THE CONSTRUCTION AND VALIDATION OF THE FULLERTON ONTOLOGICAL CONFUSION SCALE

A Thesis

Presented to the

Faculty of

California State University, Fullerton

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Psychology

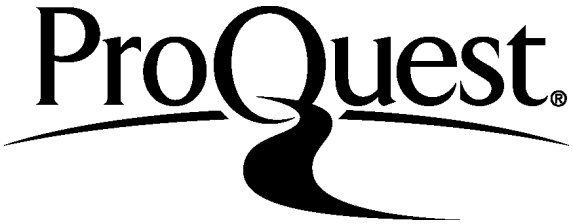
By

Fred Joseph Pasquarella

Thesis Committee Approval:

Chair: Kathleen S. J. Preston, Department of Psychology
Jessie J. Peissig, Department of Psychology
Jessica Tessler, Department of Psychology

Fall, 2015



ABSTRACT

The Fullerton Ontological Confusion (FOC) scale was constructed to address the issues of inconsistent theory and inadequate measurement regarding paranormal, superstitious, magical, and supernatural (PSMS) beliefs. For the FOC scale, PSMS beliefs were defined as confusions of ontology, or the misattribution of core knowledge belief categories. The FOC was empirically evaluated under an Item Response Theory framework using the nominal response model (NRM) to estimate item parameters and the Wald test to evaluate within-item category variation. Using the NRM and the Wald test, the FOC scale items were revised to yield a measure that was optimally formatted and informative. Correlational analysis was used to validate the FOC scale by testing the hypothesized relationships to theoretically related and unrelated constructs. The FOC scale was found to have a good degree of validity with most of the testable validation hypotheses being supported. Compared to the available existing measures, the FOC scale could be used as a more pure and informative measure for PSMS beliefs.

TABLE OF CONTENTS

ABSTRACT i

LIST OF TABLES v

LIST OF FIGURES vi

ACKNOWLEDGMENTS vii

Chapter

1. INTRODUCTION 1

Introduction 1
Review of the Current PSMS Measures 2
The Lack of a Coherent Conceptual Definition 8
PSMS Beliefs as Core Knowledge Confusions 10
Item Response Theory 15
Polytomous IRT 18
Nominal response model 18
Purpose and Hypotheses 22

2. STUDY 1: CONSTRUCTION 25

Methods 26
Participants 26
Material and Procedures 28
Results 29
Item Estimation 32
Estimation After First Revision 39
Estimation After Second Revision 41
Estimation After Third Revision 43
Discussion 48

3. STUDY 2: VALIDATION 50

Methods 52

Participants 52
Material and Procedures 54
FOC Scale 54
R-PBS 54
REI 55
Religious Commitment Inventory 55
Crown-Marlowe Social Desirability—Short Form C 56
Illusory Face Perception Task 56
Results 57
Discussion 61

4. GENERAL DISCUSSION & CONCLUSION 64

General Discussion 64
Conclusion 66

APPENDICES 69

A. INFORMED CONSENT 69
B. THE FULLERTON ONTOLOGICAL CONFUSION SCALE ORIGINAL ITEM POOL 71
C. THE 26-ITEM FOC SCALE 73
D. THE FULLERTON ONTOLOGICAL CONFUSION SCALE 74
E. THE REVISED PARANORMAL BELIEF SCALE 78
F. 10-ITEM RATIONAL-EXPERIENTIAL INVENTORY 79
G. THE RELIGIOUS COMMITMENT INVERNTORY 80
H. MARLOWE-CROWNE SOCIAL DESIRABILITY SCALE, SHORT FORM C 81
I. EXAMPLE STIMILI FOR THE ILLUSORY FACE PERCEPTION TASK 82

REFERENCES 83

LIST OF TABLES

Table		Page
1.	CBDs and Wald Test Statistics for the 26-Item FOC Scale.	34
2.	CBDs and Wald Test Statistics for the FOC Scale After First Revision.	40
3.	CBDs and Wald Test Statistics for FOC Scale After Second Revision.	42
4.	CBDs and Wald Test Statistics for FOC Scale After Third Revision.	44
5.	Discrimination and Difficulty Parameters for 2PLM Items.	45
6.	Final Scale Items with Revised Category Length and Optimal Model Type.	47
7.	Correlations for the FOC θ Scores and the Validity Criteria Measures.	60

LIST OF FIGURES

Figure		Page
1.	An example graph of a 2PLM	16
2.	An example graph of a polytomous IRT model	20
3.	Category response curves: FOC Scale Item 19. (B) Item and category information functions: FOC Scale Item 19	35
4.	Category response curves: FOC Scale Item 21. (B) Item and category information functions: FOC Scale Item 21	37
5.	Category response curves: FOC Scale Item 23. (B) Item and category information functions: FOC Scale Item 23	38
6.	Category response curves: FOC Scale Item 3. (B) Item and category information functions: FOC Scale Item 3	39
7.	Category response curves: FOC Scale Item 12. (B) Item and category information functions: FOC Scale Item 12	41
8.	Item characteristic curves: FOC Scale Item 7 and 12. (B) Item information functions: FOC Scale Item 7 and 12	43
9.	Item characteristic curves: FOC Scale Item 5, 6, 7, 12, and 17. (B) Item information functions: FOC Scale Item 5, 6, 7, 12, and 17	45
10.	Relative Test Information Function for the final FOC Scale	46

ACKNOWLEDGMENTS

I want to thank my thesis advisor and mentor, Dr. Kathleen Preston for being an incalculable source of information, guidance, and dry sarcasm that helped make the entire process seem that much less daunting. Thank you for your expertise and the abundant resources you made available to me throughout the process. I want to also thank my thesis committee members Dr. Peissig and Dr. Tessler for providing valuable feedback and assistance. To my entire thesis committee, whether it was through our personal interactions or our interactions as teacher and student, I thank you for being an invaluable part of my journey as a graduate student.

I would like to also thank my mother for being a tacit safety net during my academic career—allowing me piece of mind so I may pursue my intellectual interests unencumbered by the threat of financial hardship.

Finally, I would like to thank the person who has been by my side throughout my entire intellectual journey. My best friend, my partner in crime, and the love of my life—thank you Allie. Thank you for your constant support and encouragement. Thank you for being silly when needed and serious when necessary. Thank you for everything.

CHAPTER 1

INTRODUCTION

Introduction

Three in four Americans believe in the paranormal (Moore, 2005). This headline screams out from a 2005 Gallup Poll in which over 75% of Americans were found to hold at least one paranormal belief to be true. Beliefs in paranormal phenomena are surprisingly prevalent today and have been a part of humanity for millennia (Vyse, 1997). Indeed, 54% of the United States population believe in psychic healing, close to 50% believe in extra sensory perception, almost 40% believe that haunted houses exist, over 30% believe in ghosts or spirits, and around 30% believe in telepathy and clairvoyance (Moore, 2005; Newport & Strausberg, 2001).

Whether employed for explanatory purposes or for some rudimentary stress management, belief in the paranormal has been with humanity from its early origins. With its longevity in human culture and today's current levels of belief, one may ask why these beliefs are so enduring. In fact many questions could be raised about paranormal beliefs, and they have. Paranormal, superstitious, magical, and supernatural (PSMS) beliefs have been studied scientifically for over 100 years (Lindeman et al., 2008). However, the findings have been largely disconnected due to several factors, most notably the lack of a clear conceptual definition. The terms for PSMS beliefs are often employed or discussed as distinct phenomena; however, they are consistently used

interchangeably and have considerable conceptual overlap (Lindeman & Svedholm, 2012). Researchers tend to use the term that best fits their line of research while overlooking the analogous relationship between the terms, for instance, focusing on PSMS beliefs as magical thinking when studying clinical disorders, such as schizophrenia or schizotypal personality disorder.

Therein lies the problem. Findings from the varied lines of research remain in their specific areas and contexts, where a larger interconnected relationship may exist. In order to ask and attempt to answer any meaningful questions regarding PSMS beliefs, researchers need to establish the parameters of the construct and build empirical relationships. This can only be achieved with a valid and reliable measure for PSMS beliefs. However, the current measures for PSMS beliefs are substantially lacking (Hartman, 1999; Lawrence, 1996) and due to their varied use further divide the literature. A clear measure of PSMS beliefs may offer a much-needed tool to study this integral part of the human experience. It is the purpose of the present research to construct an informative and valid measure for PSMS beliefs that may help merge the literature and be used to answer future inquiries into the nature and correlates of PSMS beliefs.

Review of the Current PSMS Measures

There are four commonly used measures for PSMS beliefs; the Magical Ideation scale (Eckblad & Chapman, 1983), the Australian Sheep-Goat scale (Thalbourne & Delin, 1993), the Anomalous Experiences Inventory (Gallagher, Kumar, & Pekala, 1994), and the Paranormal Belief Scale (Tobacyk & Milford, 1983) and its later revision (Tobacyk, 2004). The current measures have varied use and limited overlap. This may be

due to the specific nature of each measure and the general lack of a clear conceptual definition for PSMS beliefs.

The Magical Ideation scale (Eckblad & Chapman, 1983) is a measure for PSMS beliefs that is primarily used in the context of psychopathology. PSMS beliefs in the context of this scale are akin to extreme misperceptions of causality. With the focus on psychopathology, many of the items are extreme in terms of item difficulty, that is, that they would be very hard to endorse for members of the general population. For example, "I have had the momentary feeling that someone's place had been taken by a look-alike." Indeed, Eckblad and Chapman (1983) noted significant issues with skewness and non-normality for the scale. This focus on psychopathology limits the Magical Ideation scale as a general measure for PSMS beliefs.

The Australian Sheep-Goat scale (Thalbourne & Delin, 1993) is a measure for PSMS beliefs focused specifically on belief in psychical abilities, i.e. extrasensory perception, precognition, and telepathy. It is commonly used in its 18-item true/false form (Thalbourne, 1995). Most of the items concentrate on the individual's perceived psychical abilities. For example, "I believe I am a psychic." Thalbourne (2001) has noted issues with the psychometric properties of the forced-choice version as well as reporting rather low correlations to other measures of PSMS beliefs. This may be due to the extreme nature of some of the items or the narrowness of the definition, which only includes psychical abilities.

The Anomalous Experiences Inventory (AEI; Gallagher et al., 1994) is a rather lengthy inventory for a wide range of PSMS beliefs and experiences. It is comprised of 70 items assessing paranormal belief, perceived paranormal ability, paranormal

experiences, fear of the paranormal, and general drug use. The measure is broad in scope and includes numerous popular PSMS beliefs, e.g. belief in out-of-body experience, human auras, poltergeist, aliens, fairies, etc. The expansiveness of the measure invites issues with measurement precision. The inclusion of separate emotional or cognitive types of responses obfuscates the desired construct being measured. Moreover, there seems to be no strict criteria for inclusion of the beliefs in the AEI.

Whether it is because they are too amorphous (i.e., the AEI), too narrow (i.e., the Australian Sheep-Goat scale), or too centered on psychopathology (i.e., the Magical Ideation scale), these measures are rarely employed in research outside of the author's research domain. The notably exception is the Paranormal Belief Scale (PBS) and its revised version (R-PBS). Collectively, these two measures are by far the most commonly used measure for PSMS beliefs.

A large amount of research has been conducted utilizing the PBS and the R-PBS as measures for PSMS beliefs (Tobacyk, 2004). These measures have been used to develop a nomological network of relationships between PSMS beliefs and demographic characteristics, personality traits, information processing tendencies, critical thinking ability, anxiety, and psychopathology. Additionally, the R-PBS has been integral to the construction and validation of any new conceptual definitions (Lindeman & Aarnio, 2007; Lindeman et al., 2008; see also Barber, 2014; Riekki, Lindeman, & Lipsanen, 2013; Svedholm & Lindeman, 2013; Svedholm, Lindeman, & Lipsanen, 2010). However, its pervasiveness in the literature is most likely due to a paucity of acceptable alternatives than its merit. There are significant questions concerning the scales construction and validation.

The construction of the PBS and its later revision were spurred by the lack of a clear conceptual definition, insufficient empirical evaluations in comparison to the prevalence of PSMS beliefs in society, and the deficiency with the contemporary measures (Tobacyk & Milford, 1983). As such, Tobayck (2004) defined PSMS beliefs as beliefs that fall under three criteria: 1) that they are inexplicable in terms of current science; 2) their explicability is only achieved through major revisions in the basic limiting principles of sciences; 3) that they are incompatible with normative perceptions, beliefs, and expectations of reality.

Initial factor analysis of the PBS indicated a 13-factor solution, of which, Tobayck selected seven that "appeared meaningfully interpretable" (1983, p. 1030). Twenty-five items "that clearly reflected the theme" and possessed the highest factor loading were selected to comprise the measure. The seven factors, from the greatest total variance accounted for to the least, were dubbed traditional religious belief, psi beliefs, witchcraft, superstitions, spiritualism, extraordinary life forms, and precognition. Interestingly, from the traditional beliefs subscale, the validation sample (n = 391) was heavily religious with an average score of 4.24 on a 5-point scale. The heavily religious sample may have confounded the factor analysis and biased the scale validation. That is, because of the heavily religious sample, the scale may function best as a measure for paranormal beliefs amongst religious individuals. Moreover, a more representative sample may have provided differences in factor structure and reliability. The author seemed to have overlooked this issue.

Test-retest reliability was deemed acceptable, however, the sample used comprised only twenty-five individuals. Construct validity was achieved with concurrent

and discriminant measures based on theoretical relationships. However, the sample sizes for these comparisons ranged from $n = 51$ to $n = 96$.

Tobacyk (2004) revised the PBS by changing the range of the response options from five to seven and editing certain items that were considered culturally biased or conceptually vague. Although key changes to the measure were made, Tobacyk did not provide any further statistical evidence for reliability and validity. He addressed the current criticisms with this sweeping statement, "Although there has been some disagreement about the nature and number of the belief dimensions assessed by the R-PBS…this scale appears to be a conceptually and psychometrically satisfactory measure of paranormal beliefs" (p. 97).

There are several criticisms regarding the R-PBS. For instance, the face validity of the items used in the R-PBS is questionable. Tobacyk condensed the conceptual definition of paranormal beliefs to ones that violate Broad's (1953) basic limiting principle of science, that is, any belief that contradicts the current scientific consensus is paranormal. However, certain items do not meet this criterion. Most notably are the items that comprise the extraordinary life forms; "The abominable snowman of Tibet exists," "The Loch Ness monster of Scotland exists," and "There is life on other planets." All three of these items are technically scientifically possible, however improbable. Moreover, with the substantial number of possible habitable planets and the lower limit of biological life, it is entirely possible that life exists on other planets. Indeed, most scientists would agree that life on other planets is entirely possible (Lawrence, 1996).

