scores provides only one piece of information important for the forensic psychologist to consider when evaluating factors related to the voluntariness or validity of a confession” (p. 761). They suggest that additional measures should be utilized in conjunction with the GSS-1 to make conclusions about the interaction between the interrogative situation and intrapersonal factors on the veracity of a confession (Frumkin, Lally & Sexton 2012).

Manipulating Suggestibility Scores

The addition of the GSS-1 within a forensic evaluation for a false confession is essential to include when individuals are claiming that their confession was coerced in some way throughout a police interrogation, which is the most common setting for false confession elicitation (Kassin et al., 2010). While GSS-1 scores that indicate higher levels of Interrogative Suggestibility to leading questions and/or negative feedback would lend credence to an individual’s claims to have falsely confessed to a crime, it is important to consider that obtaining an elevated score on the GSS-1 subscales would similarly assist someone who may only *claim* to have been adversely impacted by interrogative pressures. It is conceivable to believe that a factually guilty individual

would want to appear more suggestible if doing so could lead to avoiding a trial, having their sentence reduced, or overall have decreased legal repercussions (Baxter & Bain, 2002). These behaviors are commonly referred to as “faking bad” or as a non-credible performance.

Feigning symptomology is most commonly discussed in terms of malingering, which describes the exaggeration of symptom severity or falsifying psychiatric symptoms that is motivated by primary or secondary gains (Mittenberg, Patton, Canyock, & Condit, 2002). In malingering cases, a person actively exaggerates existing features of themselves or fabricates features entirely to appear a certain way; this is generally an intentional deception on the part of the individual and not a subconscious one (Ziegler & Boone, 2013). The benefits from successful feigning vary; they could be minor, such as being wrongfully granted access to a disabled parking spot, or larger, such as securing VA health care and monetary benefits. When malingering in forensic evaluations, the stakes are even higher; successful feigning could lead vast sentence reduction in criminal cases, or even to large monetary gains in civil suits. In addition, the inherent motivation that individuals could experience to exaggerate symptoms in these cases, lawyers are aware of the tests being utilized by psychologists to measure constructs such as suggestibility, and may give their clients specific instructions on how to “beat” the test to better their overall defense (Hanson, Smeets, & Jelicic, 2010). This poses a significant risk to the integrity of psychological evaluations, and must be considered when conducting any forensic evaluation that could result in suggestions of innocence.

To better understand a base rate for malingering and symptom exaggeration, Mittenberg et al. (2002) surveyed 131 neuropsychologists and asked them to identify

percentages of cases in which probable symptom exaggeration or malingering were present during their career conducting forensic cognitive evaluations. The sample reported on 33,531 criminal and civil cases, and estimated that malingering or exaggerative cognitive impairment occurs in 19-23% of criminal evaluations based expert's reports of their cases. The authors note that this statistic was comparable to statistics obtained with previous research and "representative of those [rates] observed in a variety of practice settings" (Mittenberg et al., 2002; p.1097). Larrabee, Millis, and Meyers (2009) proposed a more conservative for the base rate of malingering of cognitive dysfunction in some settings to be around 40%, plus or minus 10%.

While malingering psychiatric symptoms does not correctly encapsulate the behaviors seen when someone chooses to manipulate responses for constructs such as suggestibility, research within malingering is quite relevant when considering the similarities in motivation when conducting false confession psychological evaluations. If responding in a specific way on assessments such as the GSS-1 could lead to freedom from imprisonment or a decreased sentence, there is a strong motivation for a select population to respond in a deceptive way to reach that goal. If researchers apply even a remote percentage of this base rate toward individuals being assessed for false confession claims, it becomes imperative to be able to differentiate between genuine and feigning behaviors to avoid undercutting the validity of psychological expert witness testimony. It is vital to be able to differentiate between those who have genuine expression of symptoms and those who are instructed to exaggerate or feign suggestibility.

Validity Indicators

For psychologists who work with latent variables such as intelligence, memory,

and suggestibility, it can be difficult to readily identify genuine and non-genuine performance. Research has begun to identify ways to detect deception using functional magnetic resonance imaging (fMRI) and positron emission tomography (PET; Kingery & Schretlan, 2007) equipment that is often found in hospital settings, but this is still a developing field. However, even with significant findings utilizing these methods, it is impractical to utilize such specialized equipment in everyday psychological or forensic evaluations. Instead, psychologists rely on incorporating validity indicators into their assessment tools to ensure the person is responding to the test in a valid way. This can be imbedded in other assessment tools, or could be a standalone assessment to add to a battery. Take for example the Personality Assessment Inventory (PAI), a personality test with strong empirical support that is utilized in both clinical and forensic settings (Morey, 2007). The author found it imperative to incorporate imbedded validity indicators to assess the many ways a person may be ingenuinely responding. The PAI thus had its test items formatted in such a way that they allow the assessor to identify a variety of test taking attitudes, including assessing for those who are underreporting symptomology (i.e. presenting themselves in a falsely positive light), those who are exaggerating symptomology, and assesses for those who are responding inconsistently throughout the tests (Morey, 2007). These questions are imbedded alongside other test items and are undetectable to those who are being given the test without prior knowledge of the questions. These indicators are used in interpreting the data such that data reported in an invalid way are not inaccurately interpreted as true representations of the individuals' current level of functioning.

Gudjonsson Suggestibility Scale and Validity Indicators. While the PAI can

assess test-taking attitudes and identify disingenuous responding patterns, not all assessment tools have the ability to add in validity indicators into the questions seamlessly. The GSS-1 does not contain any validity measures in its current form, and there currently is no consistent way to assess for those motivated to feign their level of suggestibility based on the GSS-1 scores alone. At present, the questions embedded within the GSS-1 cannot to be used as a validity indicator. Therefore, a psychologist has two options when attempting to measure responding patterns: either add questions to the GSS-1 that could impact the integrity of the original assessment, or administer the assessment alongside additional measures. Gudjonsson and Sigurdsson (2004) suggest that tests such as the GSS-1 could effectively use a separate assessment with effort measures to better understand responding styles when administered alongside one another. However, to date no study has assessed the GSS-1 alongside standalone measures of validity measures.

Empirical Research Regarding Feigning Suggestibility

The GSS-1 contains no imbedded validity indicators to differentiate those who are genuinely suggestible and those who are attempting to manipulate their responses. To address whether this merits concern, researchers have tested hypotheses centered on the ability of instructed feigners to significantly manipulating GSS-1. Researchers believe that by comparing control groups and instructed feigner's responses to the GSS-1 in experimental settings, it could be determined if the GSS-1 is susceptible to feigning attempts. Researchers further hypothesized that if feigning attempts were only successful at manipulating some subscales on the GSS-1 but not others, it might be possible to identify patterns that indicate ingenuine responding without the use of additional

assessment tools.

Methods of Feigning Research

Over the past 30 years, six studies have assessed whether instructed participants can effectively manipulate their GSS-1 scores. However, there is no consensus within the literature as to whether the GSS-1 is resistant to participant's attempts to appear manipulate suggestibility scores. It therefore becomes necessary to explore each of these studies better understand the conflictual findings.

The first study addressing this hypothesis conducted by Smith and Gudjonsson (1986), in which the GSS-1 with to two groups: one group was given the GSS-1 with the standard instructions, while the manipulation group was instructed to feign at a level "below their usual ability". The instructions given to the manipulation group were intentionally undetailed so that the participants had the ability to respond however they deemed appropriate. The two groups' scores were compared, and no significant differences on the Yield, Shift, or Total Suggestibility scores were seen. However, the instructed feigning group verbalized significantly fewer aspects of the GSS-1 narrative, as measured by Immediate Recall scores; no Delayed Recall responses were collected. The authors concluded that Immediate Recall is easy to manipulate, but that the remaining GSS-1 subscales are immune to feigning attempts because the true nature of the test is obscured from the test-taker. The authors reported that participants are unsure how to appropriately respond to the leading questions or negative feedback, leading to no consistent or successful feigning pattern.

Baxter and Bain (2002) called into question these findings, arguing that the instructions given to participants in Smith and Gudjonsson (1986) were too vague to be

generalize to real investigations in which individuals would be given the GSS-1, such as criminal evaluations. Baxter and Bain (2002) attested that when being assessed in forensic environments, respondents are more likely to have introductory information about the nature of the GSS-1 or of suggestibility in general. Further, they may have an understanding that “their best hope of having their previous testimony discounted is now to appear gullible and easily pressured” (p.220, Baxter & Bain, 2002). Building upon the instructions that were given in the initial study, Baxter and Bain conducted their own study in which they provided the following script to those they placed in the “faking-bad” manipulation group:

“The interviewer is going to interview you about the content of two stories you’ll hear. What I’d like you to do is to role-play being a suspect in a criminal investigation who may get off the hook if they can convince the interviewer that they are very gullible or very susceptible to pressure, that they are likely to accept whatever is said to them uncritically, and are therefore an unreliable witness. However, please do not go in for amateur dramatics or answer wildly. Try to concentrate on appearing gullible—but not stupid—by giving the interviewer whatever she seems to want.” (p. 221; Baxter and Bain, 2002)

Participants in the faking group were informed of the nature of the test and encouraged to use this knowledge to appear genuinely suggestible. The information about what the person should expect within the evaluation adds to the instructions to appear ingenuine in their performance was a more robust way of testing this hypothesis. By adapting these more detailed instructions, Baxter and Bain (2002) found that the instructions were sufficient for participants to significantly alter GSS-1 scores compared to controls.

Baxter and Bain (2002) and the research that have followed have adapted more detailed set of instructions when presenting the GSS-1 to their instructed feigning groups, with varying findings. Researchers have adapted the information given to the manipulation group (Woolston, Bain, & Baxter, 2006; Hanson, Smeets, & Jellicic, 2010; Baxter, Bain, Pringle, Fowler, & Tafili, 2013), and adding comparison groups that were aware of the purpose of the test, but not instructed to act upon that knowledge in a feigning matter (Boon, Gozna, & Hall, 2008). See Appendix A for full instructions presented between control and manipulation groups in previous research studies cited. When comparing these studies, there is mixed support that, when given specific prompting on how to respond to the test, it would be possible to manipulate at least some of the subscales on the GSS-1. However, these trends deserve specific exploration to understand how an instructed feigner may respond to the GSS-1.

Trends in GSS Scores

With support that it was possible to feign on the GSS-1, it became imperative to determine if feigned responding increases all subscales, a select number subscales, or a single subscale on the GSS-1. If there are specific trends that are unusual when compared to expected patterns of responding on the GSS-1, those discrepancies could then be utilized to help inform a decision regarding feigning of suggestibility and an external validity measure would not be necessary. Thus, trends reported will be discussed regarding the findings of those instructed to feign suggestibility as measured by the GSS-1 within the four studies, in hopes to generalize these findings to identify patterns that feigning responses form. The trends discussed have varying levels of empirical support, thus each trend will be discussed individually and then integrated as the body of literature

regarding feigning studies is fully explored. Of note, the studies presented did not consistently present the effect sizes (Cohen's d) within their publications. Therefore, the authors calculated effect sizes for the individual effects using the individual study's reported data and are presented alongside originally presented information

Yield and Shift score relation. Baxter and Bain (2002) were the first to report participants could manipulate GSS-1 following instruction. They found significant elevations in Yield 1 scores for the manipulation group that were not accompanied by an increase in Shift or Yield 2 scores when being compared to a control group. Baxter and Bain (2002) concluded that feigning attempts would increase some subscales of the GSS-1, but that feigned suggestibility could be assessed by examining if there were incongruences between Yield 1 scores and Shift scores. The authors hypothesized that participants who were instructed to feign believed that they should initially yield to leading questions, but would be unaware that truly suggestible individuals would change their responses following negative feedback, as measured by Shift.

Boon, Gozna, and Hall (2008), and Woolston, Bain and Baxter (2006), reported similar results, supporting the belief that instructed feigning participants would demonstrate elevated Yield 1 scores without corresponding elevations in Shift and Yield 2. However, these findings were brought into question by the results by Hanson, Smeets, and Jelicic (2010), who's results suggested differences only approaching significance ($p=.06$; Cohen's $d =.62$) on Yield 1. These studies together largely suggest that instructions to feigning could impact Yield 1 scores, but the instructions may not always be sufficient to elicit this response. This raises questions why some studies detected differences between groups, while others did not.

Table 2

Yield Score Comparison

	<i>n</i>	<i>F</i>	<i>Sig.</i>	<i>Cohen's d</i>
Baxter and Bain (2002)				
Yield 1	42	5.6	0.02	0.74
Yield 2	42	1.6	0.21	0.38
Woolston, Bain and Baxter (2006)				
Yield 1*	66	10.61	< 0.001	1.39
Yield 2**	66	10.44	< 0.001	1.25
Boon, Gozna, and Hall (2008)				
Yield***	61	30.69	< 0.001	1.80 (Yield 1) 1.15 (Yield 2)
Hansen, Smeets, and Jelcic (2010)				
Yield 1****	90	2.89	0.06	0.62
Yield 2****	90	3.83	< 0.05	0.58

* Post-Hoc Tukey HSD revealed significantly higher Yield 1 scores ($p < 0.001$) in their feigning group compared to their control

** Post-Hoc Tukey HSD revealed significantly higher Yield 2 scores ($p < 0.001$) in their feigning group compared to their control

***Boon, Gozna, and Hall (2008) reported Yield 1 and 2 aggregate scores between groups, with Post-Hoc Student Newman-Keuls revealing significantly lower scores ($\alpha < 0.05$) when comparing their manipulation group to their control group; effect sizes were reported for both Yield 1 and Yield 2

**** Post-Hoc Revealed non-significant differences between the control group and the manipulation group

Table 3

Shift Score Comparison

	<i>n</i>	<i>F</i>	<i>Sig.</i>	<i>Cohen's d</i>
Baxter and Bain (2002)				
Shift	42	0.04	0.85	0.04
Woolston, Bain and Baxter (2006)				
Shift *	66	N.R.	N.S./N.R.	0.20
Boon, Gozna, and Hall (2008)				
Shift**	61	N.R.	N.S./N.R.	0.40
Hansen, Smeets, and Jelcic (2010)				
Shift	90	1.49	0.23	0.33

*Post-Hoc Tukey HSD revealed no significant differences in Shift ($p > 0.05$) between their feigning group compared to their control

**Boon, Gozna, and Hall (2008) did not report non-significant F static findings

Total Suggestibility score trends. Some research suggests that Total Suggestibility would also be elevated following instruction. Woolston, Bain, & Baxter's (2006) were the first to suggest that instructions would impact Total Suggestibility by virtue of the instructions elevating Yield 1, and therefore an examination of Shift scores would be the best indication whether the individual had true vulnerabilities. Hansen, Smeets, and Jelcic's (2010) found that Total Suggestibility was the only score significantly impacted by feigning instructions. However, not all research supported these results, further highlighting the conflict between these proposed results. Baxter and Boon (2002) found no significant impact on Total Suggestibility scores following instruction, while Boon, Gozna, and Hall's (2008) suggested impact on both Yield 1 and Shift scores, but not Total Suggestibility. Associated statistics are shown in Table 4.