The main item for the Witchcraft factor states, "Witches do exist." This item is vague and could suffer from differential interpretations. People who recognize

themselves as witches do exist just as people who call themselves magicians exist. This does not mean that there is anything supernatural about the individual or the fact of their existence. Additionally, the scale suffers from deep culturally confounding issues. Many common superstitions such as ‘unlucky number 13’ and references to black cats are used, as well as many beliefs that are more folklore than anything else (e.g., the Loch Ness Monster). These items are not culturally, or generationally, transferrable. Indeed a recent Harris Poll found that only 12% of Americans believed that the number ‘13’ is unlucky and 14% believed that breaking a mirror is bad luck (Corso, 2014), two of the three items listed as the Superstition factor in the R-PBS. These glaring issues with face validity have prompted many researchers to edit, adjust and add items to the R-PBS (e.g. Hergovich & Arendasy, 2005; Lindeman & Saher, 2007; Lindeman & Svedholm, 2012).

A reoccurring criticism of the R-PBS is that it is too narrow to measure the breadth of PSMS beliefs (Lawrence, 1996; Lindeman & Saher, 2007; Wiseman & Watt, 2004). It is questionable that seven distinct factors can be accurately assessed with only 26 items. Each subscale references a limited number of specific beliefs that may not reflect the full range of the belief. Moreover, it may lack key information to help distinguish the varying levels of the latent construct. For example, considering two of the three superstition subscale items are only endorsed by roughly a tenth of the population, only the most extreme scores for superstition would be captured.

Another contentious issue for the R-PBS is its factor structure (Hartman, 1999; Lawrence, 1996). Tobacyk (2004) insists on the 7-factor solution while other researchers have suggested a 5-factor solution (Lawrence, 1996) or a 2-factor solution (Lange, Irwin, & Houran, 2000). Moreover, Lange, Irwin, and Houran (2000) found evidence that the

measure is not additive and is heavily influenced by differential item functioning between gender and age. The multidimensionality of the R-PBS may simply be an artifact of imprecise conceptualization, cultural bias, narrowness of the items, or the reliance on factor analysis for scale creation (Clark & Watson, 1995). Tobacyk (1995) has even agreed with a large part of the scale's criticisms. Despite these shortcomings, the R-PBS continues to be the prevailing measure for PSMS beliefs (Goulding & Parker, 2001), most likely due to the lack of a sound alternative.

The absence of reliable and valid measures may be the cause of numerous inconsistencies in the literature. For example, analytical thinking was found to be positively, negatively, and unrelated to PSMS beliefs (Svedholm & Lindeman, 2012). The relationship between PSMS beliefs and religiosity has also received mixed results with evidence for positive relationships (Smith & Simmonds, 2006), negative relationships (Aarnio & Lindeman, 2007), and a curvilinear relationship (Barber, 2014). The issues regarding the relationship between religiosity and PSMS beliefs are discussed further in a later section. The lack of an adequate measure for PSMS beliefs bring into question previous findings and greatly hinders future research. It is likely that the absence of sound measures for PSMS beliefs is due to poor conceptual definition as a sound theoretical backing is the crux of good scale construction (Clark & Watson, 1995).

The Lack of a Coherent Conceptual Definition

Lindeman and Svedholm (2012) reviewed the last two decades of English, peer-reviewed articles using human subjects with the key words paranormal, supernatural, magical, and superstitious beliefs in the title. They found that the majority of assessment methods targeted the same concepts, however, those concepts were often ambiguously

defined by the authors and seemed to be rooted more in etymology than a coherent theoretical definition (see Lindeman & Svedholm, 2012, p. 246).

The conceptual overlap in the literal definition of these belief types is apparent. Magical beliefs are rooted in the phenomena of magical thinking in which one's beliefs "transcend the usual boundary between the mental/symbolic and physical/material realities" (Nemeroff & Rozin, 2000, p. 5). Paranormal is a designation for "supposed psychical events and phenomena such as clairvoyance or telekinesis whose operation is outside the scope of the known laws of nature or of normal scientific understanding" (Paranormal, n.d.). Superstitious beliefs are beliefs "credited with supernatural efficacy " (Superstitious, n.d.). Supernatural beliefs are beliefs "belonging to a realm or system that transcends nature, as that of divine, magical, or ghostly beings; attributed to or thought to reveal some force beyond scientific understanding of the laws of nature" (Supernatural, n.d.). This overlap is mirrored in much of the research into PSMS beliefs (Lindeman & Svedholm, 2012).

The separate PSMS beliefs coalesce in the literature generally as unfounded beliefs. Some simplistic definitions are that PSMS beliefs are irrational false beliefs or incorrect assessments of external reality (Beck & Forstmeier, 2007). Building on these early definitions is the notion that PSMS beliefs are violations of scientific standards. The most commonly evoked definition for PSMS beliefs draws from Broad's (1953) basic limiting principles. Broad's four principles address the establishment of cause and effect, the limited ability of mental processes on matter, the dependence of the mind on the physical brain, and the limited process of knowledge acquisition. Violation of these principles would be a violation of the "fundamental and scientifically founded principles

of nature" (Lindeman & Svedholm, 2012, p. 243). PSMS beliefs would then be anything that violates these principles (Tobacyk, 2004).

The common thread is that PSMS beliefs are beliefs that are unfounded. However this definition is too vague. For example, the belief that the sun revolves around the earth is not based in fact, however, it is hardly paranormal. The lack of an overarching conceptual definition has resulted in a largely dissociated body of literature as well as the impacting the validity of the current measures for PSMS beliefs.

PSMS Beliefs as Core Knowledge Confusions (CKCs)

After an extensive review of the literature, Lindeman and Svedholm (2012) concluded that there is no justifiable reason to treat PSMS beliefs as distinct concepts. To overcome these issues with conceptual clarity, Lindeman and Aarnio (2007) proposed a more precise definition for PSMS beliefs, namely, that PSMS beliefs are categorical errors in which the attributes of core knowledge of the physical, psychological, and biological world are confused. Building upon Chi, Slotta, and de Leeuw's (1994) ontological attributes, Lindeman et al. (2008) based CKCs in the misclassification of physical, biological, and psychological attributes. Ontological attributes of the physical world are that of material objects with an independent existence that can only interact though direct contact with other material objects derived from natural processes. Ontological attributes of the biological world include notions that entities may live, grow, become ill, and die as well as having a distinction between living and non-living entities. The psychological world is one comprised of thoughts, intentions, and emotions in which animate beings are intentional agents. Lindeman et al. (2008) define, "paranormal beliefs as beliefs in physical, biological or mental phenomena which feature the core attributes of

another of the three [ontological] categories" (p. 1312). CKCs thus conflate core knowledge categories and apply reasoning across categories, which results in common PSMS beliefs. Besides offering precise inclusion criteria, the CKC framework provides developmental and cognitive explanations for PSMS beliefs.

Lindeman and Aarnio (2007) point to three major types of early knowledge found in children, specifically, that of intuitive physics, intuitive psychology, and intuitive biology. Intuitive physics is the basic understanding that material objects have volume and exist independently in space. Intuitive psychology is the basic understanding that animate objects are intentional agents. Intuitive biology references the basic notions of health and illness. As the child develops, these intuitive understandings progress into core knowledge structures. Therefore, core knowledge is learned naturally without instruction during early development. It is important to note that PSMS beliefs are not equivalent to child-like thinking or cognitive deficits, instead, it is related to a greater reliance in an intuitive style of thinking (Lindeman et al., 2008).

CKCs are rooted in the intuitive thinking style of the dual-process theory of reasoning (Lindeman, 2011; Svedholm & Lindeman, 2013). In the dual-process model of thinking, the process in which individuals reason can be understood as the interplay of two systems, sometimes dubbed system 1 and system 2 (Kahneman, 2011), heuristic-systematic (Chaiken, 1980), or experiential-analytical (Epstein, Pacini, Denes-Raj, & Heier, 1996). These systems help individuals navigate the information dense world. The first system, known as the intuitive or experiential system is automatic, preconscious, holistic and associationistic (Epstein et al., 1996). This type of thinking is low in cognitive effort and aids in quick decision-making that is stereotypic, emotion-driven,

and based on subjective experiences. System 2, known as the rational or analytical system, is intentional, logical and conscious. This type of thinking requires much effort and aids in delayed action responses (Kahneman, 2011).

In a large-scale study, Lindeman and Aarnio (2007) found that intuitive thinking predicted PSMS beliefs better than other commonly attributed factors, such as the desire for control and emotional instability. Similar results were found in a follow-up study (Svedholm & Lindeman, 2013). What is important to note is that the paranormal beliefs are associated with the intuitive system and no consistent evidence has been found linking paranormal beliefs or CKCs with a malfunctioning analytical system (Aarnio & Lindeman, 2005; Auton, Pope, & Seeger, 2003). Indeed, Hergovich and Arendasy (2005) found no relationship between critical thinking and paranormal belief.

CKCs have been consistently found to be related to paranormal beliefs. In the seminal study for CKCs, Lindeman and Aarnio (2007) found the paranormal believers tended to *mentalize* matter, *physicalize* mental, and *biologize* mental concepts more than skeptics. That is, they tended to attribute mental characteristics to matter, physical characteristics to mental processes, and biological characteristics to mental processes. Moreover, paranormal believers assigned more purpose to natural, artificial, and random events than skeptics. Furthermore, Lindeman and Aarnio found that the best predictor of paranormal believers were CKCs.

Lindeman and Saher (2007) found that all types of PSMS beliefs were related to CKCs and that paranormal believers tended to explain biological processes more in terms of actions of an intentional agent. Lindeman et al. (2008) observed that paranormal believers cognitive responses differed significantly from skeptics when processing core

knowledge violations. Barbor (2014) obtained evidence for cross-cultural validity in an American population with results consistent with previous research. That is, the same positive relationship between CKCs and paranormal beliefs was found in an American sample.

Following from the link between PSMS beliefs and intuition is the connection of PSMS beliefs and false pattern recognitions. Falsely detecting patterns in randomness is an intuitive bias resulting from a reliance on an intuitive style of thinking (Kahneman, 2011). A common manifestation of this intuitive bias within the realm of paranormal beliefs is that of illusory face detection, referred to as pareidolia or apophenia (Shermer, 2008). Indeed, PBS scores were negatively related with the perception of randomness and were associated with a harder time recognizing random from nonrandom patterns (Dagnall, Parker, & Munley, 2007). In a study on illusory face perception, paranormal believers have been found to detect more faces in artifacts than skeptics and non-believers (Riekki, Lindeman, Aleneff, Halme, & Nuortimo, 2013). Moreover, paranormal believers tended to detect more faces whether or not they were actually present.

Despite the anecdotal link between religiosity and paranormal beliefs, there is no strong systematic evidence for a relationship**.** Traditional religious belief is consistently the strongest factor, accounting for the most variance in the R-PBS (Tobacyk & Milford, 1983; Lawrence, 1996). Interestingly, some researchers have tried to link religiosity to PSMS beliefs through comparisons of the traditional religious beliefs subscale of the R-PBS to the composite score of the remaining subscales (Smith & Simmonds, 2006). Not surprisingly, they found that religious people have greater beliefs in the paranormal. The problem rests in the fact that traditional religious belief, as measured in the R-PBS, tend

to represent an entirely different construct than one's religiosity. Certainly, there are components of religious beliefs that can be considered paranormal or supernatural, however, religiosity is primarily a result of social, cultural, and familial factors (Cornwall, Albrecht, Cunningham, & Pitcher, 1986). Religiosity is not simply the belief in a god and an afterlife. Instead it is a multidimensional construct that includes the strength of the feelings toward the beliefs and the behavioral commitment to them (Cornwall et al., 1986). Moreover, strict adherence to religious beliefs often comes at the expense of other paranormal beliefs. Dogmatic beliefs are a knowing suspension of the natural order, that is, an acceptance that, although certain facts about the nature of reality are correct, faith offers a greater, sometimes counterfactual understanding. This is in stark contradiction to other PSMS beliefs in which the individual does not acknowledge that the beliefs they hold are unsound. Indeed, Aarnio and Lindeman (2007) found that highly religious people tended to support less paranormal and superstitious beliefs. Moreover, Barber (2014) found a curvilinear relationship between religiosity and the R-PBS, in that individuals low in religiosity tended to be lower in paranormal beliefs, individuals high in paranormal beliefs tended to be low in religiosity, and individuals with medium levels of paranormal beliefs tended to be higher in religiosity. This may be an artifact of the religiously dominant measure, conceptual overlap, or both.

Although Lindeman and colleagues have provided ample evidence for the relationship of CKCs to PSMS beliefs and related constructs, they have not integrated the knowledge into a measure for PSMS beliefs. Utilizing the theoretical strength of Lindeman's CKC definition of PSMS beliefs, the Fullerton Ontological Confusions (FOC) scale was constructed to address glaring issues in the current measurement

standard. To overcome the psychometric shortcomings associated with Classical Test Theory (CTT), the FOC scale was constructed under an Item Response Theory (IRT) framework. Before discussion of the analysis can take place it is necessary to outline why IRT was used over CTT and give a brief overview of IRT.

Item Response Theory

IRT is a theoretical perspective on psychometrics that places a strong emphasis on item functioning and builds a data driven model for test construction. Nonetheless, a strong theoretical basis is still needed for initial scale construction. For IRT models, a latent trait is assumed to be continuous, in which, individuals have a true location on the trait continuum (Reise & Henson, 2003). This location is called theta (θ). θ is synonymous with CTT's true score, however θ has a probabilistic relationship to the underlying latent trait (Reise & Henson, 2003). IRT models this relationship by applying a logistic equation to the data. For example, Equation 2 displays the two-parameter logistic model (2PLM).

$$P \mid \theta = \frac{\mathbf{1}}{\mathbf{1}+e^{-a_i(\theta-b_i)}} \quad (1)$$

This equation graphs a sigmoidal curve that can vary based on two parameters, item difficulty (b) and item discrimination (a). The item difficulty parameter (b) is often referred to as the threshold parameter and is the point on the latent trait where the probability of endorsing the item is .50 (de Ayala, 2009). These parameters create item characteristic curves (ICCs). ICCs are visual representations of the items discrimination and difficulty parameters in relation to θ, as seen in Figure 1. Thus, an individual with whose standing on the latent trait above this level has a greater likelihood of endorsing that item and vice versa. The item discrimination parameter (a), or slope, is the item's

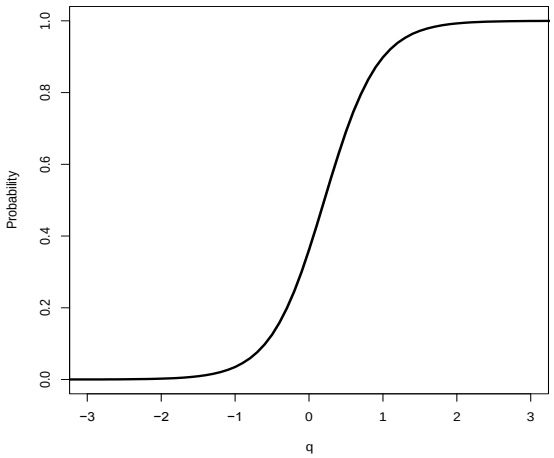


Figure 1. An example graph of a 2PLM.

Note. The probability of endorsing the item for a given θ is represented on the vertical axis and the level of θ is on the horizontal axis.

ability to discriminate between different individuals with similar levels of the latent trait (Reise & Henson, 2003). The sum of the ICCs for all items in a scale yield a test characteristic curve (TCC) that shows how the entire pool of items function over the latent trait. Under an IRT framework, error measurement, and conversely, information for each item varies across θ. To model this relationship item information functions (IIFs) are used. IIFs serve as an index for the range and the amount of information that an item provides (de Ayala, 2009). The more psychometric information provided by an item and the location on θ at which the information peaks indicates where the greatest precision of the measure is for a given θ. Under IRT, information is additive, and so a test information function (TIF) can be produced by adding all IIFs in a given test (Embretson & Reise, 2009). The TIF provides evidence for where on the trait continuum the scale is most precise. Since the TIF is the sum of all IIFs, it indicates the range and amount of information the pooled items have across the latent trait (Hambleton & Jones, 1993).

Information in IRT is synonymous to reliability coefficients in CTT in that they serve as a measure of test precision. A key difference between reliability under CTT and information in IRT is that information is a function of item characteristics and can vary based on a respondents θ, whereas reliability is constant for all respondents regardless of their actual trait level (Reise & Henson, 2003). IRT has two strong assumptions. First is the assumption of unidimensionality. That is, that a single continuous latent ability underlies performance on a test (Hambleton & Jones, 1993). Second is the assumption of local independence. That is, after controlling for the latent trait measured by the test, responses on one item are independent of the responses on other items (de Champlain, 2010).

IRT and CTT are two theories aimed at the attainment of a reliable and valid quantitative measure of a latent trait. CTT has been noted to produce weaker estimates of a latent trait than IRT (e.g., Amarnani, 2009; Borsboom, 2006; Reise & Henson, 2003). Comparing both theories side by side, IRT appears to be superior to CTT for several reasons. Specifically, CTT yields one standard error of measurement for the test that remains constant for all the scores, while IRT gives error estimates across the range of the latent trait (Hambleton & Jones, 1995; Reise, Ainsworth, & Haviland, 2005). IRT provides indications of reliability for each item as well as the test as a whole, IIFs and TIF, respectively. Moreover, reliability estimates in IRT indicate where exactly on the latent trait each individual item and the test are the most accurate (Hambleton & Jones, 1995; Reise et al., 2005). CTT is sample dependent. In order to assess item properties appropriately, it requires a representative sample. IRT is sample independent and does not require a representative sample to assess item properties (Hambleton & Jones, 1995).

Since CTT utilizes sum scores for comparisons of different response types it is test dependent. This is not the case for IRT since θ estimates are test independent (Reise et al., 2005).

Polytomous IRT

IRT can also be utilized for tests with multiple response categories, referred to as polytomous IRT models. Since the current research dealt with polytomous response sets a review of polytomous IRT is in order. Polytomous models have a distinct advantage over dichotomous IRT is that in information in dichotomous models is limited. That is, the more information, or the more peaked the information function, the narrower the range of information over the trait continuum. Polytomous IRT, by virtue of the multiple response categories and information functions, are capable of providing more information over a larger portion of the trait continuum (Ostini & Nering, 2006).

In addition to the probability of responding in a particular category, polytomous models focus on the probability of responding positively or negatively at a given category boundary distinction (Ostini & Nering, 2006). Whereas the ICC in dichotomous IRT regresses the probability of endorsing an item on the trait level, polytomous IRT uses ICC to regress the probability of responding in a category on the trait level (Embretson & Reise, 2009). Although most polytomous IRT models assume order in response categories, the nominal response model (NRM; Bock 1972) does not.

Nominal Response Model

The NRM is a polytomous model in which category responses are not assumed to be ordered. Equation 2 displays the probability of a participant responding in category x ($x = 0 \ldots m_i$) in the NRM (Embretson & Reise, 2009).

$$P_{ix} \mid \theta = \frac{e^{a_{ix}\theta + c_{ix}}}{\sum_{x=0}^{m} e^{a_{ix}\theta + c_{ix}}} \quad (2)$$

In Equation 2, the *a* and *c* parameters represent the slope and the intercept, which are estimated for each response option within an item (Preston, Reise, Cai, & Hays, 2011). To overcome an identification problem during estimation, the *a*s and *c*s are constrained by setting the intercept and slope of the first category option to zero (Preston et al., 2011). The slope represents a function of the change in the log-odds of responding to a particular category option as the trait level changes, whereas, the intercept parameters represent the relative frequency of responses in a particular category (Preston et al., 2011). As seen in Figure 2, each category response option has a category response curve (CRC) indicating where on the latent trait the particular response is psychometrically informative.

If categories are assumed to be ordered then the distinction between two adjacent categories can be modeled as a two-parameter logistic function (Preston et al., 2011). The two adjacent response options can be written as *x* and *x*-1. Equation 3 displays the probability of responding in one of two adjacent categories *x* or x'.

$$P_{ix} \mid x = x \; or \; x' = \frac{1}{1 + e^{-a^*_j \theta + c^*_j}} \quad (3)$$

Here, *j* is a *m*-1 possible response distinction of *x* and *x*-1. The category boundary discrimination (CBD) parameter is represented here as a_j^* which is equal to $a_x - a_x$ -1. This can be understood as the slope of a 2PLM. The intercept is represented here as c_j^*, which equals $c_{x-1}^* - c_x$ (Preston et al., 2011). The CBD parameters (a_j^*), in essence, provide the discrimination for distinguishing between two adjacent category response options. That is, CBD parameters measure the changes in slope from adjacent categories.

Near zero values for CBD parameters indicate that there is no meaningful difference between the two response categories. In other words, individuals are unable to make meaningful distinctions between the response options (Preston & Reise, 2015). Large positive CBD values denote highly informative category response distinctions (Preston et al., 2011).

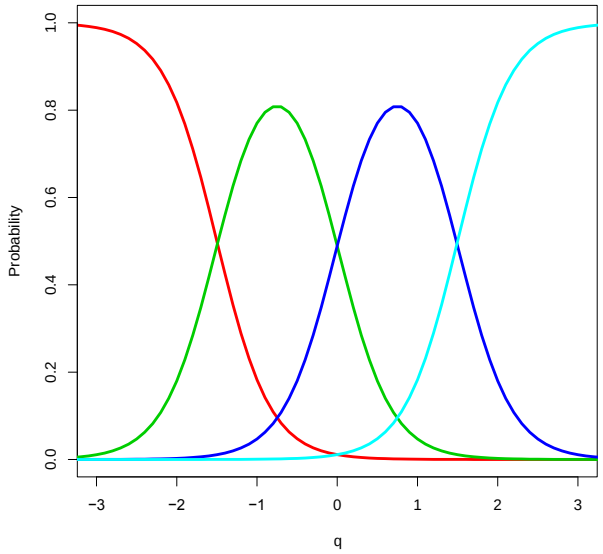


Figure 2. An example graph of a polytomous IRT model.

Note. The probability of endorsing a category for a given θ is represented on the vertical axis and the level of θ on the horizontal axis. Also, there are four response options and three CBDs.

The intercept parameter (c_j^*) can be converted into an intersection parameter that indicates the point on the latent trait where a response in one of two adjacent categories is equally likely (Preston et al., 2011). The intersection parameter is computed by dividing the intercept of two adjacent categories (i.e., c_j^*) by the differences in category slopes (i.e., a_j^*) of the same two adjacent categories (Preston et al., 2011). Thus, the intersection

is the difference in intercept divided by the difference in slope between two adjacent categories. This can be seen in Figure 2 as the point where the CRCs cross.

NRM is in the class of the divide-by-total models (Thissen & Steinberg, 1986). However, all the other divide-by-total models (Thissen & Steinberg, 1986) are nested in the NRM. That is, models like the generalized partial credit model (GPCM; Muraki, 1992) is just a more constrained version of the NRM. Unlike NRM, the GPCM specifies a single slope parameter that varies between, but not within items (Preston et al., 2011). That is, the model allows for the discrimination ability to vary, but the response category distinction is fixed to be equally differentiating and informative (Preston et al., 2011). Since the other divide-by-total models (Thissen & Steinberg, 1986) are nested in the NRM, more constrained models, like the GPCM can be applied when within-item CBDs do not vary significantly (Preston et al., 2011).

With the CBD and intersection parameters, the NRM enables latent trait distinctions for each response category and estimates of information provided by each category (Preston & Reise, 2015). Moreover, the NRM can be used to empirically test the functionality and order of the response categories (Ostini & Nering, 2006; Preston & Reise, 2015; Preston et al., 2011). This is achieved through use of the Wald test statistic (Wald, 1945), which can be used to check statistically significant within-item variations in CBD parameters. Details about how to calculate the Wald test statistic can be found in the Study 1 results section. The Wald test statistic evaluated using the chi-square distribution with degrees of freedom equal to the total number of CBDs minus one (Preston & Reise, 2015; Preston, Reise, Cai, & Hays, 2011). Significant Wald test statistics combined with near zero CBD values is an indication of non-functioning items

(Preston et al., 2011). A non-significant Wald test statistic may also suggest that a more constrained model could be applied without losing psychometric information. The NRM and Wald test statistic allows for an empirical assessment of items geared at obtaining the most optimally functioning items and item response formats.

IRT overcomes many of the shortcomings of CTT and provides sound evidence for scale revision. With CTT there is no empirical way to assess the functioning of individual response categories. Instead, categories are assumed to be ordered and of equal psychometric importance. With polytomous IRT models, specifically the NRM, one is able to assess the psychometric properties of all category responses as well as the item as a whole. Moreover, with the NRM the functioning of the responses can be assessed. An IRT framework allows for an empirical and sound evaluation of the item functionality that supersedes that enabled through a CTT framework (Embretson & Reise, 2009).

Purpose and Hypotheses

The purpose of this thesis is threefold. First, construct a psychometrically informative measure for PSMS beliefs. Second, utilize IRT to empirically revise the FOC and provide more sound evidence for structural validity and reliability. Third, validate the newly constructed FOC scale. Employing Lindeman and Aarnio's (2007) theoretically strong CKC definition of PSMS beliefs, an initial item pool was created. IRT was used to model a probabilistic relationship between item and person characteristics and the latent trait, namely the tendency to make ontological confusions. The NRM was employed to empirically evaluate the psychometric contribution of each category response within the scale items to assess their order and functioning. The Wald test was utilized to test within-item category variation and provide evidence for proper response category option

type and optimal modeling technique. The goal of the scale revision was to retain items that were optimally functional and provided generous amounts of psychometric information over a range of the latent trait continuum.

Assessment of the validity of the FOC scale followed the guidelines laid out by Whitley and Kite (2013) for examining validity evidence. The assessment focused on the content, structural, convergent, discriminant, and substantive validity. Two hypotheses are put forward in regards to convergent validity: 1) Scores on the FOC scale will be positively related to the leading measure for PSMS beliefs and 2) with the tendency to engage in intuitive thinking. Regarding substantive validity, it is hypothesized that individuals who tend to score higher on the FOC scale will falsely perceive more faces in non-face images. Three hypotheses are proposed in regards to discriminant validity: Scores on the FOC will *not* be related to 1) analytical thinking styles, 2) religiosity, and 3) social desirability.

The current research was carried out over two studies. Study 1 was conducted to construct the proposed scale and revise it under an IRT framework. In Study 1, the proposed scale items were administered to a large sample of respondents (n = 978) and the item parameter estimates produced with the NRM were evaluated in terms of psychometric functionality. Items were revised using the Wald test. Study 2 was conducted to validate the empirically revised scale with a smaller sample of respondents (n = 154). Both sets of participants responded to the proposed scale items in identical fashion, i.e. directly following the informed consent portion. Respondents in Study 1 completed a short questionnaire with only the proposed scale items and demographic

questions, while respondents in Study 2 completed a longer questionnaire that included the revised items, validation measures, and demographics.

CHAPTER 2

STUDY 1: CONSTRUCTION

The FOC scale was constructed under the CKC framework proposed by Lindeman and Aarnio (2007). That is, PSMS beliefs were limited to confusions of core knowledge and not culturally distinct knowledge. In the context of the constructed scale, ontological confusions are synonymous with CKCs, but have the added criterion of attributing purpose and intentionality to natural processes. This added criterion of *vitalism*, or an attribution of purpose, is based on findings that attributing purpose and intentionality to biological and physical properties is characteristic of PSMS believers (Lindeman & Saher, 2007). Moreover, mistakenly attributing purpose and intentionality to the biological or physical categories is, in essence, a confusion of core knowledge.

The purpose of the first study was to construct and empirically revise the FOC scale. An initial item pool was created using the aforementioned theoretical framework. The proposed scale was then assessed with IRT analysis by estimating item parameters with the NRM. Estimation using the NRM allowed for an empirically evaluation of not only item functioning, but category response functioning. Moreover, category response options were evaluated and revised based by use of the Wald test statistic (Wald, 1945). The Wald test checks for statistically significant with-item variation in CBD parameters (Preston et al., 2011). Significant Wald test statistics combined with near zero CBD values indicate that the response options are non-functioning (Preston et al., 2011). In

essence, this indicates whether respondents can make meaningful distinctions between the response options and if the response options are functioning as expected, e.g. ordered. A non-significant Wald suggests that a more constrained model could be applied without losing psychometric information. Utilizing the NRM and Wald test statistic, the proposed scale was revised retaining the most optimally functioning items and item response formats.

Methods

Participants

Participants were recruited from Amazon's Mechanical Turk (MTurk; n = 835) and California State University, Fullerton's (CSUF) Sona Systems research participant pool (n = 206)—for a total sample of N = 1041. MTurk is comprised of a highly diverse population of individuals who seek compensation for a variety of tasks in an open forum (Buhrmester, Kwang, & Gosling, 2011). MTurk respondents were compensated $0.50 for their participation. CSUF's Sona Systems participant pool was comprised of mainly undergraduate psychology students who voluntarily chose to participate in research to receive course credit. CSUF students received 1-hour participation credit for their participation.

Prior to analysis, data were screened for accuracy. Participants were removed for failing to respond correctly to any one of three attention check items. These items were placed randomly throughout the questionnaire to ascertain if the participants fully comprehended the items and were paying attention to what was being asked of them. A total of 63 (6%) participants were removed due to failure to respond correctly to any one of the attention check items.

No strict rules for the minimum number of participants needed for IRT model calibrations are available. However, a minimum of *n* = 600 to *n* = 1500 has been suggested (de Ayala & Save-Bolesta, 1999; DeMars, 2003). The final sample size (*N* = 978) falls within the suggested minimum number of respondents.