Table 4

Total Suggestibility Score Comparison

	<i>n</i>	<i>F</i>	<i>Sig.</i>	<i>Cohen's d</i>
Baxter and Bain (2002)				
Total Suggestibility	42	3.5	0.07	0.59
Woolston, Bain and Baxter (2006)				
Total Suggestibility *	66	6.87	<0.005	1.15
Boon, Gozna, and Hall (2008)				
Total Suggestibility	61	17.66	<0.001	0.75
Hansen, Smeets, and Jelcic (2010)				
Total Suggestibility**	90	3.33	<0.05	0.65

*Post-Hoc Tukey HSD revealed significantly higher Total Suggestibility scores ($p < 0.05$) in their feigning group compared to their control

**Post-Hoc Bonferroni analyses revealed significantly higher Total Suggestibility scores ($p < 0.05$) in their feigning group compared to their control

Trends regarding Free Recall scores. Smith and Gudjonsson (1986) were the first to indicate that there were no significant differences between their instructed feigning group and controls on any scale beyond Immediate Recall; their instructed

faking group had significantly lower recall scores as compared to the control. Smith and Gudjonsson (1986) argued that it was relatively easy to feign lowered recall scores, but that the feigning instructions were composed in a way that did not give them enough information about the test to adequately feign higher scores on other subscales of the GSS. Woolston, Bain and Baxter (2006) and Boon, Gozna, and Hall (2008) conducted studies with more complex instructions given to the participants, and their findings support this trend, suggesting that this decrease in free recall within the feigning group was an additional factor that could be considered when attempting to identify patterns in genuine responding pattern. However, both Baxter and Bain (2008) and Hansen, Smeets, and Jelicic (2010) found immediate recall scores were not significantly different impacted following instruction. Only Boon, Gozna, and Hall (2008) have reported the pattern of depressed Delayed Recall scores following instruction, as they are currently the only study to have administered the GSS-1 with the time delay. See Table 5.

Table 5

Free Recall Score Comparison

	<i>n</i>	<i>F</i>	<i>Sig.</i>	<i>Cohen's d</i>
Baxter and Bain (2002)				
Immediate Recall	42	1.1	0.30	0.31
Woolston, Bain and Baxter (2006)				
Immediate Recall *	66	5.99	<0.005	1.24
Boon, Gozna, and Hall (2008)				
Recall Aggregate**	61	6.21	<0.05	0.83 (Initial) 0.71 (Delayed)
Hansen, Smeets, and Jelicic (2010)				
Immediate Recall	90	0.33	0.71	$\eta_p^2=0.01$ ***

*Post-Hoc Tukey HSD revealed significantly lower Immediate Recall scores ($p<0.05$) in their feigning group compared to their control

** Recall analyses conducted together; Post-Hoc Student Newman-Keuls revealed significantly lower recall scores ($\alpha<0.05$)

*** As reported by authors; Authors did not provide Means and SD for Cohen's d calculation

Comparison to Vulnerable Populations

Although identifying elevated Yield 1 without corresponding Shift elevation is suggested by some researchers as a pattern to help identify patterns of manipulated response patterns, it is also important to consider the results of these studies outside of the experimental setting. Interestingly, the high Yield-1/low Shift trend is similar to the response pattern by those with genuine intellectual disabilities (Clare & Gudjonsson, 1993), raising the concern of how the findings of an instructed feigner would be interpreted if gathered in a real investigation.

Boon, Gozna, and Hall (2008) were the first to examine for differences in patterns of feigning-instructed participant's GSS-1 scores as compared to a genuinely vulnerable sample. Their results were surprising in that only the Immediate and Delayed Free Recall scores of their instructed feigning group were significantly different than published vulnerable/intellectually disabled individual norms, with their "Faking Bad" group participants having significantly higher recall scores; all other subscales were not significantly different. This is striking because it suggests that that not only could individuals adequately feign suggestibility, it was possible to do so in a pattern that was very similar to those with genuine disabilities. See Table 6.

These findings were challenged in part by the findings of Hansen, Smeets, and Jellicic (2010), who conducted similar comparisons between their sample and intellectually disabled norms. Their findings supported that instructed feigners obtained significantly higher recall scores, but significantly lower scores on the Yield 2, Shift, and Total Suggestibility subscales compared to truly vulnerable individuals. See Table 7. Interestingly, there was no significant differences on Yield 1 scores between the groups.

Table 6

Comparison to Vulnerable Populations' GSS-1 Scores, Boon, Gozna, & Hall, 2008

	<i>t</i>	<i>Sig.</i>	<i>Differences</i>
Immediate Recall	12.16	<0.001	Manipulation Group Higher
Delayed Recall	5.4	<0.001	Manipulation Group Higher
Yield 1	1.22	N.S.	None
Yield 2	1.67	N.S.	None
Shift	0.58	N.S.	None
Total Suggestibility	0.19	N.S.	None

Compared to published norms for "Intellectually Disabled" (I.Q. scores 57-75) scores presented in Gudjonsson (1997)

Table 7

Comparison to Vulnerable Populations' GSS-1 Scores, Hansen, Smeets, and Jelcic (2010)

	<i>t</i>	<i>Sig.</i>	<i>Differences</i>
Immediate Recall	>6.81	<0.01	Manipulation Group Higher
Delayed Recall	>6.81	<0.01	Manipulation Group Higher
Yield 1	-1.87	0.07	Not Significantly Different
Yield 2	>-3.70	<0.01	Manipulation Group Lower
Shift	>-3.70	<0.01	Manipulation Group Lower
Total Suggestibility	>-3.70	<0.01	Manipulation Group Lower

Compared to published norms for "Intellectually Disabled" (I.Q. scores 57-75) scores presented in Gudjonsson (1997)

The authors make conclusions similar to those made within the Smith and Gudjonsson's (1986) study: participants instructed to feign do not fully grasp the concepts of the GSS-1 and therefore they cannot effectively feign. Hansen, Smeets, and Jelcic (2010) thus concluded that the "GSS is relatively unaffected by attempts at faking heightened suggestibility" (p.227) but make no clear hypothesis why their findings differ from

previous findings.

Impact of Baseline Suggestibility

There is a variety of contrasting evidence for the trends in scores introduced primarily by the findings of Hansen, Smeets, and Jellicic (2010), which calls into question why such varied findings and interpretations were garnered from studies with very similar methodology. Baxter, Bain, Pringle, Fowler, and Tafili (2013) conducted a follow-up study to better understand the more recent conflictual findings, as the methods utilized by Hansen, Smeets, and Jellicic (2010) were methodologically similar to those utilized by Woolston, Bain, and Baxter (2006). Baxter et al. (2013) argued that the reason for the differences could be related to baseline interrogative suggestibility score in the samples used, which was not measured in either study. Baxter et al. (2013) gave 100 undergraduate students the GSS-2, followed instructions from Woolston et al. (2006), and then administered the GSS-1. Baseline suggestibility was determined by GSS-2 scores, and participants were separated between low, medium, and high suggestibility for statistical analysis to understand how the instructions impacted their GSS-1 results.

Their findings suggested that following instructions to become deceitful in their responses, how a person responded on the GSS-1 was moderated by their initial suggestibility levels. Participants who were partitioned into the low or medium suggestibility groups based on their GSS-2 scores were seen to have increased Yield 1 scores significantly ($p < 0.05$), where the high suggestibility group scores remained constant. Similarly, the low suggestibility group's Shift scores increased following instruction ($p < 0.01$), where the high suggestibility group scores significantly decreased ($p < 0.01$). The authors concluded that changes in GSS-1 scores following feigning

instructions depended on the mean scores of the baseline IS level of the instructed-feigning group, with lower baseline suggestibility individuals being able to successfully increase their scores and higher baseline individuals being able to suppress some subscale scores.

Summary of Findings and Literature Critique

Thus far, this paper has discussed the six studies found in the literature directly assessing the ability to feign suggestibility scores as measured by the GSS-1 to further foster the ability to differentiate between genuine and feigning responders. General trends presented have suggested that lower free recall scores, as well as high Yield-1 scores without corresponding Shift scores, could be used as markers for identification of a pattern of feigning as compared to control groups. However, more recent studies call into question these findings, which raises questions as to how to interpret the previously held consensus on how scores on the GSS-1 could be impacted by a motivation to feign suggestibility. Although Baxter et al. (2013) suggests that the conflict in the literature could be impacted by differences in baseline suggestibility levels, authors have proposed hypotheses to explain the discrepancies within the literature. The focus of this study will therefore turn to the limitations in the individual studies presented, as well as explore some of the overall methodological limitations still unaddressed within the field of feigned heightened suggestibility as measured by the GSS-1 to better understand the conflicting results and interpretations.

Administration Discrepancies

As previously outlined, the GSS-1 has very specific instructions that are formulated in a way to elicit genuine responses. Therefore, it is reasonable to argue that

by changing the administration format that there is also an impact on the pressures placed on the participant. This study argues that changing the administration of the assessment raises concerns about the validity of the individual findings relating to manipulating suggestibility scores, as well as the ability to compare the findings, as previous studies have not uniformly administered the GSS-1.

Apart from Boon, Gozna, and Hall (2008), all previously cited research utilizing the GSS-1 has foregone the 50-minute wait period between the Immediate Recall and the Delayed Recall in administration. After querying for Immediate Recall, these protocols moved directly to the questioning phase of the GSS-1. Although not explicitly addressed or explored in their articles, the elimination of the time delay raises two important questions: first, does the time delay impact the relationship between the interviewer and the participant? And second, does the change in protocol impact the ability to manipulate scores provided during the assessment? These questions require discussion in their own right.

According to Gudjonsson Suggestibility Scale manual (Gudjonsson, 1997), the 50-minute wait period is an integral part of the assessment, however there are three situations in which it would be acceptable to forgo: first, if the GSS-1 is being administered to an individual with low IQ or memory impairments such that they would not be expected to remember a large portion of the narrative after an extended period of time; second, in situations where participants are unable to verbalize a significant number of items in situations when prompted during the immediate free recall stage; or third, in the interest of time (Gudjonsson, 1997). Gudjonsson discusses that for individuals with memory impairments or learning disabilities the time delay can result in increased of

confabulation during the Delayed Recall phase and further negatively impact the results of the true suggestibility levels. However, the previous research has utilized primarily undergraduate students who are presumed to be of average intelligence based on their academic level, and therefore the removal of the wait period in the interest of shortening administration time. While minimally justified in the manual to remove the time delay, it states that the elimination of the time delay decreases the “difficulty” of the assessment in accordance to the manual according to the manual (Gudjonsson, 1997). This is not addressed by any of the literature to date.

Justification for shortening the administration length. Many of the feigning studies previously discussed cite the GSS manual (Gudjonsson 1997) as their justification for shortening their administration time. However, more recent studies have supplemented this justification with the findings presented in Smeets, Leppink, Jelicic, and Merckelbach (2009). Smeets et al. compared administration formats of the GSS-1 by comparing four different groups: a Standard GSS-1 Administration group, a No-delay/Immediate Recall group, a No Delay/No Recall group, and a 50-minute Delay/No Recall group. They found that the presence or absence of the 50-minute delay did not significantly change the participant’s scores on Yield 1, Yield 2, Shift, or Total Suggestibility compared to their control group. They also suggested that the removal of the free recall tests did not significantly impact these subscale scores, regardless of the delay time. For general research purposes, this increases the opportunity and utility of using the GSS-1; further, the authors suggest future studies should consider exploring this shortened administration within forensic populations to explore for significant differences.

While the findings of Smeets et al. (2009) could help shorten what is admittedly a lengthy administration time, they make no comment on how the adaptation of the GSS-1 protocol could impact feigning behavior and therefore their suggestions should be interpreted with caution when considered in the context of the feigning studies discussed. There are two ways this could change in a forensic setting or within the context of an instructed feigning study. First, the decreased interaction time may impact the relationship between the interviewer and the participant being administered the GSS-1. Because part of this assessment hinges on the relationship formed over time between the interviewer and interviewee, the 50-minute wait time, in conjunction with the recommendation suggestions to help form a positive bond during that time, may be significantly undermined and impact the effectiveness of both the leading questions and the negative feedback.

Second, the change in protocol does not weigh the cost of not having the free recall data available for analysis. While this may not appear initially as an important data point, it is important to consider in light of the findings of previous feigning studies that suggest that the layperson's GSS-1 scores following instructions to manipulate suggestibility were indistinguishable from intellectually disabled normative samples except for free recall data (Boon, Gozna, & Hall, 2008). Further, Hansen, Smeets, and Jelicic (2010) noted that their feigners scored below the normative means on all subscales except for free recall scores, for which they scored higher. Within these studies, the free recall scores are suggested to be an important indicator when assessing attempts to manipulate suggestibility scores, and without collecting these scores other researchers can only make limited conclusions regarding patterns of feigning following instruction.

Manipulating Cognitive Load

Beyond the factors decreasing the believability of the GSS-1 administration and decreasing the data points available, there are additional implications on the cognitive demands that are placed upon the person when being administered a shortened version the GSS-1. Gudjonsson (1997) states that the shorter administration time makes the assessment less difficult, as there are less cognitive demands between the initial presentation of the narrative and the presentation of the Yield 1 questions. It can be argued that decreasing the difficulty of an assessment as such frees cognitive resources that allow for an increased ability to alter verbal behavior when administered the GSS-1.

Drake, Lipka, Smith, & Egan (2013) addressed this question, hypothesizing that increasing cognitive load during the GSS-1 would impair the ability to feign higher suggestibility scores. They attest that because the instruction to feign is minimally demanding on working memory, it enables “effective comprehension and critical evaluation of the information in the GSS questions in light of the story” (p. 848, Drake et al. 2013). They believe that the instructions for the GSS-1 alone are not cognitively demanding enough for the test to be sufficiently difficult and therefore participants are able to process the information more easily and respond accordingly, which increases their ability to feign when responding.

Drake et al. (2013) employed a two by two design where participants were given the faking instructions outlined in Hansen, Smeets, and Jelicic (2010). Additionally, half of the participants were given a concurrent Digit Span test to increase their cognitive load during the GSS-1 questioning phase to assess the impact of additional cognitive load on the ability to feign; this was compared to a control group similarly split. Comparisons

between instructed feigners and controls supported their hypothesis that Yield 1 scores were the only scale significantly elevated within the faking-bad conditions. However, when the feigning group was given a concurrent cognitive task during GSS-1 administration, Yield 1 scores fell significantly. Although Yield 1 was still significantly higher than the control condition for feigners assigned to the digit span task, the authors argued that additional increase of cognitive load could make the assessment more resistant to feigning attempts. It should be noted that the researchers also chose to forgo the time delay as well.