The final sample comprised *N* =978 participants. Females made up the majority of the sample at 56.1% (*n* = 549). Age ranged from 18 to 83 with an average respondent age of 35.6 years old (*SD* = 13.85). Participants were from throughout the United States with 36.7% (*n* = 359) from the West, 32.1% (*n* = 314) from the South, 16.3% (*n* =159) from the Mid West, and 13.9% (*n* =136) from the Northeast.[1] The sample was primarily White/Caucasian with 70.8% (*n* = 692) of participants identifying as White, 8.9% (*n* = 69) as Mexican American, 7.5% (*n* = 73) as African American/Black, 7.1% (*n* = 69) as Asian/Asian American, and a small percentage identifying as of mixed ethnicity (1.2%, *n* = 12). Not surprising, participants overwhelming identified with American culture (82.1%, *n* =803) of participants, while only a few participants identified primarily with Latino(a) (6.4%, *n* = 63), European (5%, *n* = 49), or Asian (4.1%, *n* = 40) cultures.

A relatively large percentage of the sample identified as being Non Religious (23%, *n* =225) or Atheist (15.5%, *n* =152). However, the sample still comprised a majority of individuals who identified with either Christian (37.2%, *n* = 364) or Catholic (16.2%, *n* = 158) denominations. The sample was fairly educated with 82.8% of individuals having Some College (40%, *n* =396), College (32.5%, *n* = 318) or a Graduate

[1] Participant residency was based on of self-reported ZIP codes. Untied States regions were labeled based on United States Census Bureau (n.d.) divisions.

Degree (9.8%, *n* = 96)—while only 16.5% (*n* = 161) reported High School (or G.E.D.) as their highest level of education.

Materials and Procedure

The research was approved by California State University, Fullerton's (CSUF's) Institutional Review Board (IRB). Data were collected through the use of an online questionnaire created in Google Forms. Participants were first prompted with an informed consent explaining the details of the study, potential risks and benefits, and their rights as a research participant. A copy of the Informed Consent Form is located in Appendix A. Participants were then asked if they fully understood what was being asked of them and if they agreed to participate. Requisites for participation were that participants must be able to read and understand English and be over the age of 18 years old. After providing consent, the participants were transferred to the online questionnaire, which included the proposed scale and demographic questions. After completion, the participants were debriefed and thanked for their time.

Only the FOC scale and demographic information was administered to participants in Study 1. Demographic questions asked for participant's age, gender, ethnicity, cultural identity, highest level of education, relationship status, sexual preference, religious affiliation, and US postal ZIP code.

The FOC scale contained items (*k* = 40) assessing one's tendency to think paranormally in terms of making ontological confusions. Appendix B displays the initial-item pool for the FOC scale. Participants were instructed:

> The following are statements regarding your personal feeling towards the nature of reality and the physical world. You are being asked to rate the extent to which

you believe or disbelieve each statement. These statements are to be read literally not metaphorically. There are no right or wrong answers. Please answer as honestly as possible. Remember, your answers are anonymous.

Responses were rated with a 5-point Likert-type response format with 1 = *Strongly Disbelieve*, 2 = *Disbelieve*, 3 = *Neither Believe nor Disbelieve*, 4 = *Believe*, and 5 = *Strongly Believe*.

Results

Prior to analysis, category response frequencies for each item were examined. This served as an early indication for poor item functioning. Additionally, for IRT estimation, since parameters estimates should be obtained from a heterogeneous sample, category responses should be relatively evenly spread throughout response options (Embretson & Reise, 2009). Items that possessed extremely high response rates in only one response category, an indication of minimal within-item variability, were dropped from the item pool. Response rates greater than 70% in one category were considered extreme because, even after a collapse of two categories, a 10% minimum response rate would be impossible to attain. After examination, items 37, 39, and 40 were removed due to extreme responds rates in one or a few categories. For example, item 39, "A rock regrets that it cannot move," 80% of participants choose "Strongly Disbelieve." Category response rates are also important during the IRT estimation process. In order to achieve proper estimation during the analysis, category response options that comprised less than 10% were collapsed into the next logically consistent and adjacent category, e.g. "Believe" was collapsed with "Strongly Believe" but not with "Neutral." This ensured a heterogeneous category response pattern.

To assess the assumption of unidimensionality, an exploratory factor analysis was conducted using the Comprehensive Exploratory Factor Analysis (CEFA; Browne, 2014) with polychoric correlations specified. The initial exploratory factor analysis (EFA) provided evidence for several potentially problematic items. Specifically, several items formed a distinct factor unrelated to the larger context. One major distinct factor could be labeled as traditional religious beliefs. This factor was comprised of items 1, 14, 16, 18, 19, 22, 25, and 34. These items were not intended to be an assessment of religious beliefs, however, their specific types of ontological confusions also happened to be rooted in the beliefs of the major monotheistic religions, for example, "There is life after death." Regardless of the item context in terms of the ontological confusion definition, these eight items formed a consistent factor over several possible factor solutions. Although the ratio of the first to second eigenvalues arguably indicated the dominance of a single factor (16.1 to 3.3), the second factor accounted for a relatively modest portion of the total variance (10%). These items were therefore removed from the scale due to the potentially confounding nature of the items and the possible issues with unidimensionality. Further discussion on the justification for removal is located in the Study 1 discussion section.

In addition to the eight religious items, three extreme ontological confusion items were removed (items 2, 7, and 9). For example, item 2 stated, "Thunder is angry." Recall, that one of the four extreme ontological confusion items were removed prior to the EFA due to extreme skew and lack of heterogeneity of responses (e.g., over 80% responding "Strongly Disbelieve"). The other extreme items were satisfactorily heterogeneous in response rate, albeit still skewed. However, these items also formed a distinct factor.

Moreover, there was initial concern over participant misunderstanding of the metaphorical or literal interpretations of the statements. With all these possible confounding factors, the remaining three extreme ontological confusion statements were removed from the scale. What remained was the 26-item FOC scale, as seen in Appendix C.

An EFA was then conducted on the 26-item FOC scale. A one-factor model was specified using polychoric correlations. A common method for determining if a single factor dominants the entire set of items is to examine the ratio of the first to second eigenvalue (Reise & Waller, 1990). A large ratio, e.g. 3:1, is supportive of unidimensionality (Embertson & Reise, 2000). Moreover, the ratio of the first to second eigenvalues serves as an index for the strength of the first dimension (Reise & Waller, 1990). The ratio of the first to second eigenvalue for the one-factor solution was 7.62:1, with the value of the first eigenvalue equal to 11.3. Further support for unidimensionality came from an examination of the proportion of variance accounted for. The first factor should account for a substantial proportion of the variance compared to the other factors. The one-factor solution accounted for 43.5% of the variance. Structural validity was also assessed through an examination of the root mean square error of approximation (RMSEA; Browne & Cudeck, 1992) values. As a factor selection criterion, RMSEA provides an estimate of model misfit while controlling for sample size (Preacher, Zhang, Kim, & Mels, 2013). According to Steiger (1990), RMSEA values between .05 and .10 represent an acceptable fit. For the one-factor model, RMSEA = 0.092 with 90% CI (0.089, 0.095). Overall, the assumption of unidimensionality was met.

Item Estimation

FlexMIRT version 3.0 (Cai, 2013) was used to obtain item parameters estimates for the NRM with marginal maximum likelihood (MML) estimation and 81 quadrature points specified. With the NRM, triangle contrasts were specified to produce CBD parameter estimates with standard errors and a variance-covariance matrix appropriate for the Wald test. The marginal χ^2 values and the local dependence statistics proposed by Chen and Thissen (1997) were examined to check the assumption of local independence. No major instances of local dependence were found to be present.

Items were revised through an evaluation of several criteria; 1) item and category information, 2) the Wald test statistic, and 3) CBD parameter estimates. Item information and overall category and item functionality were visually examined using CRCs, IIFs, and CIFs plots produced in R (Preston, 2014a). The CRCs provided information on the degree of discrimination and difficulty of each response option. IIFs displayed the item information, while the CIFs indicated the information provided from each category response option. Relative IIFs were produced by scaling information by the number of estimated parameters. Extremely low relative item information (i.e., < .25) would be justification for item removal (e.g., Ura, Preston, & Mearns, 2015).

CBD parameters and the Wald test were computed using R (Preston, 2014b). In order to compute the Wald test statistic several steps were undertaken, as put forward by Preston, Reise, Cai, and Hays (2011). First, CBD parameter estimates produced during estimation were compiled in a vector (*A*). Then, orthogonal linear contrasts (*C*) were specified to compare the CBD parameters. The orthogonal linear contrasts compared the first and second CBDs, the average of the first and second with the third CBD, and the

average of the first, second, and third with the fourth. Then, *C* was post-multiplied by *A* to produce λ. Standard errors were derived by pre- and post-multiplication of the variance-covariance matrix of item parameter estimates (Σ) by the contrasts; $\Omega = C\Sigma C'$. The Wald test statistic was then calculated as $W = \lambda\, \Omega\, \lambda'$ and evaluated using the chi-square distribution with degrees of freedom equal to the total number of CBDs minus one (Preston & Reise, 2015; Preston, Reise, Cai, &Hays, 2011).

Table 1 displays the CBD parameters, Wald test statistic, degrees of freedom, and probability value for each item. After an evaluation of the relevant statistics, items 1, 9, 18, 19 were removed. Item 1 was removed due to low CBD values and extremely low overall information (< .2). Item 19 was removed because of a significant Wald test and multiple low CBD parameters. Moreover, three of the four response category options provided very low amounts of information that were not unique from each other. The CRCs, IIF, and CIFs for item 19 can be seen in Figure 3. As seen in Figure 3, the categories representing "Disbelieve," "Neutral," and "Believe" provided identical information in terms of person location of the trait continuum. Furthermore, the near zero values for CBD_2 and CBD_3 indicated that participants could not make a meaningful distinction between those three category response options.

Items 9 and 18 were removed due to a significant Wald test, a negative CBD parameter for the 4th and 5th category response options, and overall low information. Information for item 9 and 18 peaked at .25 and .23, respectively. All category response options, excluding the 1st, provided extremely low and redundant information. Multiple low CBD parameters values were further indication for redundant or non-functioning response category options. Negative CBD parameters can serve as an indication that the

category response options are not ordered as expected (Preston et al., 2011). However, when the CBD value is relatively small, it may just be a further indication that the categories are redundant. The latter possibility informed the decision to retain item 6 despite the negative CBD parameter value.

Table 1

CBDs and Wald Test Statistics for the 26-Item FOC Scale

Item	CBD_1	CBD_2	CBD_3	CBD_4	Wald	*df*	*p-value*
1	1.01	0.54	0.56	0.6	7.149	3	0.067
2	2.47	1.03	0.99	0.54	182.393	3	0.000
3	2.24	0.75	0.83	-	98.134	2	0.000
4	1.15	1.05	0.82	0.4	4.052	3	0.256
5	1.93	0.77	0.64	-	48.199	2	0.000
6	2.12	1.01	0.65	-0.15	98.324	3	0.000
7	1.82	0.64	0.2	-	82.814	2	0.000
8	1.96	1.14	0.87	0.54	60.76	3	0.000
9	1.7	0.29	0.53	-0.36	134.857	3	0.000
10	2.65	0.85	1.11	0.51	242.013	3	0.000
11	2.23	0.88	1.05	-	110.912	2	0.000
12	2.21	0.92	0.54	-	73.791	2	0.000
13	2.53	0.77	0.99	-	163.483	2	0.000
14	1.96	0.98	1.02	0.25	52.791	3	0.000
15	1.48	1.15	0.45	-	11.908	2	0.003
16	2.03	1.27	1.01	-	20.651	2	0.000
17	2.3	0.93	0.85	-	94.433	2	0.000
18	1.35	0.76	0.67	-0.32	31.936	3	0.000
19	1.81	0.29	0.15	-	103.839	2	0.000
20	1.88	0.36	0.75	-	68.64	2	0.000
21	1.52	1.16	1.19	0.34	12.249	3	0.007
22	1.36	1.06	1.16	-	2.147	2	0.342
23	2.51	1.01	1.51	-	115.694	2	0.000
24	1.61	0.95	1.03	0.55	19.957	3	0.000
25	2.78	1.76	1.33	-	48.046	2	0.000
26	1.2	0.99	0.43	-	7.518	2	0.023

Note. Items 3, 5, 7, 10, 11, 12, 15, 16, 17, 19, 20, 22, 23, 25, and 26 had only four category response options due to pre-estimation category collapsing. All items were estimated using the NRM.

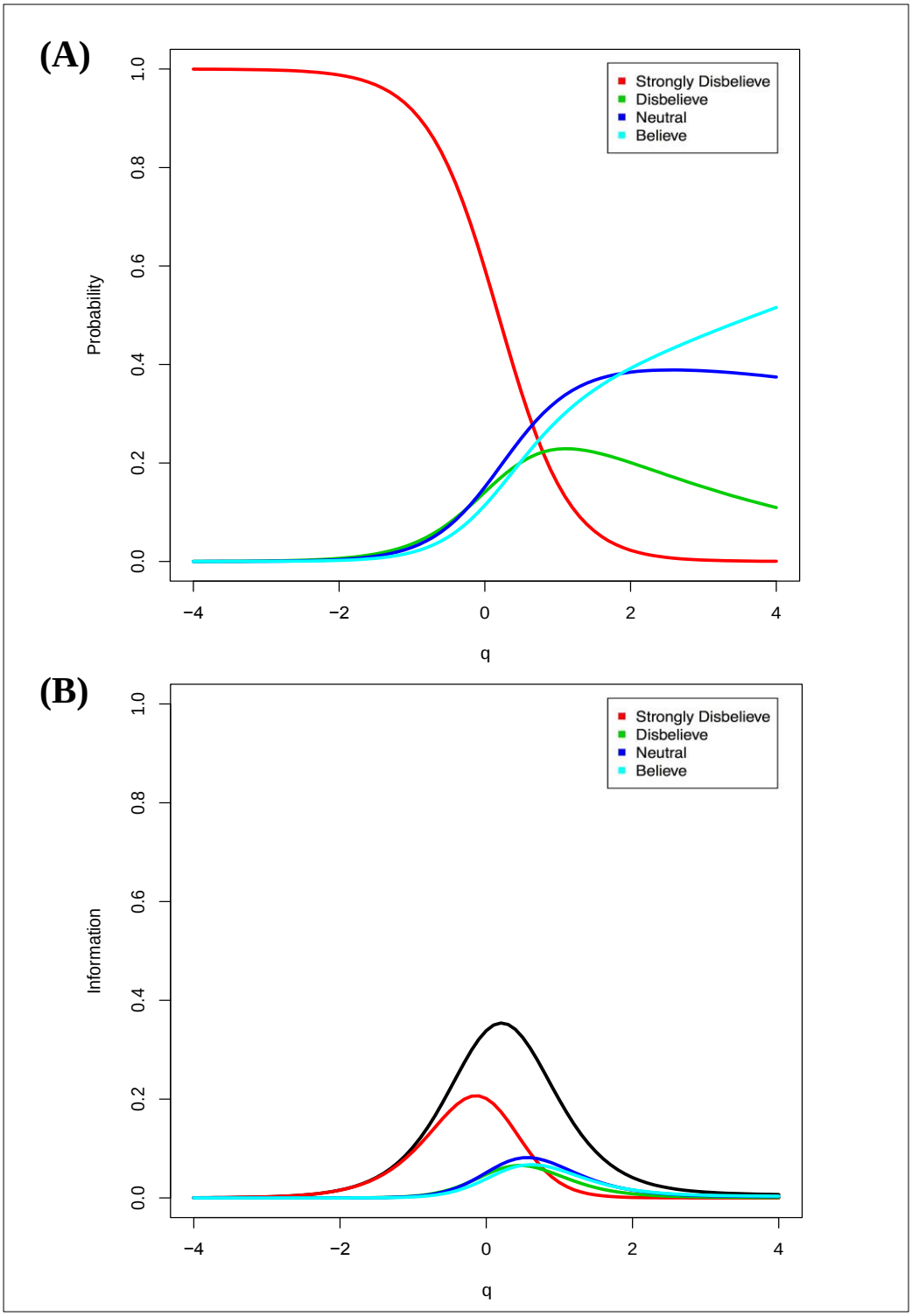


Figure 3. (A) Category response curves: FOC Scale Item 19. (B) Item and category information functions: FOC Scale Item 19.