Table 8:

Cognitive Load Comparisons

	F	Sig.	<i>d</i>
Faking/Concurrent Task X Faking/No Concurrent Task			
Yield 1	28.3 (1,38)	$p < 0.001$	0.098
Faking/No Concurrent Task X Genuine/No Concurrent Task			
Yield 1	36.8 (1,38)	$p < 0.001$	0.152
Yield 2	N.R.	$p = 0.001$	0.024
Total Suggestibility	N.R.	$p < 0.001$	0.052

Drake et al (2013; n=80)

Considering Drake et al.'s (2013) findings that altering the cognitive load is correlated with the ability to manipulate GSS-1 scores, it can be concluded that altering the GSS-1 administration by removing the delay and inherent distraction task could have a unique effect on individuals motivated to manipulate their scores. Drake et al. (2013) support that decreasing the cognitive load and increasing their ability to focus on the task at hand effectively inflates individual's baseline ability to feign. If increasing the cognitive load is empirically supported, it would be logical that utilizing methods that are

sanctioned by the assessment would be preferable to adding superfluous tasks. The goal of this research should first and foremost be to recreate conditions that are most generalizable to scenarios where the GSS-1 would be administered so the findings regarding feigning trends could be applicable. Drake et al.'s (2013) methods further reinforces that by dispensing of the wait time in research, researchers have inadvertently allowed the participants more cognitive freedom to act in accordance with the feigning instructions. However, the ability to actively participant with the manipulation and participating with the manipulation is a different matter that requires exploration.

Additional Administration Differences

Beyond the elimination of the 50-minute wait period, the studies also varied in the way they collected the data from the GSS-1. Woolston et al. (2010) is markedly different from the other studies cited because it relied on Free Recall scores to be written out by participants rather than verbally stated. The authors note that their Free Recall scores were statistically lower than the normative means, and do not discuss the motivation for this change in protocol. Although verbal responses being marked by a research assistant as correct or incorrect at the time of the response could lead to inadequate reporting, this is much easier to correct by recording the verbal responses and double-checking participant responses later rather than forcing the participants to write responses, which may limit their willingness to put forth their best effort. Further, writing responses for the GSS-1 is less likely to be utilized when giving the GSS-1 in a forensic setting. Subtle differences such as this in the administration may not cause a failure in a study, but the more minor differences arise, the more apparent the inability to generalize results becomes.

Although seemingly inconsequential on its surface, it is imperative to highlight that minor changes in administration such as changing the way data is collected may have larger consequences on results. This is even more true when there is a significant amount of these minor changes, which build upon one another exponentially to make comparison and generalizability increasingly difficult.

Undergraduate Sample Comparisons

Although there is a lack of uniformity within GSS-1 protocol administration within the feigning literature, it is important to note that all studies have primarily utilized undergraduate participants for their samples. Students in all the studies have been enticed to participate in research studies in exchange for course credit or gift cards. Often the literature ignore the impact the sample or recruitment method has on their results. While use of undergraduate research participants is not in and of itself an issue, there are some limitations that should be considered when interpreting results and generalizing conclusions to other populations. Namely, it is important to question how likely undergraduate student participants conform to manipulation instruction with enough gusto to equate how an individual would respond on the GSS-1 in a criminal justice setting if they were motivated by their potential freedom.

Baxter et al. (2013) noted that due to the population and procedure they utilized, the participants may have had no reason to engage in the manipulation when given instructions to manipulate suggestibility scores. They believed this to be a significant limitation in their study, and suggested that “Future work might offer all participants an incentive, such as a cash reward, for each false implication which they detect and accept, which may better assess their full capacity to detect discrepancies.” (P. 923, Baxter et al.,

2013). While they do not discuss their suggestions to improve methods for recruiting participants or how to address the conformity problem, this was the first article to note that the motivation of the participants in the manipulation group may be a concern.

It can be assumed that all undergraduate participants who participate for course credit or monetary compensation may be impacted by this limitation. Although the feigning instructions utilized in the previous literature may be similar to instructions that could be given to someone in a criminal investigation hoping to feign suggestibility, the motivation to follow these instructions are vastly different. Students may find it difficult to subsume the internal motivation inherently experienced by someone who may have criminal responsibility negated if they respond in a certain way on these assessments. To date, none of the studies directly addressed this lack of motivation, and have relied on only the initial motivation to participate in the study as sufficient motivation to conform to the group conditions.

Intelligence

In addition to motivation levels, considerations must be made regarding intellectual differences between the undergraduate sample and the forensic population in which the GSS would likely be used. Gudjonsson and Clark (1986) suggests that intelligence does not have an impact on suggestibility when the participant's IQ scores are within the average and above average range. This limits the expected impact that undergraduate participant's intelligence would have on suggestibility scores for within-group analyses, but dampens generalizability of findings to forensic populations. Given that that the average IQ within the criminal justice system is lower than that within the general population, this is an important consideration when attempting to generalize

findings.

Beail (2002) argued that the GSS does not properly account for intellectual disabilities within its methodology due to its reliance on memory, which the literature suggests is negatively correlated with suggestibility. Beail stated that memory plays a large role when administering the GSS-1 with the 50-minute delay because participants with intellectual disabilities have difficulty remembering the narrative presented, and will thus answer leading questions in an acquiescent manner. To better understand the difficulty posed by low intelligence, Sondena, Rasmussen, Palmstierna, and Nottestad, (2010) conducted a study to assess if the GSS-1 could be used with inmates in a Norwegian prison. Rates of intellectual disabilities were unknown, and the participants (n=133) were given the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) and the GSS-1. The authors assessed the correlation between GSS-1 scores with IQ, finding that Full Scale IQ and GSS-1 scores were significantly negatively correlated for all GSS-1 subscale except Immediate recall, which had a strong positive correlation with Full Scale IQ (see Table 8). Further, the authors identified one in ten inmates had Intelligence Quotient (IQ) scores less than 70, indicating an intellectual disability. Their sample then was divided based on Full Scale IQ scores above and below 70 and compared, revealing significantly lower GSS-1 Total Suggestibility Scores for the participants with IQ less than 70.

Table 9:

Correlation of WASI IQ with GSS and GCS

	WASI Full Scale IQ	WASI Verbal IQ	WASI Performance IQ
GSS (n=113)			
Immediate Recall	0.544	0.540	0.411
Yield 1	-0.263	-0.229	-0.223
Yield 2	-0.259	-0.237	-0.216
Shift	-0.257	-0.237	-0.233
Total Suggestibility	-0.321	-0.281	-0.281

Sondenaa, Rasmussen, Palmstierna, & Nottestad, 2010; All correlations are significant ($p < 0.05$)

Sondenaa et al (2010) conclude that intelligence had a significant impact on suggestibility scores as measured by the GSS-1, and suggest further research should explore use of GSS-1 in forensic populations with respect to individuals with intellectual disabilities. Practitioners have acknowledged the impact that cognitive deficits could have on GSS-1 responses and have included intelligence assessments to ensure the GSS-1 is administered to appropriate populations (Frumkin, Lally, & Sexton, 2012). Due to the reliance on undergraduate students as the primary source of participants, there has been limited opportunity to assess how participants with varied intelligence quotients would respond to instructions to feign heightened suggestibility levels. As feigning research expands to include samples outside of the undergraduate population, intelligence assessments should be increasingly evaluated in conjunction within the GSS-1.

Additional Assessment of Valid Responding Pattern

The current feigning literature has focused on identifying patterns that could suggest feigned suggestibility on the GSS-1 by comparing instructed malingerers to control groups using only GSS-1 scores, which is imperative because the assessment does not contain inherent validity indicators. It is interesting to note that no study has

connected the GSS with other assessment tools, and there is no mention if its use with validated malingering measurements. This further demonstrates the divide between the experimental and practical applications, as the GSS-1 is not traditionally given in isolation. Frumkin, Lally, and Sexton (2012) state that the GSS-1 should be used with other assessments, as well as within a clinical interview to better evaluate the totality of the circumstances regarding the defendant and the validity of a confession. This decreases the reliance on a single assessment and gives a clearer clinical picture when formulating a clinical report.

In real-world evaluations, is not likely that the GSS-1 would be the only assessment an evaluator would administer, as personality, intelligence, and behavioral assessments could also glean beneficial data when completing a psychological evaluation. If a person is inclined to feign suggestibility during a more comprehensive evaluation, they would likely respond to a variety of the assessments in a consistent manner. As previously discussed, Gudjonsson and Sigurdsson (2004) suggest the GSS-1 could use a separate validity indicator to understand the manner in which the individual approached the test. Taken together, it is surprising to note that no research thus far has raised this point or studied the GSS-1 in the context of other assessment tools. This is a large area that has yet to be explored, and if Baxter et al. (2013) findings are taken to be true then it could be possible that only looking at the GSS-1 responses could be too narrow to assess attempts to feign heightened suggestibility levels.

In conclusion, research focusing on instructed feigning requires a more in-depth protocol to increase the similarities to data points that would be collected in the normal course of a forensic evaluation that utilized the GSS-1.

Purpose of the Study

In an effort to bolster the empirical support for the GSS-1 and to continue to submit psychological testimony, further research is necessary to understand responding patterns when a participant intends to distort their verbal behavior on the GSS-1. The purpose of the current study is to evaluate participant's ability to manipulate their suggestibility scores as measured by the GSS-1, as well as to explore assessment tools that could help the psychologist delineate genuine and ingenuine reporting. The current study adds to the field by addressing multiple areas of concern addressed within the literature; first, by increasing the participant's motivation to follow feigning instructions by increasing incentives to reward "successful" feigning of suggestibility. This intends to foster higher motivational levels in the participants utilized that more realistically mimics those experienced by individuals within a forensic psychological evaluation as compared to the current samples utilized within the literature, and allows for more generalizable results.

Second, the current study will include the time delay within the GSS-1, during which participants will be given similar distractor tasks that they could come to expect were they given the GSS-1 during a forensic evaluation supporting a false confession claim. Third, the tasks administered alongside the GSS-1 will be analyzed. Previous research has not attempted to analyze GSS-1 scores in relation to additional psychological measures, which does not accurately reflect the practical way in which the GSS-1 would be administered. For this study, the Personality Assessment Inventory (PAI) will be administered, which includes multiple clinical and validity scales that would be useful to forensic psychologists attempting to formulate a clinical opinion. This could potentially be beneficial in understanding response patterns by assessing the impact of feigning

instruction on validity scales embedded in the PAI and similar tests administered in an assessment battery.

Next, with these adaptations in mind, the current GSS-1 findings will be compared to the means presented by Boon, Gozna, and Hall (2008) due to their study being the first to report the inclusion of the 50-minute wait period when assessing the ability to manipulate GSS-1 scores. Further, the means from the current study will be compared to Gudjonsson's normative sample (1997), Gudjonsson's intellectual disability sample (1997), as well as Frumkin, Lally, and Sexton's US forensic sample (2012) to understand what changes, if any, are able to be identified. These comparisons, as well as limitation in their comparisons, will be discussed.

Finally, by administering an additional memory malingering assessment, it is hypothesized that comparisons between participants' suggestibility scores and performance efforts could be made. Information from between group comparisons could provide valuable insight into how the instructions to feign heightened suggestibility are interpreted and their subsequent impact on responding patterns outside of the GSS-1. Additionally, information from this comparison could be used to further inform suggestions for assessments to consider when administering forensic batteries in the future in false confession cases.

Research Hypotheses

The present research aims to investigate the ability of the GSS-1 to differentiate between individuals approaching the test in a valid manner and those who are motivated to feign increased suggestibility. After a thorough literature review, the following hypothesis were derived:

Hypothesis 1: Gudjonsson Suggestibility Scale Score Differences

It is predicted that there would be a significant difference in Gudjonsson Suggestibility Scale (GSS-1) scores between the control group and the manipulation group. Based on previous research, specific hypotheses were made based on the individual subscale scores within the GSS-1:

Hypothesis 1.1. It is predicted that the GSS-1 Free Recall-Immediate scores for the manipulation group will be significantly lower than the control group Free Recall-Immediate scores.

Hypothesis 1.2. It is predicted that the GSS-1 Free Recall-Delayed scores for the manipulation group will be significantly lower than the control group Free Recall-Delayed scores.

Hypothesis 1.3. It is predicted that the GSS-1 Yield 1 scores for the manipulation group will be significantly lower than the control group Yield 1 scores.

Hypothesis 1.4. It is predicted that the GSS-1 Yield 2 scores for the manipulation group will not significantly differ between groups.

Hypothesis 1.5. It is predicted that the GSS-1 Shift scores will not significantly differ between groups.

Hypothesis 1.6. It is predicted that the GSS-1 Total Suggestibility scores for the manipulation group will be significantly higher than the control group Total Suggestibility scores.

Hypothesis 2: Personality Assessment Inventory Validity Scale Differences

It is predicted that there would be a significant difference in Personality Assessment Inventory (PAI) Validity Scale scores between the control group and the manipulation group. Specific hypotheses were made based on the individual Validity Scales within the PAI:

Hypothesis 2.1. It is predicted that the PAI Inconsistency Scale (ICN) scores for the manipulation group will be significantly higher than the control group.

Hypothesis 2.2. It is predicted that the PAI Infrequency Scale (INF) scores for the manipulation group will be significantly higher than the control group.

Hypothesis 2.3. It is predicted that the PAI Negative Impression Management Scale (NIM) scores for the manipulation group will be significantly higher than the control group.

Hypothesis 2.4. It is predicted that the PAI Positive Impression Management Scale (PIM) scores for the manipulation group will be significantly higher than the control group.

Hypothesis 3: Test of Memory Malingering Scale Differences

It is predicted that there would be a significant difference in Test of Memory Malingering (TOMM) scores between the control group and the manipulation group. Specific hypotheses were made based on the individual TOMM Scales:

Hypothesis 3.1. It is predicted that the TOMM Trial 1 scores for the manipulation group will be significantly lower than the control group TOMM Trial 1 scores.

Hypothesis 3.2. It is predicted that the Trial 2 scores for the manipulation group will be significantly lower than the control group TOMM Trial 2 scores.

Hypothesis 3.3. It is predicted that the TOMM Retention Trial scores for the manipulation group will be significantly lower than the control group TOMM Retention Trial scores.

Hypothesis 4: Gudjonsson Compliance Scale Score Differences

It is predicted that the Gudjonsson Compliance Scale (GCS) scores for the manipulation group will be significantly higher than the control group GCS scores.

Exploratory Analysis: Gudjonsson Suggestibility Scale Mean Comparisons

GSS-1 scores collected will be compared to four different samples to understand how representative they are of the current literature base's assertion. The first comparison will be between current findings and those presented by Boon, Gozna, and Hall (2008); this will help understand the impact of the increased incentive and specified distractor task for the manipulation group. It is predicted that the manipulation group GSS-1 scores will not significantly differ from mean response scores on Immediate Free Recall, Yield 1, Yield 2, Shift, Total Suggestibility those seen in the "Faking Bad" group from Boon, Gozna, and Hall (2008). Second, the means for the control group will be compared to mean scores presented by Gudjonsson (1997) for their normative sample to assess for differences in US and UK samples in line with those suggested by Frumkin, Lally, & Sexton (2012). Third, in line with previous research, the means for the manipulation group will be compared to mean scores presented by Gudjonsson (1997) for their intellectually disabled sample to assess whether instructed feigner's scores significantly differ on any GSS-1 subscales. Finally, a comparison to Frumkin, Lally, and Sexton's

(2012) proposed U.S. forensic population will be conducted. These explorations intend to further understand the implications of the current study's findings in the context of the current literature.