Note. The categories for "Strongly Believe" and "Believe" were collapsed prior to estimation.

Significant Wald test statistics in combination with low CBD parameter values and overlapping category information was grounds for collapsing adjacent category response options for items 2, 3, 6, 7, 8, 10, 12, 13, 14, 15, 20, 21, 24, and 26. For these items, categories with the lowest CBD parameters were collapsed. As an example of the general reasoning behind this revision, please refer to Figure 4, which displays the CRCs, IIF, and CIFs for item 21. As can be seen in Figure 4, the 4^{th} and 5^{th} categories were functioning similarly, where the 5^{th} category provided no additional information that was not being provided by the 4^{th} category. The combination of the significant Wald test, low CBD and overlapping information functions provided amble empirical evidence to support a reduction of the response categories.

Generally, significant Wald test and low CBD parameters were grounds for category collapse. However this was not always the case. Items 11, 16, 17, 23, and 25 were not revised, despite the significant Wald. For instance, Figure 5 displays the CRCs, IIF, and CIFs for item 23. Although the CBDs varied significantly, category information indicated that each category provided unique information and served the overall scale in that it helped provide information over a wider range of the latent trait. Instead, the significant Wald test was an indication that a less constrained model could result in distorted person parameter estimates (Preston et al., 2011). Thus, these items require modeling using the NRM. Items 4 and 22 did not have significantly varying CBD parameters. This suggested that accurate person parameter estimate could be obtained with a more constrained model. As such, these items were modeled using the GPCM during the first revision estimation process.

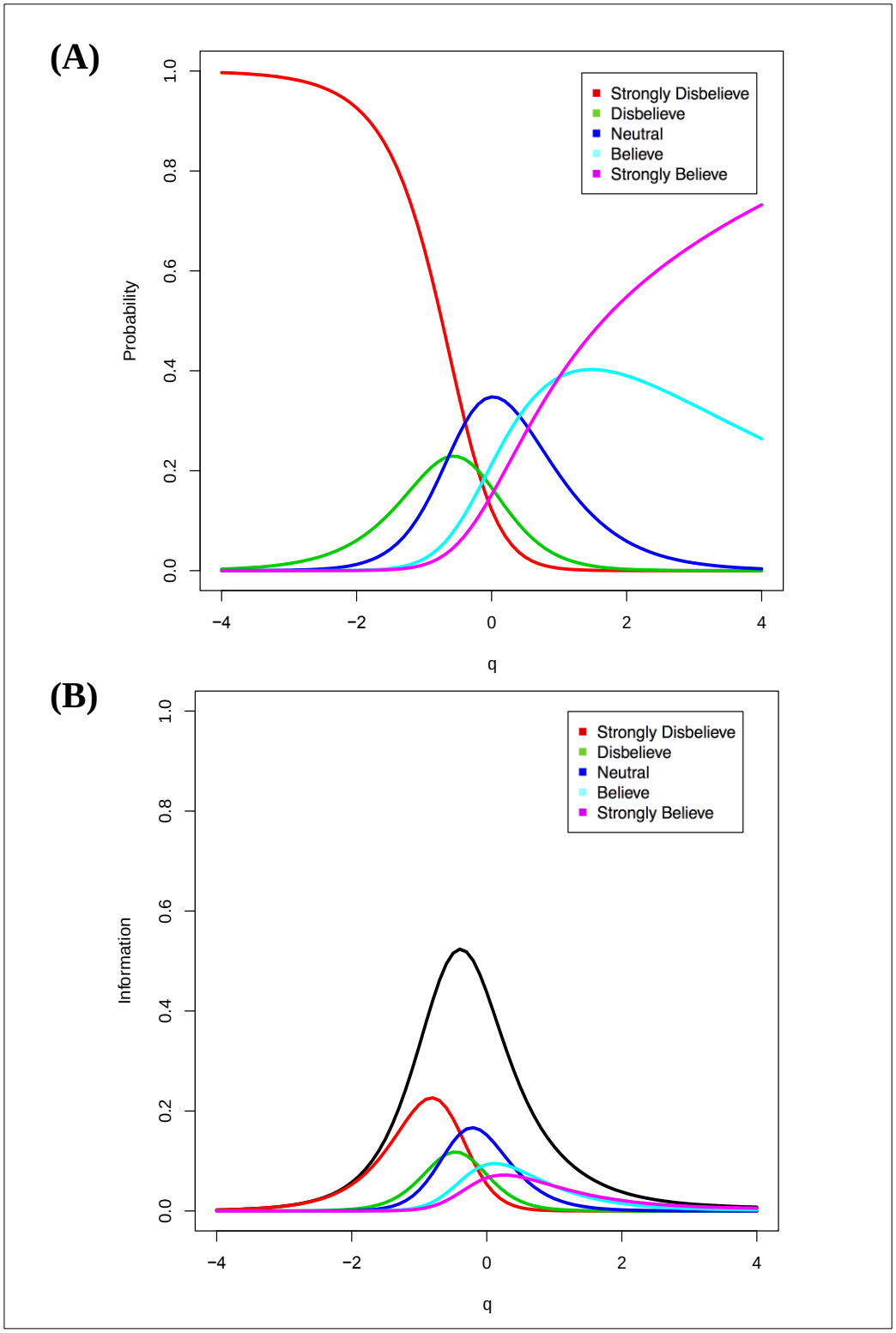


Figure 4. (A) Category response curves: FOC Scale Item 21. (B) Item and category information functions: FOC Scale Item 21.

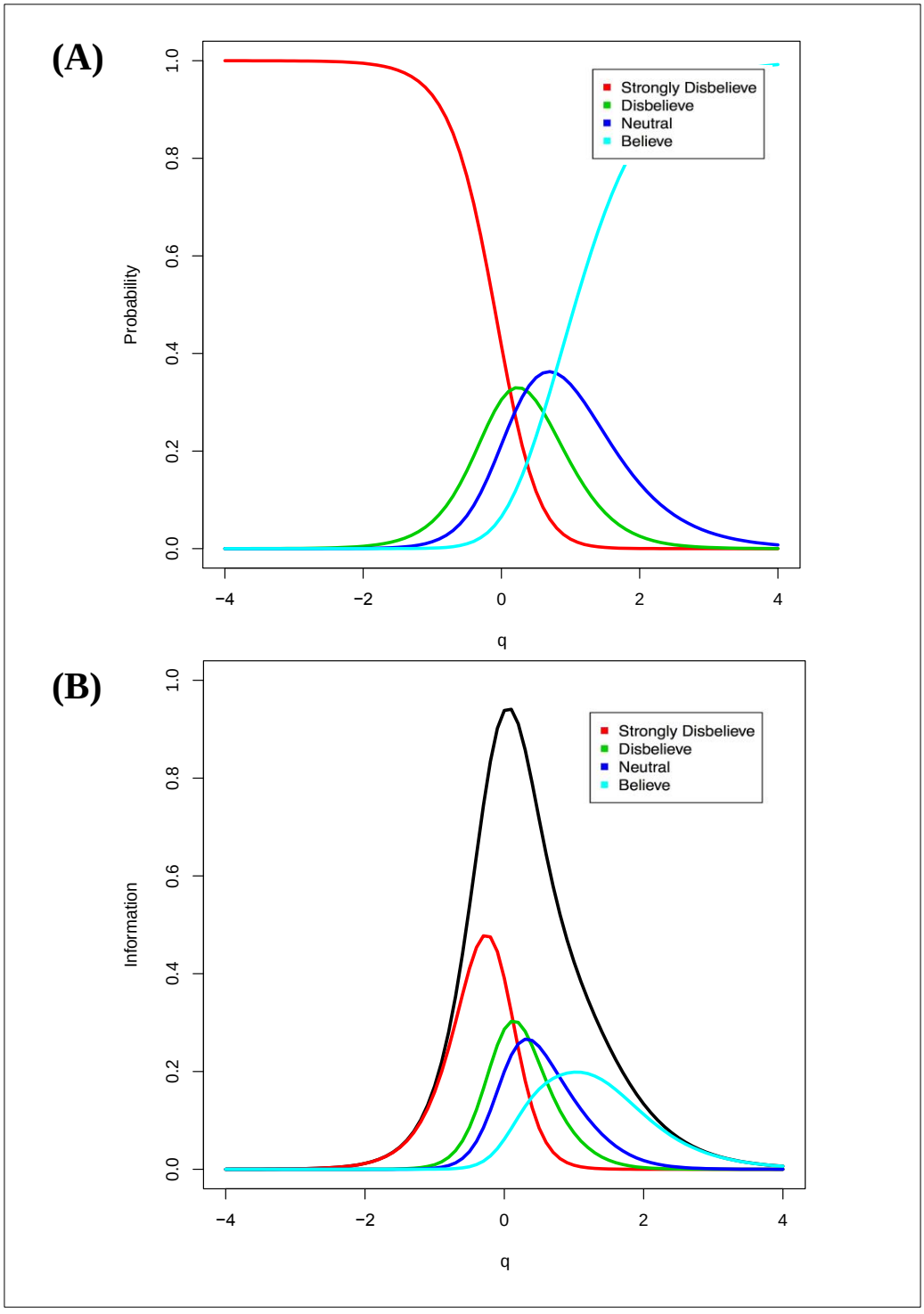


Figure 5. (A) Category response curves: FOC Scale Item 23. (B) Item and category information functions: FOC Scale Item 23.

Note. The categories for "Strongly Believe" and "Believe" were collapsed prior to estimation

Estimation After First Revision

Item parameters were then estimated with the aforementioned revisions. Item parameters for items 4 and 22 were estimated using the GPCM, while all other items were estimated using the NRM. Table 2 displays the CBD parameters, Wald test statistic, degrees of freedom, and probability value for each item. Non-significant Wald test statistics for items 15 and 26 indicated that the item and person parameter estimates could safely be obtained using the GPCM. Wald test statistics and CIFs indicated that items 3, 8, 11, 13, 16, 20, 23, and 25 should be estimated using the NRM. For example, Figure 6 shows the CRCs, IIF, and CIFs for item 3. As can be seen in Figure 6, all category response options provide unique information in regards to the latent trait continuum. Moreover, the item provided information over a wide range of the latent trait. This is especially important in light of the fact that many of the items, and categories, seemed to provide the most information around the center of the latent trait continuum, or a $\theta = 0$.

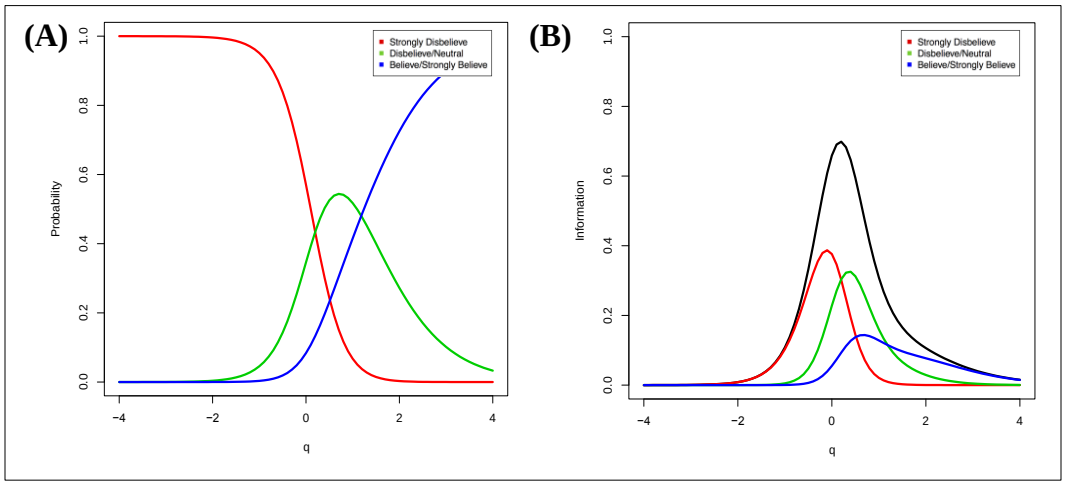


Figure 6. (A) Category response curves: FOC Scale Item 3. (B) Item and category information functions: FOC Scale Item 3.

Note. Categories representing "Strongly Believe" and Believe were collapsed after pre-estimation screening while "Disbelieve" and "Neutral" were collapse in the first revision.

Table 2

CBDs and Wald Test Statistics for the FOC Scale After First Revision

Item	CBD 1	CBD 2	CBD 3	CBD 4	Wald	*df*	*p-value*
1	-	-	-	-	-	-	-
2	2.48	1.0	1.2	-	136.88	2	0.000
3	2.54	1.2	-	-	25.244	1	0.000
4	0.91	0.91	0.91	0.91	-	-	-
5	1.96	0.74	0.65	-	53.672	2	0.000
6	2.04	0.91	0.52	-	55.884	2	0.000
7	1.71	0.7	-	-	17.931	1	0.000
8	1.94	1.05	1.03	-	34.907	2	0.000
9	-	-	-	-	-	-	-
10	2.62	0.82	1.37	-	191.827	2	0.000
11	2.3	0.83	1.09	-	139.236	2	0.000
12	2.22	1.17	-	-	11.972	1	0.001
13	2.87	1.29	-	-	30.312	1	0.000
14	2.07	0.98	1.12	-	60.968	2	0.000
15	1.5	1.39	-	-	0.121	1	0.728
16	2.04	1.24	1	-	23.009	2	0.000
17	2.22	0.87	0.79	-	87.023	2	0.000
18	-	-	-	-	-	-	-
19	-	-	-	-	-	-	-
20	2	0.92	-	-	19.317	1	0.000
21	1.59	1.14	1.33	-	6.468	2	0.039
22	1.19	1.2	1.19	-	-	-	-
23	2.47	1.01	1.52	-	105.892	2	0.000
24	1.58	0.88	1.16	-	19.239	2	0.000
25	2.86	1.67	1.34	-	70.166	2	0.000
26	1.22	1.22	-	-	0	1	1.000

Note. Items 1, 9, 18, and 19 were removed from the item set. Items 4 and 22 were estimated using GPCM.