CHAPTER III

Methodology

Participants

The sample for the current study consisted of 32 undergraduate students recruited through flyers posted throughout a local college campus. Participants were offered a \$20 gift card as an incentive for participating in the study. Researchers excluded potential participants who were under the age of 18, as well as any individuals that could not read and write in English fluently. Participant age ranged from 18-30 ($M = 20.06$, $SD = 2.21$). Eight participants were male, and twenty-four participants were female. Participants identified their race/ethnicity as followed: Black, 39.4%, ($n=13$); Caucasian, 24.2% ($n=8$); Hispanic, 18.2% ($n=6$), Asian, 15.2% ($n=5$), and 12.5% ($n=4$) participants identified their race as Mixed. No participants failed to complete the study once they consented to participate.

Protection of Human Participants

The study was approved by the Institutional Review Board (IRB) of Nova Southeastern University. This study was determined to have minimal risk to the participants and all participants were given an informed consent form prior to testing. Due to the use of deception within the study embedded within the Gudjonsson Suggestibility Scale, all participants were debriefed regarding the use of deception and given the opportunity to remove their data if they wished without losing any recruitment incentive. No participant chose to withdraw their data from the study when given this opportunity. This debriefing also included providing participants with the contact information for both the research team and student counseling if they had further

questions regarding the study or if they wanted to discuss thoughts or emotions brought about through their participation.

Measures

Gudjonsson Suggestibility Scale-1 (GSS-1). The Gudjonsson Suggestibility Scale (Gudjonsson, 1997) is an assessment tool which assists in measuring an individual's susceptibility to interrogative pressures during a simulated investigative interview. The assessment is presented as a memory test by the evaluator to disguise the nature of the assessment, which includes asking the participant questions about their subjective accounts of their memory. The participants are then presented a narrative of a factitious robbery. Immediately following hearing the narrative, participants are asked to recall everything they are able. Participants are then presented an unrelated distractor task that lasted up to fifty minutes, after which they are asked to recall everything they can regarding the narrative once again. Participant are then asked twenty questions regarding the content of the narrative; of these questions, fifteen questions are considered leading questions, as they ask about information that is not present in the narrative and are presented in ways in which convey an expected answer. Regardless of how the participants respond, after responding to all of the questions they are told that the following verbatim: "You have made a number of errors. It is therefore necessary to go through the questions once more, and this time try to be more accurate." Participant are asked the same twenty questions again. All responses are suggested to be recorded using audio-tape for a secondary check of scoring. Scoring of the GSS-1 is "nondiscretionary" (Frumkin, Lally, & Sexton, 2012), and was completed following the instructions outlined in the administration manual (Gudjonsson, 1997).

The GSS-1 Manual (Gudjonsson, 1997) has collected normative data for a variety of populations from the United Kingdom, including for adults in the general population (see Table 10), adult and juvenile criminal offenders, and adults with intellectual disabilities (see Table 11). To assess test-retest reliability, Gudjonsson (1987) compared scores obtained on the GSS-1 and the GSS-2, an alternative form that follows the same administrative format of the GSS-1 but changes the narrative presented to the participant and adjusts the questions accordingly. He found that test-retest correlations on the scales ranged from 0.73-0.90 for individuals within a general population (n=28; see Table 12). While no current normative sample exists from the United States, Frumkin (2008) suggests that conservative interpretation of the scores as compared to the UK norms is appropriate.

Table 10

GSS Normative Sample, Gudjonsson (1997)

	<i>N</i>	<i>Mean</i>	<i>S.D.</i>	<i>Range</i>
Immediate Recall	157	21.3	7.1	4-36
Delayed Recall	135	19.5	7.5	4-34.5
Yield 1	157	4.6	3.0	0-13
Yield 2	157	5.6	3.8	0-15
Shift	157	2.9	2.5	0-12
Total Suggestibility	157	7.5	4.6	0-21

Table 11

GSS Norms for Individuals with Intellectual Disabilities, Gudjonsson (1997)

	<i>N</i>	<i>Mean</i>	<i>S.D.</i>	<i>Range</i>
Immediate Recall	68	8.1	4.9	0-24
Delayed Recall	68	6.0	4.9	0-23.5
Yield 1	68	9.8	3.5	0-15
Yield 2	68	9.4	3.5	0-15
Shift	68	4.8	3.0	0-13
Total Suggestibility	68	14.6	4.6	0-24

Table 12

GSS Test-Retest Correlation, Gudjonsson (1987)

<i>Immediate Recall</i>	<i>Delayed Recall</i>	<i>Yield 1</i>	<i>Yield 2</i>	<i>Shift</i>	<i>Total Suggestibility</i>
0.77	0.73	0.84	0.86	0.79	0.90

Test of Memory Malinger. The Test of Memory Malinger (TOMM; Tombaugh, 1996) is a 50-item forced-choice recognition assessment that aims to differentiate individuals with genuine memory disturbances and those who are feigning memory impairment. Subjects are informed they are being assessed on their ability to learn and remember pictures of common objects. Participants are shown 50 items, seeing each item for three second. During the subsequent two trials, the subject is shown 50 pages, each of which contains one previously seen item and a new picture. Participants are instructed to identify the item they have previously seen. Between Administration 1 and 2, the participant is given a second learning trial of the initial items. Following the second trial, a retention trial can be administered. This retention trial is administered after a 15 minute wait period, and does not include a learning trial. When administering the TOMM, both positive and negative feedback is given by the researcher following each response.

The TOMM utilizes a high number of visual stimuli that make the assessment appear more difficult than it truly is, and therefore participants who are attempting to perform poorly within a battery of tests may modify their responses on the TOMM (Tombaugh, 1996). In addition, implementing feedback for correct and incorrect responses increases performance for genuine-performing participants. The TOMM (Tombaugh, 1997). was validated on a sample of 158 participants which included

participants from a Veterans Administration Medical Center and participants who had head-injuries who had participated in previous research. Participants were divided into five groups: No Cognitive Impairment (n=13), Cognitive Impairment (n=42), Aphasia (n=21), Traumatic Brain Injury (TBI; n=45), or Dementia group (n=37). Results indicated that the Dementia group scored significantly lower than the other groups on Trial 1, Trial 2, and on the Retention Trial. Tombaugh (1997) concluded that a cutoff score of 45 could correctly identify the majority of participants as not malingering (95% correct classification rate), with decreased discernibility for only the Dementia group (91%). Similar comparisons have utilized the 45 cutoffs with results indicating 100% specificity and as high as 93% sensitivity rating (Tombaugh, 1997). Therefore, scoring less than 45 on trial 2 or the retention trial “raise[s] concern that the individual is not putting forth maximum effort and is likely malingering” (p.2, Tombaugh, 1996). Evaluations of internal consistency of the TOMM has found high coefficient alpha (.94-.95; Spreen & Strauss, 1998).

Personality Assessment Inventory (PAI). The Personality Assessment Inventory (PAI) is a 344-item self-report inventory that was developed to adult psychopathology (Morey, 2007). The PAI consists of 22 scales, including four validity scales, eleven clinical scales, five treatment scales, and two interpersonal scales (Morey, 2007). The PAI is used to aide in screening, diagnosing, and planning treatment for psychopathology. The PAI has advantages over similar personality measures such as a low reading level required for administration (4th grade reading level), scaled responses compared to force-choice responses, and fewer overall test questions. Item response utilizes a four-point Likert-Scale: False, Slightly True, Mainly True, and Very True

(Morey, 2007). Responses are analyzed and each scale is reported as a t-scores that range from 20-110; mean scores fall between a t- Score of 50 and 70, with scores outside of that range indicating significant differences (Morey, 2007). Reported good internal consistency for college samples, α ranges .66 to .94, with an internal consistency of .82 for college samples; test retest reliability ranges from .60-.94 (Morey, 2007).

Of significant importance to the current research are the validity scales embedded into the PAI questions, which evaluate reporting styles to ensure an interpretable report. The Inconsistency Scale (ICN) consists of ten paired items that are highly correlated to assess consistent responding throughout the assessment, whereas the Infrequency (INF) scale utilizes eight questions that assesses for careless responding. Negative Impression Management (NIM) focuses on identifying exaggerated endorsement of symptoms, whereas the Positive Impression Management (PIM) assesses for responding to show oneself in an overly favorable light. These scales can be utilized to either invalidate PAI scores or can be utilized to alter interpretations of clinical scale scores considering responding patterns (Morey, 2007).

Gudjonsson Compliance Scale. The Gudjonsson Compliance Scale (GCS; Gudjonsson, 1989) is a twenty question self-report measure of compliance. Compliance differs from suggestibility in that, while suggestibility infers an internal acceptance of actions or statements being made as being true, compliance merely indicates that “the person makes a conscious decision to carry out the behavior proposed or requested, even if he or she privately does not agree with it” (p.5, Gudjonsson, 1997). This construct helps researchers distinguish between the level in which participants may simply comply with instructions presented rather than internalize the belief in their responses. The has

modest statistical support, with an alpha coefficient of 0.71. Research suggests the GCS loads onto three different factors: avoidance of confrontation/conflict, eagerness to please, and compliant behaviors (Gudjonsson, 2007). The test-retest reliability for the GCS is modest ($r=0.88$ $p<0.001$; Gudjonsson, 1989). The GCS mean normative scores for UK university students is 7.8 (S.D.= 4.1; Gudjonsson, 1989), which will be used for comparison for the purposes of this study.

Demographic Questionnaire. A brief demographics form was utilized in this study to collect participant's age, gender, race/ethnicity, and educational level. See Appendix D for the Demographics Questionnaire presented.

Procedure

This study used a between-group design. Participants were randomly assigned to two conditions, with each group containing sixteen participants. During the recruitment, all participants were informed that they would be taking multiple assessments as part of the research (see Appendix B for the recruitment flyer). If prompted, participants were told by recruiters they would spend up to two hours taking both verbal and pen-and-paper assessments. Participants were scheduled to meet with research assistants based on communal availability. Assessment administration were conducted in private rooms within a local public library. All research assistants were doctoral psychology candidates trained in clinical assessment administration. In preparation for this this study, assistants were familiarized with the assessment tools and observed in administration of the assessments prior to data collection to ensure consistent administrations of the GSS-1 negative feedback. Additionally, administrations were reviewed to provide feedback to ensure consistent administration over time and research assistants.

When a participant first met with a research assistant, they were given an informed consent that was reviewed collaboratively (See Appendix C). The consent informed participants that they would be taking a series of written and verbal psychological assessments to better understand their connection to one another. Upon signing the consent, participants were presented with an envelope that contained one of two instructions; these instructions varied by condition and established comparison groups. Participants in the control group received the following instructions, which has a Flesch-Kincaid reading level of 7.0:

“You are going to be given multiple assessments today. Please complete the assessments as you normally would. You should attempt to complete the assessments to the best of your ability. The interviewer is unaware of the instructions that you have been given, so please do not let them know what you have read here. When you have finished reading this letter, please fold it and return it to the envelope.”

Participants in the manipulation group were given a different set of instructions that has a Flesch-Kincaid reading level of 9.6, which read as follows:

“You are going to be given multiple assessments today. You should attempt to play the role of someone who is a suspect in a criminal investigation. You have the chance of getting off the hook if you can convince the interviewer that you are easily influenced by pressure from other people. You should try to go along with everything the interviewer says to you, as though you accept everything that is said to you without question. However, please do not go in for amateur dramatics or answer wildly. Try to concentrate on appearing gullible—but not stupid—by

giving the interviewer whatever they seem to want. If you are determined to be the most credible faker in the research study, you will be compensated with a \$50 Publix Gift card. The interviewer is unaware of the instructions that you have been given, so please do not let them know what you have read here. When you have finished reading this letter, please fold it and return it to the envelope.”

Research assistants were unaware of the instructions or the intent of the group differentiation. This was debriefed to all research assistants following completion of data collection. Following the presentation of the group instructions, participants were evaluated using the four clinical assessment tools discussed and a demographics questionnaire. Research assistants recorded verbal responses with an audio recorder for subsequent review of scoring.

Administration began with the beginning of the GSS-1. The research assistant informed the participant that they were being presented with a memory test, and asked them to rate their memory and explain their rationale. The participants were then played a standardized audio recording of the GSS-1 narrative. When the recording was completed, participants were asked to recall as many aspects of the story as they could without encouragement. The GSS-1 then calls for a 50-minute wait period, thus participants were administered the PAI as they waited. If the participant completed the PAI before the wait period was completed, the GSS-1 was resumed; no participant took the full wait period to complete the PAI.

When the GSS-1 was resumed, the participants were asked to again recall as many aspects of the narrative as they could. When the participant is unable to recall additional information, they were presented with 20 forced-choice questions about the

narrative. After the participant had answered the questions, they were told the following verbatim: "You have made a number of errors. It is therefore necessary to go through the questions once more, and this time try to be more accurate" (Gudjonsson, 1997).

Participants were then asked the same 20 forced-choice questions. This concluded the GSS-1 administration. Participants were then administered the first and second trials of the TOMM with standard instruction. After the TOMM administration, participants were administered the GCS and the Demographics Questionnaire (see Appendix D).

Following completion and submission of all assessment responses participants were debriefed about their participation, with emphasis on the standardized negative feedback regarding their accuracy when responding to the GSS-1 questions. After being informed this response were scripted and not related to actual responses, participants were informed that they could withdraw their data without penalty due to the deception involved. Notably, no participant elected to withdraw their data following being informed of the study. Participants were given contact information for the lead researcher, as well as the counseling center contact information. Participants were then given the \$20 incentive.

Statistical Analysis

A series of analyses were conducted for this study to address the five main hypotheses. A series of One-Way Analysis of Variance (ANOVA) models was performed to evaluate each of the hypotheses for the effect of instruction (Control vs. Manipulation groups) on GSS-1 six subscale scores (Yield 1, Yield 2, Shift, Total Suggestibility, Immediate Recall, Delayed Recall), PAI's validity scales (Inconsistency, Infrequency, Negative Impression management, and Positive Impression Management scales), TOMM

Trial Scores (Trials 1 and 2) and the GCS scores. Additional analyses were conducted when assumptions failed that did not allow ANOVAs to be conducted. Finally, a series of *t*-tests were completed to compare GSS scores to previous research findings. In Chapter's 4 and 5, the results of these analyses will be discussed.

CHAPTER IV

Results

This study was a comparison between participants instructed to feign heightened suggestibility levels ($n = 16$) and a control group ($n = 16$) on the following assessments: Gudjonsson Suggestibility Scale (GSS-1), Personality Assessment Inventory (PAI), Test of Memory Malingering (TOMM), and Gudjonsson Compliance Scale (GCS). In this chapter, results of a series of one-way Analyses of Variance (ANOVAs) are presented. Additionally, t-tests were conducted to enlighten how these results compared to previous feigning research and published GSS-1 samples. Analyses were conducted utilizing SPSS 20.0 (SPSS Inc., Chicago, IL). Given the study's small sample size, effect size (Cohen's d) was utilized to supplement decisions regarding statistical significance. Although some hypotheses did not reach statistical significance, interpretations of results were supplemented by the presence or absence of a meaningful effect size. For the current study, Cohen's (1992) guidelines for effect size interpretation were utilized, defining $|d|$ equal to .2 as a small effect, d equal to .5 as a medium effect, and a d equal to .8 as a large effect.