Significant Wald test statistics and non-unique information for items 5, 6, 7, 10, 12, 14, 17, 21, and 24 were an indication that categories with the lowest CBD should be collapsed. This further reduction led to items 7 and 12 being collapsed into a dichotomous response option. As can be seen in Figure 7, categories three and four provided no unique information for item 12. Since these revised items resulted in only

two response categories, item parameters for these items must be estimated using the 2PLM for the second revision.

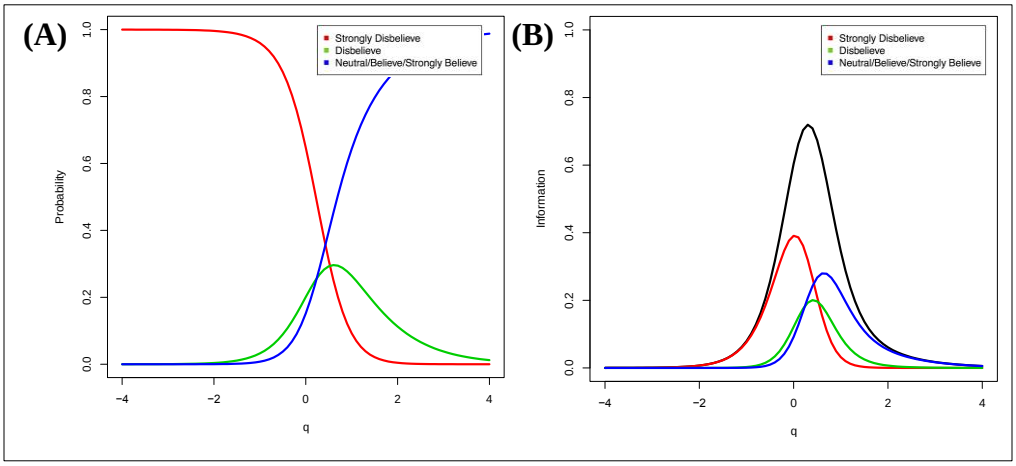


Figure 7. (A) Category response curves: FOC Scale Item 12. (B) Item and category information functions: FOC Scale Item 12.

Note. Categories representing "Strongly Believe" and "Believe" were collapsed after pre-estimation screening while the collapsed "Strongly Believe/Believe" category was further collapsed with "Neutral" in the first revision.

Estimation After Second Revision

Item parameters were then estimated with the aforementioned revisions. Item parameters for items 4, 15, 22, and 26 were estimated using the GPCM. Item parameters for items 7 and 12 were estimated using the 2PLM. All other item parameters were estimated using the NRM. Table 3 displays the CBDs and Wald test statistics for the FOC scale items after the second revision. Wald test, CBD parameters and CIFs for items 5, 6, and 17 indicated that categories 2 and 3 should be collapsed. This resulted in items 5, 6, and 17 being revised as a dichotomous response option between "Strongly Disbelieve" and all other response options collapsed together. Wald test, CBD parameters and CIFs

for items 2, 3, 8, 10, 11, 13, 14, 16, 20, 21, 23, 24, 25 indicate that the NRM would be the optimal method of estimating person parameters.

Table 3

CBDs and Wald Test Statistics for FOC Scale After Second Revision

Item	CBD 1	CBD 2	CBD 3	CBD 4	Wald	*df*	*p-value*
1	-	-	-	-	-	-	-
2	3.01	1.5	-	-	24.815	1	0.000
3	2.53	1.17	-	-	29.008	1	0.000
4	0.91	0.9	0.91	0.9	-	-	-
5	1.94	0.99	-	-	11.203	1	0.001
6	2.02	1.16	-	-	7.514	1	0.006
7	-	-	-	-	-	-	-
8	2	1.01	1	-	47.126	2	0.000
9	-	-	-	-	-	-	-
10	3	1.68	-	-	20.63	1	0.000
11	2.33	0.77	1.08	-	166.709	2	0.000
12	-	-	-	-	-	-	-
13	2.88	1.24	-	-	37.705	1	0.000
14	2.65	1.34	-	-	17.569	1	0.000
15	1.41	1.41	-	-	-	-	-
16	2.08	1.16	0.95	-	32.893	2	0.000
17	2.21	1.1	-	-	15.634	1	0.000
18	-	-	-	-	-	-	-
19	-	-	-	-	-	-	-
20	2	0.87	-	-	22.768	1	0.000
21	2.23	1.61	-	-	5.023	1	0.025
22	1.18	1.19	1.18	-	-	-	-
23	2.47	0.97	1.47	-	116.144	2	0.000
24	2.06	1.42	-	-	5.332	1	0.021
25	2.91	1.57	1.32	-	97.506	2	0.000
26	1.2	1.2	-	-	-	-	-

Note. Items 1, 9, 18, and 19 were removed from the item set. Items 4, 15, 22 and 26 were estimated using the GPC model.

Figure 8 shows the ICCs and IIFs for the 2PLM items 7 and 12. Item difficulty, or *b* parameter for item 7 and 12 was -0.67 and 0.21, respectively. This indicated where on the latent trait continuum the probability for endorsing each item is .50. Item

discrimination, or *a* parameter for item 7 and 12 was 2.07 and 2.76, respectively. Item discrimination can be seen in Figure 8 (A) as the slopes of each ICC. Both items had relatively steep slopes. To illustrate, item 12 serves to make fine distinctions for individuals around a θ of .21. That is, individuals that possess $\theta > .21$ are almost certain to endorse the item.

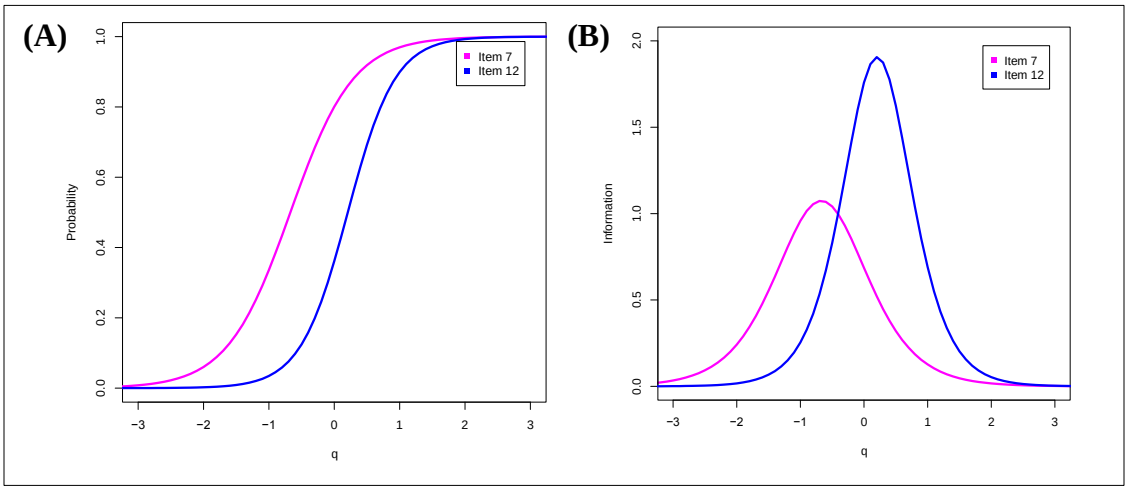


Figure 8. (A) Item characteristic curves: FOC Scale Item 7 and 12. (B) Item information functions: FOC Scale Item 7 and 12.

Note. (A) For both items, ICCs represent the dichotomous distinction between "Strongly Disbelieve" and the collapse of all other category options. (B) Information is scaled from 0.0 to 2.0.

Estimation After Third Revision

Item parameters were then estimated with the aforementioned revisions. Items 5, 6, 7, 12, and 17 were estimated using the 2PLM. Item parameters for items 4, 15, 22, and 26 were estimated using the GPCM. Item parameters for all other items were estimated using the NRM. Table 4 displays the CBDs and Wald test statistics for the FOC scale items after the third revision. Since NRM and GPCM items were satisfactory after the last revision, only the final 2PLM items are discussed.

Table 4

CBDs and Wald Test Statistics for FOC Scale After Third Revision

Item	CBD 1	CBD 2	CBD 3	CBD 4	Wald	*df*	*p-value*
1	-	-	-	-	-	-	-
2	3.01	1.48	-	-	25.668	1	0.000
3	2.53	1.14	-	-	29.733	1	0.000
4	0.91	0.9	0.91	0.9	-	-	-
5	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-
8	2.02	0.97	1.01	-	54.676	2	0.000
9	-	-	-	-	-	-	-
10	3.02	1.68	-	-	19.439	1	0.000
11	2.37	0.74	1.09	-	188.497	2	0.000
12	-	-	-	-	-	-	-
13	2.89	1.21	-	-	36.593	1	0.000
14	2.67	1.34	-	-	23.295	1	0.000
15	1.41	1.41	-	-	-	-	-
16	2.1	1.1	0.94	-	39.336	2	0.000
17	-	-	-	-	-	-	-
18	-	-	-	-	-	-	-
19	-	-	-	-	-	-	-
20	1.99	0.83	-	-	23.36	1	0.000
21	2.23	1.61	-	-	5.14	1	0.023
22	1.18	1.17	1.18	-	-	-	-
23	2.48	0.91	1.45	-	130.146	2	0.000
24	2.06	1.41	-	-	6.396	1	0.011
25	2.96	1.49	1.31	-	123.549	2	0.000
26	1.19	1.19	-	-	-	-	-

Note. Items 1, 9, 18, and 19 were removed from the item set. Items 4, 15, 22 and 26 were estimated using the GPC model. Items 5, 6, 7, 12, and 17 were estimated using the 2PLM.

Table 5 displays the discrimination and difficulty parameters for all 2PLM items. Discrimination was relatively high with items 6, 12, and 17 being capable of making the finest discriminations. Figure 9 shows the ICCs and IIFs for all the 2PLM items. As seen in Figure 9 (A), items 5, 12, and 17 functioned similarly, in that individuals with a θ of around 0.20 shared an equal probability of endorsing each item

Table 5

Discrimination and Difficulty Parameters for 2PLM Items

	a	*b*
Item 5	2.46	0.22
Item 6	2.74	-0.34
Item 7	2.05	-0.67
Item 12	2.75	0.21
Item 17	2.74	0.14

As seen in Figure 9 (B), all of the items provided a good amount of information over the range of the latent trait.

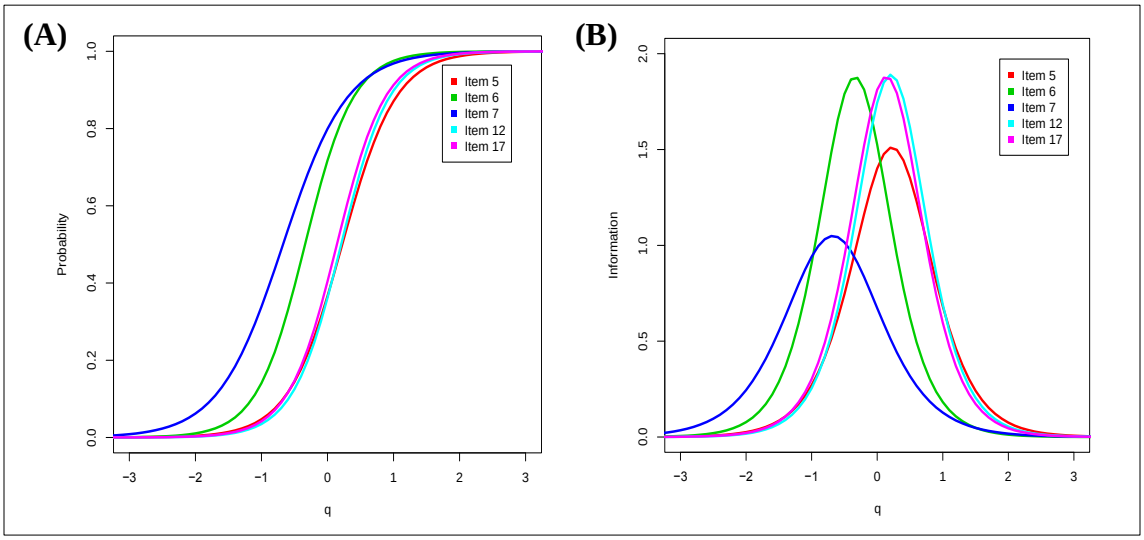


Figure 9. (A) Item characteristic curves: FOC Scale Item 5, 6, 7, 12 and 17. (B) Item information functions: FOC Scale Items 5, 6, 7, 12 and 17.

Note. (A) For all items, ICCs represent the dichotomous distinction between "Strongly Disbelieve" and the collapse of all other category options. (B) Information is scaled from 0.0 to 2.0.

After the third revision, all items were most optimally modeled and functioned satisfactorily. Therefore, the estimation for the third revision represented the final scale version. Table 6 displays the final scale items, revised category lengths and optimal model type. Figure 10 displays the TIF, or the relative combined information for all FOC scale items. As seen in Figure 9, information for the entire scale peak with .85 at approximately θ of 0. Moreover, the scale provided a modest amount of information for persons with θ between -1.5 and 1.5. Limited information fit statistics for the final model indicated an acceptable fit with RMSEA = 0. 05, M_2 (1007) = 3056.79, p = 0.0001.

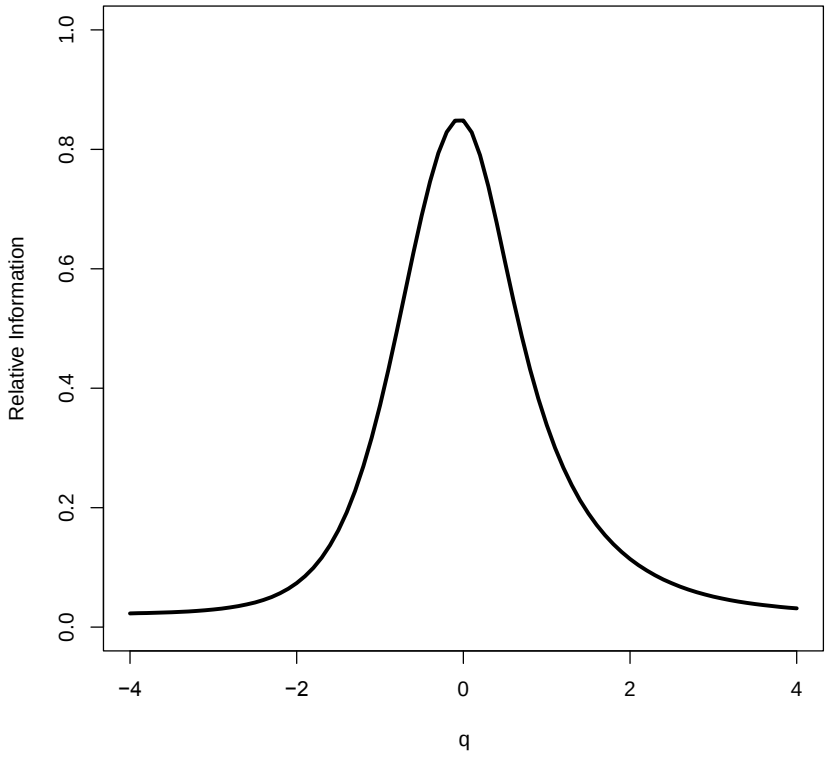


Figure 10. Relative Test Information Function for the final FOC Scale.