Hypothesis 1

Hypothesis 1 aimed to understand differences group assignment would have on GSS-1 scores. Measures of central tendency and variability (Mean, Standard Deviation, and Range) are presented in Table 13. Tables 14 and 15 present results of Levene's Test of Equality of Error Variances and Shapiro-Wilk Test of Normality, respectively. Results of the series of ANOVAs conducted are presented in Table 16. Results for individual hypotheses are discussed below.

Table 13

Central Tendency and Variability for the GSS-1

	<i>Control</i>			<i>Manipulation</i>		
	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
Immediate Recall	20.531	5.143	13.5-31.5	15.969	4.436	7.5-24.5
Delayed Recall	19.781	4.764	13.0-32.0	15.094	4.083	7.0-25.5
Yield 1	5.25	2.113	1-9	5.34	2.241	2-10
Yield 2	6.25	2.696	2-10	7.47	2.907	3-14
Shift	4.69	3.610	0-10	4.56	2.707	1-9
Total Suggestibility	10.00	4.913	2-16	9.938	4.106	3-18

Table 14

Levene's test for the GSS-1

	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
Immediate Recall	.866	1	30	.359
Delayed Recall	.595	1	30	.446
Yield 1	.040	1	30	.843
Yield 2	.023	1	30	.881
Shift	.235	1	30	.475
Total Suggestibility	.110	1	30	.642

Table 15

Shapiro-Wilk Test of Normality for the GSS-1

	<i>Control</i>			<i>Manipulation</i>		
	<i>Statistic</i>	<i>df</i>	<i>p</i>	<i>Statistic</i>	<i>df</i>	<i>p</i>
Immediate Recall	.951	16	.501	.982	16	.975
Delayed Recall	.917	16	.148	.925	16	.200
Yield 1	.974	16	.898	.958	16	.633
Yield 2	.961	16	.678	.957	16	.606
Shift	.896	16	.071	.896	16	.068
Total Suggestibility	.933	16	.268	.959	16	.652

Table 16

ANOVAs Results for the GSS-1

	<i>df</i>	<i>F</i>	<i>p</i>	η_p^2	<i>Observed Power</i>
Immediate Recall	1, 30	7.220	.012	.194	.739
Delayed Recall	1, 30	8.929	.006	.229	.824
Yield 1	1, 30	.015	.904	.000	.052
Yield 2	1, 30	1.512	.228	.048	.222
Shift	1, 30	.012	.913	<.001	.051
Total Suggestibility	1, 30	.002	.969	<.001	.050

Hypothesis 1.1. This hypothesis proposed that Immediate Free Recall scores would be significantly lower in the Manipulation group when compared to the Control group. A one-way ANOVA revealed a statistically significant difference between the Control and Manipulation group on Immediate Free Recall ($p = .012$). Cohen's effect size ($d = .95$) indicated a large practical difference between the groups, with Immediate Free Recall scores being significantly lower in the Manipulation group as compared to the control group. This finding supported the hypothesis.

Hypothesis 1.2. This hypothesis proposed that Delayed Free Recall scores would be significantly lower in the Manipulation group when compared to the Control group. A one-way ANOVA revealed a statistically significant difference between the Control and Manipulation group on Delayed Free Recall scores ($p = .006$; $d = 1.06$). This indicates a large practical difference between the groups, with Delayed Free Recall scores being significantly lower in the Manipulation group as compared to the control group. This finding supported the hypothesis.

Hypothesis 1.3. This hypothesis proposed that Yield 1 scores would be significantly lower in the Manipulation group when compared to the Control group. A

one-way ANOVA revealed no significant differences between groups with respect to Yield 1 scores ($p = .904$). The Cohen's d of .04 indicates no practical differences between the groups, which did not support the hypothesis.

Hypothesis 1.4. This hypothesis proposed that Yield 2 scores would not significantly differ between groups. A one-way ANOVA was conducted, which revealed no significant difference between groups with respect to Yield 2 scores ($p = .228$). Although the result was nonsignificant, the Cohen's d of .44 suggests a moderate between-group difference. This finding did not support the hypothesis.

Hypothesis 1.5. This hypothesis proposed that Shift scores would not be significantly differ between groups. A one-way ANOVA revealed no significant differences between groups with respect to Shift scores ($p = .913$). The Cohen's d of .04 indicates no practical differences between the groups, providing support for this hypothesis.

Hypothesis 1.6. This hypothesis proposed that Total Suggestibility scores would be significantly higher in the Manipulation group when compared to the Control group. A one-way ANOVA revealed no significant differences between groups with respect to Total Suggestibility scores ($p = .969$). The Cohen's d of .01 indicates no practical differences between the groups, which did not support the hypothesis.

Hypothesis 2

Hypothesis 2 proposed that instructions to feign heightened suggestibility would significantly impact PAI validity scale scores when compared to the control group. Measures of central tendency and variability (Mean, Standard Deviation, and Range) are

presented in Table 17. Tables 18 and 19 presents results of Levene's Test of Equality of Error Variances and Shapiro-Wilk Test of Normality, respectively, which test assumption that allow for the ANOVAs to be conducted; cases in which analyses failed to meet these assumptions will be discussed in further detail. Results of the series of ANOVAs are presented in Table 20. Results for individual hypotheses are discussed below.

Table 17

Central Tendency and Variability for the PAI

	<i>Control</i>			<i>Manipulation</i>		
	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
Inconsistency	45.06	6.816	37-64	55.38	7.658	40-67
Infrequency	52.94	6.728	44-67	60.00	19.121	40-110
Negative Impression Management	49.69	6.416	44-66	55.50	12.253	41-81
Positive Impression Management	45.69	15.217	15-66	42.13	11.598	22-66

Table 18

Levene's test for the PAI

	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
Inconsistency	.595	1	30	.446
Infrequency	4.949	1	30	.034
Negative Impression Management	8.105	1	30	.008
Positive Impression Management	.738	1	30	.397

Table 19

Shapiro-Wilk Test of Normality for the PAI

	<i>Control</i>			<i>Manipulation</i>		
	<i>Statistic</i>	<i>df</i>	<i>p</i>	<i>Statistic</i>	<i>df</i>	<i>p</i>
Inconsistency	.876	16	.033	.960	16	.655
Infrequency	.921	16	.177	.805	16	.003
Negative Impression	.793	16	.002	.867	16	.024
Management Positive Impression	.939	16	.341	.980	16	.965

Table 20

ANOVAs Results for the PAI

	<i>df</i>	<i>F</i>	<i>p</i>	η_p^2	<i>Observed Power</i>
Inconsistency	1, 30	16.188	.000	.350	.973
Positive Impression Management	1, 30	.555	.462	.018	.111

Table 21

Welsh's ANOVAs Results for the PAI

	<i>df</i>	<i>F</i>	<i>p</i>
Infrequency	1, 18.658	1.942	.180
Negative Impression Management	1, 22.650	2.826	.107

Hypothesis 2.1. This hypothesis proposed that Inconsistency (INC) scores would be significantly higher in the Manipulation group as compared to the Control group. The control group had one outlier on INC as assessed by inspection of a boxplot. This score was retained, as it was not a significant departure from expected response patterns on the PAI (t=64). The Shapiro-Wilk test of normality indicated that INC scores were not normally distributed for the control group ($p < .05$). However, a one-way ANOVA was

conducted to better understand the sample's characteristics and mean differences. There were statistically significant differences between groups with respect to INC scores, ($p = .000$), with instructed feigners endorsing more items within the Inconsistency scale. The Cohen's d of 1.42 indicates this is a large practical difference between the groups, which supports the hypothesis.

Hypothesis 2.2. This hypothesis proposed that Infrequency scores would be significantly higher in the Manipulation group when compared to the Control group. The manipulation group had two outliers on INF as assessed by inspection of a boxplot. These scores were retained, as they were not significant departures from expected response patterns on the PAI. The Shapiro-Wilk test of normality indicated that INF scores were not normally distributed for the manipulation group ($p > .05$), while Levene's Test of Equality of Error Variance indicated that the error variances for Infrequency scores were not homogeneous ($p < .05$). A Welsh's t-test was conducted to better understand the sample's characteristics and mean differences. The Welsh's t-test results indicate that the group means were not statistically significantly different ($p > .05$; See Table 21). However, the Cohen's d of .49 suggests a medium effect, with more frequent endorsement of items on the Infrequency scale for the manipulation group. This finding supports the hypothesis.

Hypothesis 2.3. This hypothesis proposed that Negative Impression Management (NIM) scores would be significantly higher in the Manipulation group when compared to the Control group. The Shapiro-Wilk test of normality indicated that NIM scores were not normally distributed for both groups ($p > .05$). A one-way Welsh's ANOVA was conducted, which revealed no statistically significant differences on NIM scores (p

>.05; See Table 21). The Cohen's d of .59 indicates a medium effect, with instructed feigners endorsing more items within the NIM scale than the control group. This finding supported the hypothesis.

Hypothesis 2.4. This hypothesis proposed that Positive Impression Management (PIM) scores would be significantly higher in the Manipulation group when compared to the Control group. A one-way ANOVA revealed no statistically significant differences between groups with respect to PIM scores ($p = .462$). However, the Cohen's d of .26 indicates a small effect size, with instructed feigners endorsing fewer items within the PIM scale than the control group. This finding did not support the hypothesis.

Hypothesis 3

Hypothesis 3 aimed to understand differences group assignment would have on TOMM scale scores. Measures of central tendency and variability (Mean, Standard Deviation, and Range) are presented in Table 22. Due to the means being both statistically and practically identical across both trials, no further analyses were conducted. This finding did not support the hypothesis.

Table 22

Central Tendency and Variability for the TOMM

	<i>Control</i>			<i>Manipulation</i>		
	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
Trial 1	49.38	.957	47-50	49.00	1.633	44-50
Trial 2	50.00	0	N/A	50.00	0	N/A

Hypothesis 4

Hypothesis 4 aimed to understand differences group instructions would have on

GCS scores. Control group scores ($M=8.688$, $SD=1.740$) were compared to Manipulation group scores ($M=10.375$; $SD= 3.01$). Levene's Test of Equality of Error Variances ($p=.088$) and Shapiro-Wilk Test of Normality ($ps=.416$ and $.180$) indicated the assumptions for ANOVA testing were met. A one-way ANOVA was conducted, which revealed no significant difference between groups with respect to GCS scores ($F(1,30) = 3.722$, $p = .062$). However, the Cohen's d of $.69$ indicates a medium effect, with instructed feigners endorsing more items on the GCS than the control group. This finding supports the hypothesis.

Exploratory Analyses

To contrast current findings with those within the literature, additional exploratory analyses were conducted. These compared current findings to previous findings regarding the ability to feign on the GSS, as well as comparisons to normative samples.

Comparison to previous literature. First, the current GSS-1 subscale scores and GCS scores were compared to those collected within Boon, Gozna, and Hall (2008) through a series of Independent Sample t-tests. Both control groups were compared (See Table 23), while the "Faking Bad" group was compared to the current study's manipulation group (See Table 24). Results indicated no significant differences between control groups. An examination of Cohen's effect size suggests a moderate effect size for Immediate and Delayed Recall ($d = .64$ and $d = .56$, respectively), with higher recall scores garnered within the current study. An examination of Cohen's effect size suggests a small effect size for Yield 1 ($d = .18$), Yield 2 ($d = .21$), Shift ($d = .31$), and Total Suggestibility ($d = .29$), with lower scores within Boon, Gozna, and Hall's study.

Table 23

Independent-Sample t-test comparison to Boon, Gozna, and Hall (2008) “Standard Procedure” group

	<i>“Standard” group</i>		<i>Control Group</i>		<i>Analysis</i>	
	<i>Mean</i>	<i>S.D.</i>	<i>Mean</i>	<i>S.D.</i>	<i>t-value</i>	<i>p</i>
Immediate Recall	17.14	5.38	20.531	5.143	1.948	0.060
Delayed Recall	16.86	5.6	19.781	4.764	1.712	0.096
Yield 1	5.76	3.38	5.25	2.113	-0.562	0.578
Yield 2	6.9	3.43	6.25	2.696	-0.645	0.523
Shift	5.67	2.69	4.69	3.610	-0.910	0.369
Total Suggestibility	11.43	4.92	10.00	4.913	-0.877	0.387

Table 24

One-Sample t-test comparison to Boon, Gozna, and Hall (2008) “Faking Bad” group

	<i>“Faking Bad” group</i>		<i>Manipulation Group</i>		<i>Analysis</i>	
	<i>Mean</i>	<i>S.D.</i>	<i>Mean</i>	<i>S.D.</i>	<i>t-value</i>	<i>p</i>
Immediate Recall	12.48	5.79	15.969	4.436	2.075	0.045
Delayed Recall	12.61	5.3	15.094	4.083	1.610	0.116
Yield 1	10.76	2.17	5.34	2.241	-7.389	0.000
Yield 2	10.76	2.74	7.47	2.907	-3.496	0.001
Shift	4.45	3.52	4.56	2.707	0.107	0.915
Total Suggestibility	14.81	4.16	9.938	4.106	-3.555	0.001

When comparing manipulation groups, the current study had significantly higher Immediate Recall scores ($d = .66$), and significantly lower Yield 1 ($d = 2.46$), Yield 2 ($d = 1.17$), and Total Suggestibility scores ($d = 1.18$). There were no significant differences on Delayed Recall ($d = .52$) or Shift ($d = .03$) scores between manipulation groups.

Comparison to GSS-1 Normative sample. The current GSS-1 subscale scores for the control group were compared to the normative sample presented in the GSS manual (Gudjonsson, 1997; see Table 25). Results indicate that scores did not significantly differ on any GSS-1 subscales. Of note, in line with previous literature (Frumkin, Lally, & Sexton, 2012), despite not being significantly different, the Total

Suggestibility and Shift scores for the current study's sample were approximately one-half the SD above the mean for the UK normative sample, as seen by their approaching meaningful differences. However, these differences were less apparent in the current analysis when examining Yield 2 scores.

Table 25

One-Sample t-test comparison to the normative sample for the GSS-1 (Gudjonsson, 1997)

	<i>Gudjonsson Normative sample</i>		<i>Control Group</i>		<i>Analysis</i>	
	<i>Mean</i>	<i>S.D.</i>	<i>Mean</i>	<i>S.D.</i>	<i>t-value</i>	<i>p</i>
Immediate Recall	21.3	7.1	20.531	5.143	-.753	.463
Delayed Recall	19.5	7.5	19.781	4.764	.236	.817
Yield 1	4.6	3.0	5.25	2.113	1.230	.238
Yield 2	5.6	3.8	6.25	2.696	.965	.350
Shift	2.9	2.5	4.69	3.610	1.981	.066
Total Suggestibility	7.5	4.6	10.00	4.913	2.036	.060

Comparison to intellectually disabled sample. The current GSS-1 subscale scores were compared to those presented in the GSS manual (Gudjonsson, 1997) for participants with low IQ (IQ<75; See Table 26). Results indicate that scores vary significantly on all GSS-1 subscales, save for Shift scores. The current sample recalled significantly more aspects of the GSS-1 narrative on both Immediate and Delayed Recall. Further, they yielded significantly less during both Yield trials. Finally, Total Suggestibility scores were significantly lower in the current study's manipulation group.