Table 6

Final Scale Items with Revised Category Length and Optimal Model Type

Item	Statement	Category Length	Model
2	Ghosts or spirits can interact with the physical world	3	NRM
3	Certain people can cast spells (e.g. do magic, curse someone, etc.)	3	NRM
4	It is possible for people to have lived in a previous life	5	GPCM
5	It is possible to move material objects with only one's thoughts.	2	2PLM
6	When I make a wish I believe something or someone can hear it	2	2PLM
7	The universe has a purpose for me	2	2PLM
8	It is possible for some people to experience future events before they happen	4	NRM
10	It is possible for people to communicate with the dead	3	NRM
11	Bad things happen to certain people because they attract negative energy	4	NRM
12	Looking up a person's zodiac sign is a good way to gather information about who they are	2	2PLM
13	Certain items (e.g. crosses, magnetic bracelets, amulets, etc.) can help protect people	3	NRM
14	Good things happen to some people because of positive energy	3	NRM
15	Energy can be good or bad	3	GPCM
16	The position of the planets can influence the events in my life	4	NRM
17	Certain natural events such as eclipses can be an omen, or signs of something bad to come	2	2PLM
20	Weather can be affected by the emotions of individuals or groups	3	NRM
21	People can feel the presence of friends or family after they have died	3	NRM
22	Deep thought, meditation, or concentration can cure other peoples illnesses	4	GPCM
23	A person can influence the physical world solely with their thoughts	4	NRM
24	Positive thoughts can cleanse the body of toxins	3	NRM
25	The essence, or spirit, of a person can remain in objects he or she came into contact with when they were alive, such as, articles of clothing, books, etc.	4	NRM
26	Energy lives in nature	3	GPCM

Note. Items 1, 9, 18, and 19 were removed from the item set. Model labels 2PLM is the two-parameter logistic model, NRM is the nominal response model, and GPCM is the general partial credit model.

Discussion

The FOC scale construction was achieved through several steps. First, an extensive literature review, where a theoretical framework was expounded to define PSMS beliefs. With this, an initial item pool was created and administered to a large sample. Then, the scale was empirically evaluated and revised under an IRT framework using the NRM. With the NRM and Wald test, item category response options were revised. The finalized scale format can be seen in Appendix D. Table 6 displays the final items, category lengths, and optimal model type.

Initial factor analysis indicated the presence of a religious factor that had the potential to confound the scale. The items that comprised the religious factor were constructed as ontological confusions, but as a whole, ended up being characteristic of the root beliefs for the major monotheistic religions. For example, the item that states, "Human beings possess something non-physical (e.g. a soul) that makes them different from animals" in theory would be linked to other ontological confusions in which physical permanence is attributed to the mental concept (e.g., consciousness). Additionally, the item that states, "Praying for people can heal them" should be linked to confusions of the mental and the biological. Although apparent in retrospect, it was assumed that items with religious links would blend in with the overall ontological confusions.

The FOC scale was intended to be a measure for PSMS beliefs and not a measure of religiosity. It is assumed that religious individuals would hold some PSMS beliefs to be true, however, PSMS beliefs are not what defines religious individuals. Since the religious factor had the potential to influence the modeling of the PSMS belief trait

estimation if the scale was assumed to be unidimensional, and because the measure was not constructed to be multidimensional, the religious factor items were removed from the scale.

Item 19 was removed during the revision process due to highly varying CBDs and low category information. Moreover, after reviewing the item it is conceivable that participant had trouble understanding the item wording. Item 19 states, "Natural disasters, such as earthquakes or tsunamis, can occur as a form of retribution for certain human's actions or beliefs." There is the possibility that some participants had trouble understanding the meaning of the word *retribution* being that it is high school level vocabulary.

Items 9 and 18 were removed because of negative CBD parameters. However item 6 had a negative CBD parameter and was not removed. The difference in rationale was that items 9 and 18 also displayed low and redundant category information. Although negative CBD parameters tend to indicate unordered categories, if the magnitude is small it may just reflect the utter lack of ability for the respondent to distinguish between response categories.

An important finding of Study 1 is that assuming that category response options function similarly for all items can be incorrect. Recall, only one item retained the original 5-category response format. The majority of items category response formats were reduced either during the initial item screening or over the course of the three revisions. Future research should be cautious to arbitrarily choose category response options formats. Moreover, it is advised that researchers empirically assess the functioning of each category response.

CHAPTER 3

STUDY 2: VALIDATION

The revised FOC scale was validated following the guidelines put forward by Whitley and Kite (2013). In order to offer evidence for validity for the FOC scale, the content, structural, external, discriminant, and substantive validity were reviewed and evaluated. The content, structural, external and substantive validity, together, are what Campbell and Fiske (1959) called convergent validity. These separate lines of evidence come together to provide support for the degree of validity of the measure (Whitley & Kite, 2013; p. 158). Discriminant validity is another facet of validity where the extent to which a measure is *not* measuring what it is *not* suppose to measure (Whitley & Kite, 2013; p. 158). Taking all the evidence together provides a means for determining the degree of validity.

Content and structural validity were evaluated during the construction and revision process found in Study 1. Content validity addresses whether the content, or items, are relevant and adequately represent the entire range of the latent trait (Whitley & Kite, 2013; p. 155). With the IRT analyses carried out in Study 1, items that were not functioning as a measure for the FOC scale were removed. Moreover, the analyses indicated that items provided information over a range of the latent trait and possessed acceptable fit as a model for estimating individual's tendency to make ontological confusions. Therefore, the FOC scale was found to possess content validity. Structural

validity addresses the dimensionality of the measure (Whitley & Kite, 2013; p. 156). The FOC scale was found to be unidimensional and therefore possessed structural validity.

External validity is the notion that measures should be related to external criteria that is theoretically similar (John & Soto, 2007). The external validity for the FOC scale was evaluated by analyzing the relationship between the FOC scale and the external validity criterion measures, i.e., the R-PBS (Tobacyk, 2004) and the Rational Experiential Inventory (REI; Epstein et al., 1996). Since the proposed measure is intended to represent a more complete conception of PSMS beliefs, FOC scale θ scores are expected to have a significant positive relationship with scores on R-PBS. Additionally, a greater tendency towards an intuitive thinking style has been found to be related to PSMS beliefs and mistakes in ontology (Lindeman & Aarnio, 2006; Svedholm & Lindeman, 2013). Therefore it is expected that REI intuitive scores would be positively related with FOC scale θ scores.

Discriminant validity was evaluated by testing the relationship between the FOC scale and theoretically unrelated constructs, i.e. social desirability, intuitive thinking, and religiosity. Social desirability is an individual's bias to respond to questions in a socially acceptable way instead of honestly and accurately (Holtgraves, 2004). As a construct, socially desirability should have no relation to the FOC θ scores. As previously mentioned, a confusion of ontology is not evidence of a detriment in analytical thinking, instead it is related to a reliance on an intuitive style of thinking (Aarnio & Lindeman, 2005; Auton, et al., 2003). Therefore, scores on the REI rational scale should *not* be correlated with FOC scale θ scores. Religiosity is a more contentious construct in terms

of its relationship to PSMS beliefs. Although some conflicting evidence exists, with a valid measure of PSMS beliefs, religiosity is predicted to be unrelated.

Substantive validity is achieved by testing propositions derived from the theoretical underpinnings of the measure (John & Soto, 2007). One such relationship is that of the PSMS believer's tendency to falsely perceive patterns in randomness (Dagnall et al., 2007). Conceptually, ontological confusions are rooted in an intuitive style of thinking (Riekki et al., 2013). Intuitive thinkers have been found to falsely perceive patterns in randomness (Gagnall, Parker, & Munley, 2007; Riekki et al., 2013). Therefore, individual with a greater tendency to hold ontological confusions should falsely perceive patterns. Examining the relationship between FOC θ scores and false alarms during a face perception task was used to test this theoretical relationship. That is, greater FOC θ scores should be positively related with a greater tendency to falsely perceive faces in non-face images.

The goal of Study 2 was to validate the FOC scale by testing the aforementioned hypothesized relationships. Please reference the validation hypotheses in the purpose and hypotheses section of the Introduction.

Methods

Participants

Participants (N = 170) were obtained from the CSUF's Sona Systems research participant pool. The Sona Systems Research Pool was comprised primarily of undergraduate Psychology students that were allowed to participate in research of their choosing for course credit. Participants received a 1-hour credit for their participation.

Data were screened for accuracy prior to analysis. Three attention check items were placed randomly throughout the questionnaire. Participants were removed for failing to respond correctly to any one of three attention check items. A total of 16 (9.4%) participants were removed due to failing to respond correctly to any one of the attention check items. The resulting sample size was $N = 154$.

A power analysis was conducted regarding the use of correlational analysis during the validation process. The statistical computing language R (2014) was utilized to conduct a power analysis using a power analysis package, pwr (Champely, 2012). Specifying the desire for a large effect size ($r = .5$), a significance level at $\alpha = 0.05$, and power = .90, a minimum of $n = 38$ was needed for correlational analysis. Therefore, the sample size ($n = 154$) was more than adequate for the desired power and effect size for the correlational analysis.

The final sample comprised $N = 154$ participants. The sample was predominately female (62.9%, $n = 97$). Age ranged from 18 to 64 with an average respondent age of 24.6 years old ($SD = 8.4$). The sample was ethnically diverse with White/Caucasian with 44.2% ($n = 68$) of participants identifying as White, 22.1% ($n = 34$) as Mexican American, 22.7% ($n = 35$) as Asian/Asian American, 2.6% ($n = 4$) as African American/Black, and 5.2% ($n = 8$) identifying as mixed ethnicity. The majority of participants identified with American culture (64.3 %, $n = 99$), while 16.2% ($n = 25$) identified primarily with Latino(a) culture, and 14.3% ($n = 22$) with Asian cultures.

The sample was fairly religious with a little over half identifying as being Catholic (37%, $n = 57$), Christian (20.8%, $n = 32$), or Muslim (2.6%, $n = 4$). The remaining participants were not religious with 25.9% ($n = 40$) self-reporting as Non

Religious, 8.4% as Atheist (*n* = 13), and 3.9% as Agnostic (*n* = 6). All of the participants had at least a High School education (24.7%, *n* = 38), with 40.9% having Some College (*n* =63), 28.6% having a College Degree (*n* = 44), and 5.8% having a Graduate Degree (*n* = 9).

Materials and Procedure

The procedure for Study 2 was identical to that of Study 1. The online questionnaire contained the revised FOC scale, the R-PBS (Tobacyk, 2004), the Rational Experiential Inventory (Epstein et al., 1996), the Religious Commitment Inventory (Worthington et al., 2003) and the Crown-Marlowe Social Desirability—Short Form C (Reynolds, 1982). Furthermore, participants rated several images in an illusory face perception task. Participants also answered demographic questions, such as, age, gender, ethnicity, cultural identity, highest level of education, relationship status, sexual preference, religious affiliation, and US postal ZIP code.

FOC Scale. The FOC scale was constructed and empirically revised in Study 1. The fully revised 22-item version was administered to participants in Study 2. Appendix D contains the text version of the scale. The online version was identical in content but formatted to fit the online questionnaire. Participant's θ scores were derived from the parameter estimates produced during the final scale estimation in Study 1. Item and person parameters for items 1-5 were estimated with the 2PLM. Items and person parameters for items 11, 15, 19, 22 were estimated using the GPCM. Item and person parameters for the remaining items were estimated with the NRM.

R-PBS. The most commonly used measure for PSMS beliefs; the R-PBS (Tobacyk, 2004) was utilized to assess the external validity of the FOC. Refer to

Appendix E for the full scale. The R-PBS consists of 26-items assessing 7 factors of PSMS beliefs, which are Traditional Religious Belief, Psi, Witchcraft, Superstition, Spiritualism, Extraordinary Life Forms, and Precognition. Items are scored with a 7-point scale; 1 = *Strongly Disagree*, 2 = *Moderately Disagree*, 3 = *Slightly Disagree*, 4 = *Uncertain*, 5 = *Slightly Agree*, 6 = *Moderately Agree*, and 7 = *Strongly Agree*. The sum scores for each subscale represent each factor score. The sum score of the entire scale reflects general paranormal belief. The author reported a reliability coefficient of $\alpha = .95$ for the entire scale.

REI. The degree to which someone favors an intuitive style or an analytical style of thinking was measured with the 10-item REI developed by Epstein, Pacini, Denes-Raj, and Heier (1996). Refer to Appendix F for the full scale. Participants were instructed to rate the extent to which they believe the statements are characteristic of themselves using a 5-point scaled anchored with 1 = *Completely False* to 5 = *Completely True*. The 10-item REI is comprised of two scales, five items for Need for Cognition (NFC) and five items for Faith in Intuition (FI). The authors reported good reliability for both the NFC ($\alpha = .73$) and the FI ($\alpha = .72$) scales. The 10-item REI is a shortened version of the original 31-item REI. Epstein, Pacini, Denes-Raj, and Heier (1996) found nearly identical results in terms of the predicted relationships to other constructs between the 10-item version and the original REI.

Religious Commitment Inventory (RCI-10). Religiosity was measured in terms of religious commitment using the 10-item RCI-10 developed by Worthington et al. (2003). Refer to Appendix G for the full scale. Participants were asked to rate the extent to which they believe the statements are characteristic of themselves using a 5-point scale

anchored with 1 = *Not at all* and 5 = *Totally*. Worthington and colleagues (2003) found evidence that the measure was largely unidimensional, reliable and valid. The authors reported a reliability coefficient of $\alpha = .92$ for the RCI-10.

Crown-Marlowe Social Desirability—Short Form C. Reynolds (1982) explicated a valid short form version of the Crowne-Marlowe Social Desirability scale. The short form version consists of 13 true/false statements from the original Crowne-Marlowe Social Desirability scale (1960). Refer to Appendix H for the full scale. Reynolds (1982) reported evidence for unidimensionality, reliability, and validity for the short form version. Kuder Richardson formula twenty indicated reliability ($r_{KR20} = .76$) and the short form versions were significantly positively correlated to the original scale ($r = .93$, $p < .001$).