Table 26

One-Sample t-test comparison to the intellectually disabled normative sample for the GSS-2 (Gudjonsson, 1997)

	<i>Intellectually Disabled norms</i>		<i>Manipulation Group</i>		<i>Analysis</i>	
	<i>Mean</i>	<i>S.D.</i>	<i>Mean</i>	<i>S.D.</i>	<i>t-value</i>	<i>p</i>
Immediate Recall	8.1	4.9	15.969	4.436	7.095	<.001
Delayed Recall	6.0	4.9	15.094	4.083	8.908	<.001
Yield 1	9.8	3.5	5.34	2.241	-7.953	<.001
Yield 2	9.4	3.5	7.47	2.907	-2.658	.018
Shift	4.8	3.0	4.56	2.707	-.351	.731
Total Suggestibility	14.6	4.6	9.938	4.106	-4.542	<.001

Comparison to Frumkin, Lally, and Sexton (2012) US Forensic norms.

Finally, both the control group (Table 27) and manipulation group (Table 28) GSS scores were compared to scores presented by Frumkin, Lally, & Sexton (2012) to explore how they compared to their U.S. forensic normative adult sample. The control group scored consistently lower on all GSS-1 scores, save for Shift. When comparing the US sample to the current study's manipulation group, the manipulation group was noted to have significantly lower Total Suggestibility scores. No other subscales were significantly different when comparing these groups.

Table 27

One-Sample t-test comparison between control group means and US Forensic sample (Frumkin, Lally, & Sexton, 2012)

	<i>US Forensic Adults</i>		<i>Control Group</i>		<i>Analysis</i>	
	<i>Mean</i>	<i>S.D.</i>	<i>Mean</i>	<i>S.D.</i>	<i>t-value</i>	<i>p</i>
Yield 1	6.4	3.2	5.25	2.113	-2.177	.046
Yield 2	8.3	3.6	6.25	2.696	-3.042	.008
Shift	5.8	3.9	4.69	3.610	-1.233	.237
Total Suggestibility	12.2	5.5	10.00	4.913	-1.791	.093

Table 28

One-Sample t-test comparison between manipulation group means and US Forensic sample (Frumkin, Lally, & Sexton, 2012)

	<i>US Forensic Adults</i>		<i>Manipulation Group</i>		<i>Analysis</i>	
	<i>Mean</i>	<i>S.D.</i>	<i>Mean</i>	<i>S.D.</i>	<i>t-value</i>	<i>p</i>
Yield 1	6.4	3.2	5.53	2.247	-1.885	.079
Yield 2	8.3	3.6	7.78	3.005	-1.144	.271
Shift	5.8	3.9	5.13	3.612	-1.828	.087
Total Suggestibility	12.2	5.5	10.750	5.066	-2.204	.044

CHAPTER V

Discussion

The current research study was conducted to inform the use of the GSS-1 in the context of evaluations of false confession claims, specifically whether GSS-1 scores could be successfully manipulated following instruction. The study aimed to clarify what factors, if any, should be considered when trying to differentiate between genuine and ingenuine responders using only the data within the GSS-1. The second purpose of the study was to determine if manipulation instructions impacted additional assessments tools that could independently differentiate the groups. Given that the GSS-1 is a unique assessment tool that could be used to support false confession claims, empirical support is necessary to differentiate valid and invalid responders to bolster scientific reliability and utility.

Interpretation of Hypothesis 1: Gudjonsson Suggestibility Scale Findings

It was hypothesized that participants in the manipulation condition would suppress free recall scores and would yield more freely to leading questions. Based on previous research (Boon, Gozna, & Hall, 2008; Woolston, Bain & Baxter, 2006), it was proposed that there would be significant differences between groups with respect to Immediate Recall, Delayed Recall, Yield 1, and Total Suggestibility scores following instruction to feign on the GSS-1. Concurrently, it was proposed that Yield 2 and Shift scores would not be significantly impacted due the absence of shifting responses following the negative feedback phase within the GSS-1.

The findings of the current study indicate that both the initial and delayed free recall scores were successfully suppressed following instruction. However, participants

were not able to successfully manipulate any other GSS-1 subscales. This raises interesting questions regarding the impact of the feigning instructions on the other domains of the GSS-1.

Free Recall Scores. Beginning with Smith and Gudjonsson's 1986 study, Free Recall scores have consistently been cited as being the most susceptible to manipulation efforts following instruction. This finding was almost unanimous within the feigning literature for Initial Recall scores, save for the findings of Boon and Baxter (2002). The current findings support that instructed feigners can successfully suppress their Initial Recall scores. Only Boon, Gozna, and Hall (2008) included a time delay when administering the GSS-1, which prevented conclusions to be drawn regarding the impact of instruction on Delayed Recall scores. Findings of this study, in conjunction with Boon, Gozna, and Hall (2008), further support that participants can consistently suppress the number of aspects of the GSS-1 narrative they verbalize during both recall phases when compared to control groups.

Despite this ability to suppress free recall scores, the current instructions utilized efforts were not sufficient to mimic recall rates exhibited by the intellectually disabled normative sample presented by Gudjonsson (1997). Manipulation group participants verbalized significantly more aspects of the narrative than a truly vulnerable group on both Immediate and Delayed Recall scores, suggesting that the methods in which participants chose to manipulate their verbal behavior were not sufficient to adequately mimic patterns exhibited by truly vulnerable individuals. In line with previous interpretations, participants likely were unable to accurately conceptualize how a truly vulnerable individual may respond to the GSS-1 during free recall queries, and when

weighing the request to maintain believability in their responses chose to only moderately suppress their free recall verbalizations.

Therefore, in the context of a forensic evaluation, the presence of both free recall scores could help further the interpretive process when comparing scores to the appropriate normative sample. However, these recall scores must be interpreted considering the GSS-1 primary scales to best understand these results of the present study.

GSS-1 Primary Scales. Contrary to initial hypotheses, manipulation group participants were not able to successfully manipulate their responses to score significantly higher on any of the remaining GSS-1 subscales. Further, although literature suggested that the manipulation group scores would not significantly differ from the truly vulnerable group, the current study's manipulation group had significantly lower suggestibility scores on all GSS-1 primary subscales when compared to truly vulnerable individuals, save for Shift scores. Interpretations of GSS-1 data from a participant within the current study's manipulation group within the US criminal justice system would suggest the individual would not be particularly vulnerable to interrogative pressures, with significantly lower Total Suggestibility scores limiting more favorable interpretations. Thus, the efforts to feign by the current study's manipulation are considered unsuccessful.

The findings of the current study indicate that even when participants are taking efforts to appear vulnerable, they cannot successfully differentiate themselves from a control group or accurately replicate the response patterns of truly vulnerable individuals when administered the GSS-1. These findings support those presented in Hanson,

Smeets, and Jellicic's 2010 study, which was the first to suggest that the GSS-1 "is relatively unaffected by attempts at faking heightened suggestibility" (p. 227). Initial interpretations of the findings would support these claims. However, these results are surprising in the larger context of the literature base and are particularly interesting given the additional incentive given to the participants within the manipulation group to appear more suggestible, as well as the inclusion of the time delay.

Comparison to Previous Literature

Given that the findings related to the primary subscales (e.g. Yield 1, Yield 2, Shift, and Total Suggestibility) of the GSS-1 differ from the findings of Baxter and Bain (2002), Woolston, Bain and Baxter (2006), and Boon, Gozna, and Hall (2008), it becomes important to delineate why the current findings, which support those presented in Hanson, Smeets, & Jellicic's (2010) study, more accurately reflect the patterns in which participants respond to feigning instructions. By examining the differences in the methodology and samples within the study, this study aims to make light of the conflictual information presented.

The current study's means were compared to normative samples to assess whether there were confounds relating to the participants recruited. Control group GSS-1 scores from the current study were not statistically different from the normative sample presented in the GSS-1 Manual (Gudjonsson, 1997), or from the control group from Boon, Gozna, and Hall (2008). Based on these analyses, it is unlikely that the findings in the current study's findings were related to sampling errors.

Thus, differences in findings must relate to the manipulation group, namely their response to the instructions and response to the procedures of the study. It could be

hypothesized that the manipulation group chose to ignore the manipulation instructions presented, or simply did not engage with the instructions. However, the significant decrease in Initial Free Recall scores compared to the control group suggests that this is not the case. Further, the significant difference on the Delayed Recall scores indicate that the participants were aware of, and attempting to engage with, their manipulation both before and after the time delay. It can be concluded that, at least to some degree, the participants within the manipulation group were engaged with their manipulation during the majority of the administration time.

The current study's manipulation group had significantly higher Immediate Recall scores, while scoring significantly lower on Yield 1, Yield 2, and Total Suggestibility compared to Boon, Gozna, and Hall's (2008) "Faking Bad" group. It is important to consider that the instructions given to the groups were quite similar, specifically in wording regarding appearing gullible without appearing as if they were acting or dramatizing their responses. One of the main differences in the instructions was the addition of the incentive given to participants in the current study to engage in the manipulation as suggested by Baxter et al. (2013).

Proposed Interpretations. There are two factors that may have accounted for such significant differences on primary scales of the GSS-1 between the current study and that presented in Boon, Gozna, and Hall (2008). The first was the implementation of the PAI during the delay between the Initial and Delayed Recall phases. Boon, Gozna, and Hall (2008) did not note how they had their participants utilized this time, while participants within the current study were presented with 344 questions of the PAI in which they had to complete. It could be that, in line with the findings of Baxter et al.

(2013), that the added cognitive load associated with responding to the PAI would have decreased resources available to manipulate responses during the Yield 1 questioning phase. In all previous studies, the participants were presented the feigning instructions, the GSS-1 narrative, and the Yield 1 questions in quick succession. Their findings would indicate that when this procedure is followed, participants may be more cognitively prepared to manipulate their scores. However, within the current study the delay and the associated cognitive demands depleted these resources. The current findings suggest that the GSS-1 has a built-in opportunity to increase the cognitive load and thus the defense against feigning efforts by using the 50-minute delay for additional assessment tools such as personality assessments.

A second factor that could have accounted for the differences between the current findings and the previous literature could relate to the inclusion of the extra incentive suggested by Baxter et al. (2013) to foster additional adherence to the manipulation instructions. One hypothesis proposed is that this addition may have had a paradoxical impact on responding patterns on the primary subscales on the GSS-1, such that the added attention to the desire to feign led to suppressed feigning efforts as seen in the high Interrogative Suggestibility (IS) group in Baxter et al.'s (2013) study. Baxter et al. suggested that the baseline IS level impacted how an undergraduate participant responded to instructions to feign; participants with mild to moderate IS levels were successful in increasing their suggestibility scores, while High IS participants paradoxically garnered lower scores following instruction. The authors interpreted these results to indicate that the high IS group, while “more trusting and value trust, may be uncomfortable with the deception and the invitation to be party to it” (p.921), which led to them become less

trusting within the confines of the study and subsequently lower IS scores. When applying these interpretations to the current study, the addition of the added incentive may have generalized the discomfort experienced by the high IS group to all participants, which led to a decrease in Yield 1, Yield 2, and Shift responses.

It should be noted that if this were the main factor influencing this change, the results suggesting that the GSS-1 is resistant to feigning efforts would not be generalizable to the forensic setting because, as Baxter et al. (2013) writes, “people who come to the test with established deceptive intent will not be affected [by the introduction of deceptive intent] in this way” (p. 922-923). While this could play a factor in the differences between the studies, the secondary hypotheses, and the assessments included to test them, may be a more important influencing factor to consider.

Hypotheses 2-4: Personality Assessment Inventory, Gudjonsson Compliance Scale, and Test of Memory Malingered Findings

The current study was the first to include additional assessments administered alongside the GSS-1, which both helped replicate the data available when administering a forensic battery for false confession cases and allowed for the time delay within the GSS-1 to be included without the inclusion of inconsequential distractor tasks. It was believed that the PAI, GCS, and TOMM could help differentiate between the control group and instructed feigners. Secondly, responses were anticipated to shed light on how participants chose to engage with the manipulation instructions. The only significant differences between groups that were identified on the PAI were on the Inconsistency scale; participants within the manipulation group responded more inconsistently to questions throughout the PAI. Infrequency, Positive Impression Management, and

Negative Impression management were not significantly different between groups. No significant differences were found between group's GCS or TOMM scores.

However, the current study's sample size may have negatively impacted the ability to detect group differences and therefore effect sizes were examined. The results suggest large practical differences on the Inconsistency scale, medium practical differences on the Infrequency and Negative Impression Management (NIM) scales, and a small practical effect on the Positive Impression Management (PIM) scale. Save for PIM scores, the feigning groups all responded in the expected direction as compared to the control group. This suggests a pattern of increased responding of uncommon symptomology, and a decrease in desire to present themselves in a positive light. Of note, almost none of these instructed feigning responders elevated any validly scale scores to a degree that would invalidate the test. In examination of the individual scores, only one participant within the feigning group responded in such a way that could invalidate their PAI profile (INF=110; Morey, 2007).

It becomes important to consider whether participants were engaging with their manipulation during the PAI. As discussed, the suppression of the Initial and Delay Recall scores within the manipulation group suggest that both before and following the completion of the PAI participants were engaging with their manipulation. Then the question becomes did the participants ignore their instructions while completing the PAI, or were these insufficient to adequately impact how participants responded to the PAI stimuli. When examining the verbiage utilized within the instructions, the participants were told to apply the instruction to "multiple assessments," but the focus was on responding in conjunction with the interviewer.

Participants were asked to “convince the interviewer that [they] are easily influenced by pressure from other people.” Then asked to “go along with everything the interviewer says to you.” There is strong language that focuses on the verbal interaction with the researcher presenting the assessments, but less focus on responding to pen and paper assessments that were provided by the assessor by potentially not assessed as being applicable. This could suggest that the written assessments were not approached in the same way in which the GSS-1, an interactive assessment, was approached by the participants in the feigning group. It is unclear whether additional instructions that specifically mention feigning on written assessments would impact the scores in a similar manner.

Alternatively, the participants could have manipulated their scores, but chose a pattern that was not sufficient. The instructions to avoid overt detection may have led to some inconsistent attempts at changing their responses that did not result in significant changes in scores. For example, the sliding scale of response options within the PAI and the instructions to maintain believability may have led to only mild alterations of reporting. Although the current study did not intend to assess clinical scales presented in assessments such as the PAI, future studies could assess for any differences that arise on clinical scales, especially in a more clinically diverse sample. Future research is necessary to see if the absence of significant findings was related in part to the sample size, or if the manipulation as presented is truly insufficient to cause the desired change.

With respect to GCS scores, while no significant differences were found, a medium practical effect was seen for the manipulation group. This was the expected direction of the results due to the face-validity of the test and the ease in which

participants could anticipate the nature of the questioning. It is interesting that groups did not significantly differ, but these findings lend support to the hypothesis that the pattern chosen to manipulate scores was simply ineffective, not absent.

Interestingly, there were no differences on TOMM scores between groups, indicating that participants did not attempt to feign memory deficits when attempting to comply with the manipulation instructions. Because this assessment is interactive with the researcher administering the assessments, and is presented as a memory task similar to how the GSS-1 is framed, these results are surprising and give an interesting look into the patterns that the individuals chose to respond. This may be related to the specificity of the feedback given after each TOMM response. Having the GSS-1 and its negative feedback administered prior to the TOMM administration, participants may have been more comforted by the continuous positive and negative feedback presented within the TOMM protocol, which inclined participants to respond in a more genuine manner. Further, when choosing a method in which to respond in accordance with the manipulation instructions, it is possible that the participants did not believe that appearing to have memory deficits would assist in their cause, thereby not being caught in the TOMM's trap.

Integration of Findings and the Statement of the Problem

Frumkin, Lally, & Sexton (2012) explored factors relevant to findings during a suspected police-induced false confession evaluation to increase effectiveness and acceptability of expert witness testimony within the United States. They suggest that it is imperative to examine both interrogative factors and personal characteristics to best understand whether the conditions were right to elicit a false confession. The authors

proposed that after a throughout investigation of the interrogation is conducted, a psychological evaluation of the false confessor is necessary; this includes a clinical interview that is supplemented by a battery of clinical assessment tools to substantiate qualitative claims (Frumkin, Lally, & Sexton, 2012). However, they make no suggestions as to which assessments should be utilized and do not mention areas of concern that could arise during these evaluations, such as indicators to assess for individuals with ulterior motives that claim they have falsely confessed. This could undercut the validity of false confession literature as assumed by the Judge when accepting false confession testimony under *Daubert* rules. Therefore, this study was conducted to bolster the continued use of the Gudjonsson Suggestibility Scale within the forensic realm.

Findings support the prevailing belief within the literature that individuals can suppress both Immediate and Delayed Recall scores following instruction to feign with relative ease. Contrary to the bulk of the literature, the current findings suggest that the primary subscales of the GSS-1 are relatively immune to feigning efforts when administered with the time delay and a sufficiently demanding distractor task, as indicated by inability for participants to significantly alter their Yield 1, Yield 2, Shift, or Total Suggestibility scores within the present study. This could be in part due to participants within the study not being able to “fully grasp the idea behind the GSS.” (p. 227, Hansen, Smeets, & Jelicic, 2010), such that the response patterns chosen did not sufficiently attend to the means in which the GSS-1 measures suggestibility. However, the findings of the current study suggest that the addition of the full-time delay, in conjunction with a sufficiently engaging cognitive tasks such as a personality assessment tool, is sufficient in decreasing ability to manipulate suggestibility scores. This is in line

with the findings of Drake et al. (2012), who emphasize the need for increased cognitive demands to thwart feigning attempts.

This finding also suggests that, when administered the full GSS-1 protocol, the response patterns of instructed feigners are able to be differentiated from truly vulnerable individuals, decreasing the concern for false identification raised in previous research for the primary subscales of the GSS-1. This does not mean that considerations from previous literature that should be discarded, especially if the GSS-1 is administered following protocols that decrease the cognitive demands of individuals intending to feign. For example, Boon, Baxter, and Hall (2008) suggest that there are warning flags that may indicate feigning such as significant yielding to the initial presentation of leading questions, as indicated by higher Yield 1 scores, with lack of response to negative feedback, as indicated by low Shift scores. While the current study does not support these hypotheses, the qualitative differences in instructions between the feigning groups mimics the differences in how the assessment could be presented to an instructed feigner, thereby eliciting different response patterns. Further, if participants had additional knowledge on the GSS-1 beyond that presented in the current study, that may bring about findings similar to those in previous literature

Limitations of the Study

While the results of the study add to the literature in the field, there are several limitations that should be considered when interpreting the results. First, although efforts were taken to account for the small size of the current sample, the current study was limited the ability to find significant group differences. This is more apparent with the larger effect sizes identifying group differences on the PAI rather than the GSS-1, which

should be interpreted with caution until a replication study can be performed with adequate sampling. However, the current findings regarding the response to the GSS-1 following instruction should be used to inform both clinical and research practices.

Second, recently concerns have been raised regarding the generalizability of feigning literature utilizing undergraduate participants to forensic cases in which they are intended to be compared. It is undeniable that the motivation to feign in a research study and the motivation to feign within the clinical setting are vastly different. When discussing the limitations of their findings, Baxter et al. (2013) are the first authors to consider that their undergraduate participants “had no reason to work that hard” (p.923) when engaging in the manipulation, and propose that additional incentives are necessary to increase engagement with the manipulation. Although the present study was the first to offer additional incentive to those within the feigning condition, it is unclear whether the motivation utilized could garner similar manipulation engagement seen within individuals who truly desire to feign.

Finally, the present study did not include a full battery of assessments that one could administer during a forensic evaluation exploring the veracity of a false confession claim. Although assessment tools were utilized that specifically assessed for memory malingering, this may not best account for all approaches an individual could take to appear more suggestible. There are qualitative differences between attempting to feign psychiatric symptomology and/or memory deficits as traditionally conceptualized by malingering and the behaviors that are necessary in order to feign heightened suggestibility. While similar, assessing for malingering may not be sufficient to understand these disingenuous responding patterns. Therefore, a broader array of

assessments could contain be utilized to better account for these differences. This should continue to be examined in future research.

Implication for Future Research

The findings of the current study can be used to influence both future research within the field of false confession research, as well as impact the way in which experts evaluate false confession claim evaluations. It has been difficult to generalize past research findings to clinical applications due to the shifting GSS-1 administration that have not followed the recommended procedures. The administration manual suggests that by eliminating the delay the overall difficulty of the assessment is decreased due to the recently of the narrative administration (Gudjonsson, 1997). By removing the 50-minute delay and shortening the administration time of the GSS-1, researchers may have inadvertently lowered the cognitive demand required to feign, thereby artificially increasing the ability or desire to appear more vulnerable when presented with the Yield 1 questions. Future research may be inclined to replicate Smeets et al. (2009) to assess the impact of the various administration changes when manipulation intent of the administration is also considered.

In addition, researchers should continue to assess whether individuals instructed to feign heightened suggestibility scores would also be likely to change their verbal and written behavior on a variety of assessment tools that would likely be administered in the forensic setting alongside the GSS-1. These should include malingering assessments and similar response validity assessments to assess for the variety of means in which the general population may choose to appear more vulnerable. If possible, this should also be examined with a forensic population to bolster generalizability.

Clinical Implications

Presently, admissibility of expert witness testimony supporting false confession claims are mixed at best. While there is a robust peer reviewed literature base that highlights empirically-supported individual and interrogative facets that increase the likelihood of false confession rates (See Kassin et al., 2010), the admissibility of this testimony only more recently accepted (*Crane v. Kentucky*, 1986). Under *Daubert* rules of evidence (*Daubert Merrell Dow Pharmaceuticals Inc.*, 1993), the presiding judge acts as the gatekeeper for admitting confession expert witness testimony. While this has increased the admission of testimony in cases where claims of false confessions are made, it also raises the burden of proof needed to be deemed as supported by the scientific community. While the tides turn towards increased acceptance of confession literature as indicated by the numerous cases presented by Frumkin, Lally, and Sexton (2012), research must continue to explore the best clinical practices and recommendations to increase the precision of these evaluation.

This study supports the findings of the most recent feigning literature (Hansen, Smeets, & Jelicic, 2010). The present study further proposes that the GSS-1 is resistant to feigning efforts primarily when the participant is administered the GSS-1 in its full format, with a relevant distractor task such as a personality assessment too. While no significant findings were seen on the primary subscales beyond Initial and Delayed Recall scores, it is still imperative for forensic psychologists employ caution and good clinical practice when utilizing the Gudjonsson Suggestibility Scale during a false confession assessment. When interpreting individual assessment responses, especially in cases where there is a heightened motivation to manipulate verbal and written behavior,

researchers should continue to assess for inaccuracies within the GSS-1 scales.

It is important to consider that within the course of law, “items of evidence do not appear in a vacuum; rather, they appear through the presentations of attorneys and the theories they espouse in their opening statements and closing arguments.” (p137, Appleby & Kassin, 2016). As such, it is imperative that GSS-1 scores are not interpreted without the context of other information. In forensic false confession cases, detailed analyses of the interrogation are also necessary for the psychologist to understand what factors that the individual might have reacted to during an interrogation. Gudjonsson suggests that the GSS-1 scores “should not be interpreted in isolation from other information, including that obtained during a clinical interview” (p.29; 1997). The unique combination of interrogation factors and personal factors are the focus of these forensic evaluations to help determine if someone’s suggestibility level contributed to a false confession (Frumkin, Lally, & Sexton, 2012). These can be incorporated into other clinical information gathered through other psychological testing and through a clinical interview with the alleged false confessor. Overall, this information should be used to better understand how the individual might behave in a brief interrogation considering the totality of the information collected during a clinical interview.

Finally, research unanimously supports the addition of a variety of safeguards to protect against wrongful convictions in cases where false confessions are elicited, which include adapting police interrogation procedures and policies to increasing support for expert witness testimony supporting false confession claims (Appleby & Kassin, 2016; Kassin et al, 2010). The current study aims to help forensic psychologists conducting these forensic evaluations increase their precision to further benefit those who falsely

confess or are wrongly convicted.

Conclusions

This study aimed to assess whether participants instructed to feign higher suggestibility could significantly alter their verbal behavior to better understand if similar feigning could occur in the context of a false confession evaluation. Findings indicate that instructed feigners cannot successfully manipulate responses to the GSS-1 beyond the recall scores. Instructions to feign elevated participant's motivation to respond inconsistently, as indicated by the PAI, however the instructions did not significantly impact other clinical assessment tools. Further, despite free recall scores being most consistently cited within the GSS-1 literature as the most susceptible measure to feigning instructions, the TOMM was unable to differentiate these participant from the control group. This indicates that the suppression strategy utilized by these individuals is different than those traditionally utilized by memory malingerers. Given these findings, future research should continue to explore assessments that are utilized alongside clinical evaluations for false confession claims to assess for suspected feigning.

References

- Appleby, S.C., & Kassin, S. M. (2016). Self-report trumps science: Effects of confessions, DNA, and prosecutorial theories on perceptions of guilt. *Psychology, Public Policy, and Law*, 22, (2), 127-140. DOI: 10.1037/law0000080
- Baxter, J.S., & Bain, S.A. (2002). Faking interrogative suggestibility: The truth machine. *Legal and Criminal Psychology*, 7, 219-225.
- Baxter, J.S., & Bain, S.A., Pringle, A., Fowler, H., & Tafili, T. (2013). Opposing determinants of compliance and interrogative suggestibility. *Personality and Individual Differences*, 54, 918-924. DOI: 10.1016/j.paid.2013.01.004
- Beail, N. (2002). Interrogative suggestibility, memory, and intellectual disability. *Journal of Applied Research in Intellectual Disabilities*, 15, 129-137.
- Binet, A. (1900). *La Suggestibilite*. Paris: Doin.
- Brown v. Mississippi, 297 U.S. 278 (1936).
- Cattell, J. M., & Farrand, L. (1896). Physical and mental measurements of the students of Columbia University. *The Psychological Review*, 3(6), 618-648.
- Clare, I., & Gudjonsson, G. (1993). Interrogative suggestibility, confabulation, and acquiescence in people with mild learning difficulties (mental handicap): Implications for reliability during police interrogation. *British Journal of Clinical Psychology*, 32, 295-301.
- Colorado v. Connelly, 479 U.S., 157 (1986)
- Crane v. Kentucky, 476 U.S. 683, 106 S.Ct. 2142 (1986)

Daubert v. Merrell Dow Pharmaceuticals, Inc. 509 U.S. 579 (1993).

Drizin, S.A., & Leo, R.A. (2004). The problem of false confessions in the post-DNA world. *North Carolina Law Review*, 82, (3), 891-1003.

Drake, K.E., Lipka, S., Smith, C., & Egan, V. (2013). The effect of cognitive load on faking interrogative suggestibility on the Gudjonsson Suggestibility Scale. *Personality and Individual Differences*, 54, 845-849. DOI: 10.1016/j.paid.2012.12.011

Eysenck, H. J., & Furneaux, W. D. (1945). Primary and secondary suggestibility: An experimental and statistical study. *Journal of Experimental Psychology*, 35, 485-503.

Federal Rules of Evidence. (1975). St. Paul, MN: West.

Frumkin, I.B. (2010). Evaluations of competency to waive Miranda rights and coerced/false confessions: Common pitfalls in expert testimony. In: Lassiter, D.G. & Meissner, C. (Eds.) *Police Interrogations and False Confessions: Current Research, Practice, and Policy Recommendations*, Washington, D.C.: American Psychological Association.

Frumkin, I.B., Lally, S.J., & Sexton, J.E. (2012). A united states forensic sample for the Gudjonsson Suggestibility Scales. *Behavioral Sciences and the Law* 30, 749-763. DOI: 10.1002/bsl.2032

Frye v. United States, 293 F. 1013 (D. C. Cir., 1923).

Grabner, S.S., Conte, C.B., Walker, L.E., Nagle, M.H., & Shapiro, D. (August 2015).

Women and false confessions: Broadening the typology. Presentation given at the 123rd annual meeting of the American Psychological Association, Toronto, Canada

Gould, J. B., & Leo, R. A. (2010) One hundred years later: Wrongful convictions after a century of research. *Journal of Criminal Law and Criminology*, 100, (3), 825-868.

Gudjonsson, C.H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences*, 5, 303-314.

Gudjonsson, CH. (1987). A parallel form of the Gudjonsson Suggestibility Scale. *British Journal of Clinical Psychology*, 26, 215-221.

Gudjonsson, G. H. (1997). *The Gudjonsson Suggestibility Scales Manual*. Hove, UK: Psychology.

Gudjonsson, G. H., & Clark, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behavior*, 1, 83–104.

Gudjonsson, G. H. (1989). Compliance in an interrogative situation: A new scale. *Personality and Individual Difference*, 10 (5), 535-540. DOI: 10.1016/0191-8869(89)90035-4

Gudjonsson, G.H., Sigurdsson, J.F. Bragason, O.O., Einarsson, E., & Valdimarsdottir, E.B. (2004). Confession and denials and the relationship with personality

- Gudjonsson, G.H., & Sigurdsson, J.F. (2004). The relationship of suggestibility and compliance with self-deception and other-deception. *Psychology, Crime, & Law*, 10, (4), 447-453. DOI: 10.1080/10683160310001634278
- Gudjonsson, G. H., Rutter, S., & Clare, I. (1995). The relationship between suggestibility and anxiety among suspects detained at police stations. *Psychological Medicine*, 25, 875–878.
- Inbau, F. E., & Reid, J. E. (1962). *Criminal interrogation and confessions*. Baltimore, MD: Williams & Wilkins.
- Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2013). *Criminal interrogation and confessions* (5th ed.). Burlington, MA: Jones & Bartlett.
- Henkel, L. A., Coffman, K. A., & Dailey, E. M. (2008). A survey of people's attitudes and beliefs about false confessions. *Behavioral Sciences & the Law*, 26, 555–584. DOI: 10.1002/bsl.826
- Kassin, S.M., & Wrightsman L. (1985) Confession evidence. In Kassin S., & Wrightsman, L. (Eds.) *The Psychology of Evidence and Trial Procedure*. Beverly Hills, CA: Sage Publications.
- Kassin, S.M. & Neumann, K. (1997). On the power of confession evidence: An experimental test of the fundamental difference hypothesis. *Law and Human Behavior*, 21, (5), 469-484. DOI:10.1023/A:1024871622490

- Kassin, S.M., Redlich, A.D., Alceste, F., & Luke, T.J. (2018). On general acceptance of confession research: Opinions of the scientific community. *American Psychologist*, 73, (1), 63-80. DOI: 10.1037/amp0000141
- Kassin, S.M., Drizin, S.A., Grisso, T., Gudjonsson, G.H., Leo, R.A., & Redlich, A.D. (2010). Police-induced confessions: Risk factors and recommendations. *Law and Human Behavior*, 34, p.3-38. DOI: 10.1007/s10979-009-9188-6
- Kassin, S.M. (2008). False confessions: Causes, consequences, and implications for reform. *Current Directions in Psychological Science*, 17, 249-253. DOI: 10.1111/j.1467-8721.2008.00584.x
- Kassin, S.M. (2012). Why confessions trump innocence. *American Psychologist*, 67, (6), 431-445. DOI: 10.1037/a0028212
- Kassin, S.M. (2017). False confessions: How can psychology so basic be so counterintuitive? *American Psychologist*, 72, (9), 951-964. DOI: 10.1037/amp0000195
- Kingery, L.R., & Schretlen, D.J. (2007) Functional neuroimaging of deception and malingering. In K. B. Boone (Ed.) *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York, NY, US: Guilford Press.
- Kumho Tire v Carmichael*, 526 U.S. 137 (1999).
- Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009). 40 plus or minus 10, a new magical number: Reply to Russell. *The Clinical Neuropsychologist*, 23(5), 841-849. <http://dx.doi.org/10.1080/13854040902796735>

- Leo, R.A., & Ofshe, R.J. (1998). The consequences of false confessions: Deprivations of liberty and miscarriages of justice in the age of psychological interrogation. *The Journal of Criminal Law & Criminology*, 88, (2), 429-496.
- Leo, R. A. (2008). *Police interrogation and American justice*. Cambridge, MA: Harvard University Press.
- McCann, J.T. (1998). A conceptual framework for identifying various types of confessions. *Behavioral Sciences and the Law*, 16, (4), 441-453.
- Merckelbach, H., Muris, P., Wessel, I., & van koppen, P. J. (1998). The Gudjonsson suggestibility scale (GSS): Further data on its reliability, validity, and metacognition correlates. *Social Behavior and Personality*, 26, 203–210.
- Mittenberg, W., Patton, C., Canyock, E.M., & Condit, D.C. (2002). Base rate of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, (8), 1094-1102.
- Morey, L.C. (2007). *The Personality Assessment Inventory professional manual-2nd Ed.* Odessa, FL: Psychological Assessment Resources.
- Perske, R. (2008). False confessions from 53 persons with intellectual disabilities: The list keeps growing. *Intellectual and Developmental Disabilities*, 46, (6), 365-373.
DOI: 10.1352/2008.46:468-479
- Perske, R. (2011). Perske's list: False confessions from 75 persons with intellectual disability. *Intellectual and Developmental Disabilities*, 49, (5), 468–479. DOI: 10.1352/1934-9556-49.5.365

- Polczyk, R. (2005). Interrogative suggestibility: Cross-cultural stability of psychometric and correlational properties of the Gudjonsson Suggestibility Scales. *Personality and Individual Differences, 38*, 177–186.
- Reason, J. T. (2000). Human error: Models and management. *British Medical Journal, 320*, (7237): 768-770. DOI:
- Redlich, A. D., Ghatti, S., & Quas, J. A. (2008). Perceptions of children during a police interview: A comparison of suspects and alleged victims. *Journal of Applied Social Psychology, 38*, 705–735. <http://dx.doi.org/10.1111/j.1559-1816.2007.00323.x>
- Risinger, D. M. (2007). Innocents convicted: An empirically justified factual wrongful conviction rate. *Journal of Criminal Law & Criminology, 97*, (3), 761-806. DOI:
- Roos, C., & Gow, K. (2007). The effect of emotional arousal on recall and interrogative suggestibility. *Australian Journal of Clinical and Experimental Hypnosis, 35*, (2), 150-168.
- Saks, M. J., & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. *Science, 309*, 892– 895. DOI:10.1126/science.1111565
- Smeets, T., Leppink, J., Jelicic, M., & Merckelbach, H. (2009). Shortened versions of the Gudjonsson Suggestibility Scale meet the standards. *Legal and Criminological Psychology, 14*, 149-155. DOI:10.1348/135532507X190207

- Smith, K., & Gudjonsson, G. H. (1986). Investigation of the responses of 'fakers' and 'non-fakers' on the Gudjonsson Suggestibility Scale (GSS). *Medicine, Science and the Law*, 26, 66–71
- Sondenaa, E., Rasmussen, K., Palmstierna, T., & Nottestad, J.A., (2010). The usefulness of assessing suggestibility and compliance in prisoners with unidentified intellectual disabilities. *Scandinavian Journal of Psychology*, 51, 434-438. DOI: 10.1111/j.1467-9450.2010.00811.x
- Spreen, O., & Strauss, E. (1998). Test of memory malingering (TOMM). In *A compendium of neuropsychological tests: Administration, norms, and commentary (2nd. ed.)*, p. 676-677. New York: Oxford University Press
- SPSS 20.0, SPSS Inc., Chicago, IL
- Suggestible. (n.d.). In *Merriam-Webster Dictionary*. Retrieved from: <http://www.merriam-webster.com/dictionary/suggestibility>
- Tombaugh, T. N. (1996). *TOMM: Test of Memory Malingering*. North Tonawanda, NY: Multi-Health Systems.
- Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, 9(3), 260-268.
- Walker, L.E., Conte, C., & **Grabner, S.S.** (2014). Women uniquely vulnerable in criminal justice system. *The National Psychologist*, 23, (12). Accessible at:

<http://nationalpsychologist.com/2014/12/women-uniquely-vulnerable-in-criminal-justice-system/102697.html>

Wallace, D. B., & Kassin, S. M. (2012). Harmless error analysis: How do judges respond to confession errors? *Law and Human Behavior*, *36*, 151–157.
DOI:10.1037/h0093975.

Wechsler, D. (1999). Wechsler Abbreviated Scale of Intelligence Manual. San Antonio, TX: Harcourt Brace & Company.

Woolston, R., Bain, S. A., & Baxter, J. S. (2006). Patterns of malingering and compliance in measures of interrogative suggestibility. *Personality and Individual Differences*, *40*, 453–461. DOI:10.1016/j.paid.2005.10.010

Ziegler, E.A., & Boone, K. B. (2013). Symptom invalidity on neuropsychological testing. In Arnett, P.A. (Ed.), *Secondary influences on neuropsychological test performance*. New York, NY: Oxford University Press.

APPENDICIES

APPENDIX A

Previous Research Group Instructions

Baxter and Bain (2002): “Faking” Group Instructions:

“The interviewer is going to interview you about the content of two stories you’ll hear. What I’d like you to do is to role-play being a suspect in a criminal investigation who may get off the hook if they can convince the interviewer that they are very gullible or very susceptible to pressure, that they are likely to accept whatever is said to them uncritically, and are therefore an unreliable witness. However, please do not go in for amateur dramatics or answer wildly. Try to concentrate on appearing gullible—but not stupid—by giving the interviewer whatever she seems to want. The interviewer doesn’t know who is being given this information and who isn’t, so it’s important that you don’t let her know what I’ve said to you.” (p. 221)

Woolston, Bain, and Baxter (2006): “Misled” Group Instructions

“You are going to be interviewed regarding the content of a story that you will be told. You should attempt to play the role of someone who is a suspect in a criminal investigation. You should try to comply and go along with everything the interviewer says to you, as though you accept everything that is said to you without question. The interviewer is unaware of the instructions that you have been given, so please do not let them know what you have been told.” (p. 456)

Boon, Gozna, and Hall (2008): “Faking Bad” Group Instructions:

“Your role is that of a suspect in a police investigation. Acting in that role you are about to take a memory test as part of that investigation. Your role is to behave to the best of your ability as a suspect would who believed that she/he would be more likely to be found innocent if she/he appeared highly gullible, i.e. easily led and likely to accept anything that was said to them. The interviewer does not know I’m telling you this. Try to convince the interviewer into believing that you are highly gullible. Try to do this as if your freedom depended upon it i.e. try not to give the interviewer any clue that you are acting.” (p. 267-268)

Hansen, Smeets, and Jelcic (2010): “Misled” Group Instructions

“You are going to be interviewed regarding the content of a story that will be told. You should attempt to play the role of someone who is a suspect in a criminal investigation. You have the chance of getting off the hook if you can convince the interviewer that you are easily influenced. You should try to trick the interviewer into believing that you are susceptible to accepting leading questions, that is,

questions that may assume or imply information that is not completely true or is misleading. You will also be given a couple of statements with which you can either agree or not agree. Again your chances to get off the hook rise, if you respond in a certain way. Let the interviewer believe that you are eager to please other people and you try to avoid conflicts with others. The interviewer is unaware of the instructions that you have been given, so please do not let him know what you have read.” (p. 224)

Current Study: Feigning Group Instructions

“You are going to be given multiple assessments today. You should attempt to play the role of someone who is a suspect in a criminal investigation. You have the chance of getting off the hook if you can convince the interviewer that you are easily influenced by pressure from other people. You should try to go along with everything the interviewer says to you, as though you accept everything that is said to you without question. However, please do not go in for amateur dramatics or answer wildly. Try to concentrate on appearing gullible—but not stupid—by giving the interviewer whatever they seem to want. If you are determined to be the most credible faker in the research study, you will be compensated with a \$50 Publix Gift card. The interviewer is unaware of the instructions that you have been given, so please do not let them know what you have read here. When you have finished reading this letter, please fold it and return it to the envelope.”

Current Study: Control Group Instructions

“You are going to be given multiple assessments today. Please complete the assessments as you normally would. You should attempt to complete the assessments to the best of your ability. The interviewer is unaware of the instructions that you have been given, so please do not let them know what you have read here. When you have finished reading this letter, please fold it and return it to the envelope.”

APPENDIX B
Recruitment Flyer

**Attention Nova Southeastern
Undergraduate Students!!**

We are looking for undergraduate students to participate in a research study entitled "*Evaluation of Forensic Assessment Tools.*"

Students must be currently enrolled at Nova Southeastern University as an undergraduate student, speak and read English fluently, and must be over the age of 18.

The study should take about 1.5 hours to complete, and participants will receive a \$20 Publix gift card after participating as a thank you. If you are interested, please email sg1543@mynsu.nova.edu to set up a date and time to participate. Thank you!

APPENDIX C

Informed Consent

Consent Form for Participation in the Research Study Entitled

Evaluation of Forensic Assessment Tools

IRB protocol #

Principal investigator

Stephen Grabner, M.S.
3301 College Ave.
Fort Lauderdale, FL 33314
(954) 262-5705

Co-investigator

Jonathan Shook, M.S.
3301 College Ave.
Fort Lauderdale, FL 33314
(954) 262-5705

Co-investigator

David Shapiro, Ph.D
3301 College Ave,
Fort Lauderdale, FL 33314
(954) 262-5705

For questions/concerns about your research rights, contact:
Human Research Oversight Board (Institutional Review Board or IRB)
Nova Southeastern University
(954) 262-5369/Toll Free: 866-499-0790
IRB@nsu.nova.edu

Site Information

Nova Southeastern University

College of Psychology

3301 College Ave,

Fort Lauderdale, FL 33314

What is the study about?

You are invited to participate in a research study aimed to better understand how responses on four different assessment tools used by psychologists in forensic evaluations could be related.

Why are you asking me?

We are inviting you to participate because you are an undergraduate student at Nova Southeastern University, you are over the age of 18, and you speak and write English fluently.

What will I be doing if I agree to be in the study?

Over the course of approximately two hours you will be answering a set of four assessment instruments given to you by a research team member. These assessments include two paper-and-pencil assessments which require you to answer a number of questions about yourself and how you view the world. The other two assessments are memory tests that are administered verbally and visually by a research assistant.

Is there any audio or video recording?

This research project will include audio recording as part of the two memory assessments given in order to accurately score your responses. This audio recording will be available to be heard by the lead researchers, personnel from the IRB, and research assistants. When listening to the recording, research assistants will use earphones to ensure your privacy. The recording will be kept securely in Dr. Shapiro's office in a locked cabinet and will not be shared with individuals that are not involved in the study. The recording will be kept up to 36 months from the end of the study. The recording will be destroyed after that time by deleting all copies of the recording. Because your voice will be potentially identifiable by anyone who hears the recording, your confidentiality for things you say on the recording cannot be guaranteed, although the researcher will try to limit access to the tape as described in this paragraph.

What are the dangers to me?

Risks to you are minimal, meaning they are not thought to be greater than other risks you experience every day. Being recorded means that confidentiality cannot be promised. However, the precautions regarding privacy of the statements that you make, as described above, will be taken to reduce any risk others hearing your responses. Reading some of the questions in the assessments may bring back unhappy memories. If this happens, the research assistant will try to help you. If you need further help, they will suggest someone you can see in connection with the university through the student counseling center. If you have questions about the research, your research rights, or if you experience an injury because of the research, please contact Dr. Shapiro at (954) 262-5894. You may also contact the IRB at the numbers indicated on page 1 with questions about your research rights.

Are there any benefits to me for taking part in this research study?

There are no benefits to you for participating at this time.

Will I get paid for being in the study? Will it cost me anything?

There are no costs to you to participate. You will be compensated through a \$20 Publix gift cards given after the completion of the assessments.

How will you keep my information private?

The questionnaires will not ask you for any information that could be directly linked to you, and the recordings will be stored in a locked box. All participants will be assigned a code to replace their name, and all information will be stored under that code to minimize the risk of specific participant identification. As mentioned, the recordings will be destroyed 36 months after the study ends. All information obtained in this study is strictly confidential unless disclosure is required by law. The IRB, regulatory agencies, or Dr. Shapiro may review research records.

What if I do not want to participate or I want to leave the study?

You have the right to leave this study at any time or refuse to participate. If you do decide to leave or you decide not to participate, you will not experience any penalty or loss of services you have a right to receive. If you choose to withdraw, any information collected about you **before** the date you leave the study will be kept in the research records for up to 36 months from the conclusion of the study and may be used as a part of the research.

Other Considerations:

If the researchers learn anything which might change your mind about being involved, you will be told of this information.

Voluntary Consent by Participant:

By signing below, you indicate that

- this study has been explained to you
- you have read this document or it has been read to you
- your questions about this research study have been answered
- you have been told that you may ask the researchers any study related questions in the future or contact them in the event of a research-related injury
- you have been told that you may ask Institutional Review Board (IRB) personnel questions about your study rights
- you are entitled to a copy of this form after you have read and signed it
- you voluntarily agree to participate in the study entitled *Evaluation of Forensic Assessment Tools*

Participant's Signature: _____ Date: _____

Participant's Name: _____ Date: _____

Signature of Person Obtaining Consent: _____

Date: _____

APPENDIX D

Demographics Questionnaire

Please complete the following Demographic Questions

Age: _____ Gender: _____

Race/Ethnicity: _____

Level of education: _____