Illusory Face Perception Task (IFPT). Illusory face perception, or apophenia, was measured through a series of degraded images with and without facial features. Refer to Appendix I for examples of degraded face and non-face images. Face and non-face images were degraded by 80%. Degradation was achieved by adding the specific percentages of random noise to the images (Ramos, 2013). To attempt to prevent ceiling or floor effects with the face perception stimuli, 43 face and non face stimuli were piloted with a sample of $N = 52$ participants recruited from Amazon's MTurk. Of the original 40 stimuli, 15 stimuli had a miss rate (or incorrect detection) above 5%. The top 8 stimuli had a miss rate between 15% and 30%, while the bottom 7 stimuli had a miss rate between 5% and 15%. Actual face images comprised 7 of the 15, while non-face images were 8 of the 15. The final face and non-face images shared similar proportions for miss rates. The 15 face and non-face stimuli were randomly ordered in the online

questionnaire. The participants viewed the 15 images and were asked to rate whether of not they perceive an image of a face.

Signal Detection Theory (SDT) was used to model the participant's responses to the detection tasks. SDT postulates that all decision-making involves the presence of uncertainty that inevitably biases decisions (Macmillan & Creelman, 2005). In the context of the current research, participants' responses to the face detection task could be one of four possibilities. The participant could either correctly detect a face images (*hit*), fail to detect a face image (*miss*), incorrectly detect a face image (*false alarm*), or correctly reject a non-face image (*correct rejection*). Percentages for each response possibility were calculated for all the images. Since response bias is an issue during decision-making under uncertainty, signal detection analysis was employed to adjust for response bias (Stanislaw & Todorov, 1999). Taking the difference between the normed probabilities for false alarms and hits provides a measure of detection sensitivity (*d'*; Addi, 2007; Stanislaw & Todorov, 1999). Essentially, *d'* is an indication of the strength of the signal relative to the noise (Addi, 2007) and served as an unbiased indication of the participants ability to discriminant between the face and non-face images.

Results

FlexMIRT version 3 (Cai, 2013) was used to produce individual FOC θ scores. Item parameters were not re-estimated; instead scale scoring was achieved by using the item parameter estimates produced during the final FOC scale calibration. All other items for each measure were summed to form their specific scale score based on the procedure laid out by their authors. For scores on the IFPT the percentage of false alarms (IFPT-FA) and the *d'* (IFPT-D) were calculated. Signal detection analysis was utilized to calculate *d'*

for each individual by taking the standardized difference between false alarms and hits. A *d'* = 0 would indicate an utter lack of ability to distinguish between face and non-face images, whereas larger values of *d'* would indicate a greater ability to distinguish face from non-face images (Addi, 2007; Stanislaw & Todorov, 1999).

Prior to analysis, data were screened using R. Cases were examined for univariate and multivariate outliers. Two cases were found to be both mulitivariate and univariate outliers and were removed from the data set. Little's (1988) missing completely at random (MCAR) test was performed on the data and indicated that the data were MCAR, $\chi^2(39) = 28.169$, $p = 0.901$. Moreover, the data contained less than 5% missing data, thus pairwise deletion was used during the analyses. Pairwise plots were examined to check for nonlinearity and heteroscadasticity. No major violations were found to be present.

Skewness, kurtosis, and normality plots were examined to check for issues with non-normality distributed variables. RCI-10, IFPT-FA and IFPT-D were found to have issues with normality. Specifically, each variable was heavily skewed to the extent to which issues with floor and ceiling effects were possibly present. The Box-Cox method was used to find the optimal transformation for each variable (Box & Cox, 1964). After transformations, a notable ceiling and floor effects were still present for both the IFPT-FA and the IFPT-D. These effects most likely represented issues with the IFPT itself. Interpretations of the correlations with these variables are to be taken with caution. A floor effect was present for RCI-10; however, this was most likely a natural response pattern considering the amount of non-religious individuals in the sample.

Pearson's product-moment correlation coefficient between the obtained θ scores and the validity criterion measures were calculated and are displayed in Table 7.

Supporting the first validation hypothesis, FOC θ scores were significantly positively correlated to scores on the R-PBS ($r = 0.77$, $p < 0.001$). Furthermore, the second validation hypothesis regarding external validity was supported with FOC θ scores being significantly positively correlated with scores on the REI-FI scale ($r = 0.32$, $p < 0.001$). Both correlations offered support for the external validity of the FOC scale.

In regards to discriminant validity, only one of the three hypotheses was initially supported. FOC θ scores showed no relationship with the measure for social desirability ($r = 0.12$, $p = 0.142$). However, there were relationships found between FOC θ scores and the measure for analytical thinking and religious commitment. There was a significant positive relationship between both REI-NFC and FOC θ scores ($r = 0.32$, $p < 0.001$) and RCI-10 and FOC θ scores ($r = 0.25$, $p < 0.01$).

After further investigation, the issue with the floor effect of the RCI-10 was deemed problematic. Specifically, the tendency for non-religious individuals to score a 10 (essentially zero) on the RCI-10 and lower on the FOC scale caused the correlation between the two variables to be attenuated. The relationship was then analyzed after removing individuals who self reported as Non-Religious, Atheist, or Agnostic ($n = 59$, 38.8%). Among the self-reported religious, religious commitment was not related to FOC θ scores ($r = 0.08$, $p = 0.457$). Thus, a total of two of the three discriminant validity hypotheses were supported.

Due to the ceiling and floor effects found in the IFPT variables, the present findings regarding the substantive validity were not used as an indication of evidence for validity. However, the results could be an indication that future research into the relationship between FOC scores and illusory face perception are warranted.

Table 7.

Correlations for the FOC θ Scores and the Validity Criteria Measures

		1	2	3	4	5	6	7
1	FOC							
2	R-PBS	0.77***						
3	REI-NFC	0.32***	0.30***					
4	REI-FI	0.32***	0.30***	0.35***				
5	RCI-10	0.28**	0.30**	0.08	0.12			
6	CMSD	-0.12	-0.09	-0.06	0.16	0.07		
7	IFPT-FA	0.27***	0.25**	0.11	0.19*	0.17	-0.01	
8	IFPT-D	-0.25**	-0.23**	-0.11	-0.16*	-0.15*	0.01	-0.96***

Note. (1) FOC is the Fullerton Ontological Confusion scale. (2) R-PBS is the Revised Paranormal Belief Scale (Tobacyk, 2004). (3) REI-FI is the Faith in Intuition subscale while the (4) REI-NFC is the Need for Cognition subscale for the Rational Experiential Inventory (Epstein et al., 1996). (5) RCI-10 is the 10-item Religious Commitment Inventory (Worthington et al., 2003). (6) CMSD-C is the Crown-Marlowe Social Desirability—Short Form C (Reynolds, 1982). (7) IFPT-FA is the Illusory Face Perception Task false alarms while the (8) IFPT-D is the sensitivity score.

*$p < .05$. ** $p < .01$. ***$p < .001$.

Participants with greater FOC θ scores tended to make more false detections of faces in non-face images ($r = 0.27$, $p < 0.001$). Furthermore, participants with greater FOC θ scores tended to be less capable of making distinctions between face and non-face stimuli as indicated by *d'* ($r = -0.25$, $p < 0.01$). Regardless, no definitive evidence for substantive validity could be identified.

Discussion

The FOC was found to possess external validity with both paranormal beliefs and faith in intuition being positively related to FOC θ scores. As for discriminant validity, need for cognition were related despite being theoretically unrelated. The substantive validity of the FOC was unable to be determined due to the ceiling and floor effects relating to the issues with the IFPT. Overall, the FOC scale was found to possess a moderate degree of validity with four of the five testable validation hypotheses being supported.

FOC θ scores were significantly positively related to scores on the R-PBS, indicating that individuals tended to score similarly on each scale. That is, individuals who tended to make ontological confusions also tended to hold more paranormal beliefs to be true. This relationship provided support for the external validity of the FOC scale. Moreover, the very large relationship between the two provides support for the notion that FOC θ scores are representative of PSMS beliefs.

As predicted, FOC θ scores were significantly positively related to the tendency to prefer an intuitive style of thinking. However, FOC θ scores were also positively related to a preference in analytical thinking. The degree of both of these relationships was moderate ($r = .32$). The relationship between the intuitive style of thinking and FOC

scores was anticipated, however, the relationship with analytical thinking was not. A speculation would be that making mistakes of ontology would be related to detriments in analytical styles of thinking. However, previous research indicated that no relationship between analytical thinking and CKCs (Aarnio & Lindeman, 2005; Auton et al., 2003) or PSMS beliefs (Hergovich & Arendasy, 2005) should exist.

The question is whether the current finding is evidence of a problem with the validity or evidence of something else. As for whether this relationship is damaging to the validity of the FOC scale, the fact that the R-PBS showed similar patterns of correlation with all other measures was encouraging.

Religious commitment was initially found to be positively related to FOC scores, in that the greater commitment seemed to be related to greater FOC scores. However, this may have been better conceived of as lower FOC scores being related to lower religious commitment. On the face of it, this made intuitive sense. People who are not religious at all probably do not hold many paranormal beliefs to be true. However, the concern with religiosity and the FOC scale was that it was constructed to be a measure for PSMS beliefs and not religiosity. Since individuals who were not religious at all would inevitably score low on the RCI-10, this flooring effect confounded the correlation between the FOC scale and the RCI-10. In the context that the RCI-10 was employed to provide evidence that the FOC scale was not a measure for religiosity, analyzing the relationship with only individuals who self-reported as being apart of a religion was warranted. By doing this, the floor effect was removed and a clearer picture of the relationship between religiosity and the FOC scale was made. Specifically, of the

religious, FOC scores and religiosity had no relationship. This provides evidence for the discriminant validity of the FOC scale.

The ceiling and floor effects displayed by the IFPT measures were caused by an overly easy task. Despite the extensive piloting of the face and non-face stimuli, the measures still produced ceiling and floor effects. Specifically, a third of the participants had no false alarms, while another fifth had very few false alarms. This could have been a result of the medium in which the images were viewed. Although no notable differences were found when looking at the different platforms (i.e., smartphone or desktop computer), it is possible that the ability of the participants to view the images at their leisure could have influenced the results. The face and non-face stimuli were piloted with an MTurk sample. Being that MTurk respondents get compensated the same amount regardless of the time they spent, the motivation to complete the task quickly could have been present. However, with the CSUF sample, it is possible that participant took time to scrutinize the image before giving an answer. Furthermore, there was no restriction on whether or not the participants could manipulate the image, e.g. adjust the screen zoom. All of these possibilities could have been the cause of the extremely low false alarm rate.

Despite the inability to adequately assess the substantive validity of the FOC scale, the evidence provided in Study 2 offers much support for the overall validity of the FOC scale.

CHAPTER 4

GENERAL DISCUSSION & CONCLUSION

General Discussion

The purpose of this thesis was to construct an informative and valid measure for PSMS beliefs. The findings of Study 1 and Study 2 provided promising evidence for the overall reliability and validity of the FOC scale. IRT was utilized to empirically revise the scale and yielded a measure with only the most psychometrically informative items. Moreover, IRT analysis using the NRM and the Wald test provided evidence for the most optimal format for the FOC scale. The construction of the FOC was the result of both strong theoretically defined item creation and data driven revisions. Both of these methods yielded a sound and informative measure for PSMS beliefs. Study 2 provided evidence for both external and discriminant validity. Taking the finding from both studies, it is of the opinion of the researcher that the FOC scale provides a more accurate and useful measure for PSMS beliefs than any of the existing measures.

The current research also provided a means of conceptual unification for PSMS beliefs by constructing a theoretically sound, psychometrically informative, and empirically valid measure for them. This was achieved by defining PSMS beliefs as confusions of ontology, or CKCs (Lindeman & Aarnio, 2007). This definition set guidelines for what could be considered PSMS. Moreover, it controlled for cultural biases (e.g., references to the Loch Ness monster or other culturally distinct concepts) and

subjective opinions (e.g., whether the possibility of aliens are a supernatural belief or a scientific possibility). With these qualities, the FOC scale possessed content and structural validity.

The NRM and the Wald test were utilized to empirically revise the FOC scale. This process yielded a measure that was both psychometrically informative and parsimonious. The FOC was found to provide information over the range of the latent trait. The FOC was parsimonious because redundant and uninformative category responses were eliminated. Moreover, the choice of optimal response option formats were data driven and not subjective. This resulted in reduced cognitive demand on the participant while taking the FOC scale.

The external, discriminant, and substantive validity were assessed. While issues obtaining a clear assessment of substantive validity were present, overall, the FOC was found to be valid. That is, evidence for external and discriminant validity were found indicating that the FOC measured what it is suppose to measure.

Importantly, the FOC was not a measure of religiosity, despite the conceptual overlaps in certain beliefs. The researcher was careful to make those distinctions clear. Ontological confusions that also happened to be central beliefs to some of the monotheistic religions were removed during the initial EFA. Assessing religiosity as religious commitment among the religious, the FOC scale scores were found to have no relationship to religiosity. Thus, the FOC served as a measure of ontological confusions not religiosity.

A major limitation of the current research was the test of substantive validity in Study 2. This does not call into question the validity of the FOC scale as a whole, but

simply reduces the amount of validation evidence. In order to correct for that issue in the future, a control experiment should be undertaken in which the ability to manipulate the image and the time spent on each image could be controlled for. Moreover, the scale needs to have predicted relationships tested in a more rigorous fashion to add more substantive support for the validity of the FOC scale.

Conclusion

The research set out to produce a better measure for PSMS beliefs than the leading measure, the R-PBS. The FOC scale surpassed the R-PBS in several ways. First of all, the FOC scale is unidimensional. While the author of the R-PBS insists that the overall sum score is a good measure of general PSMS beliefs (Tobacyk, 2004), no empirical justification has been provided. In fact, the scale has seven subscales assessing the different factors of the R-PBS. What is troubling is that the R-PBS and its subscales may only serve as a measure for whether or not individuals believe in certain popular superstitions or supernatural phenomena and not as a measure of an overriding tendency toward paranormal thinking.

This leads to the second improvement of the FOC scale. The FOC scale was constructed to have reduced biases and to be culturally neutral. While it is possible that some cultural bias may be present, the FOC does not contain items that reference specific cultural beliefs like the R-PBS. For example, the R-PBS makes reference to black cats, unlucky numbers, the Loch Ness monster, and the abominable snowman—all items with clear cultural and possible generational biases. Furthermore, the vague inclusionary criteria of the R-PBS allowed for items that are questionably PSMS. Specifically, the item stating, “There is life on other planets.” As previously mentioned, the issue of life on

other planets is a question entertained by science as a distinct possibility and considering the existence of human life, is plausible and potentially discoverable. Thus, the inclusion of that item as a PSMS belief is questionable. By using a strict theoretical definition of inclusion, the FOC scale limited internal confounds and yielded a more pure measure of PSMS beliefs.

The inability of the R-PBS to disentangle religious beliefs and paranormal beliefs is another limiting factor for the R-PBS. As found in previous research (Aarnio & Lindeman, 2007; Barber, 2014) and in the current study, religiosity and belief in the paranormal are not synonymous, nor even necessarily related. This distinction is especially important in light of the present emergence of the non-religious. Over the last several decades, individuals have been self-identifying away from religions (Glenn, 2013). These "non-religious" currently comprise about a fifth of the United States population (Pew Research Center, 2012). This is important because the non-religious are not necessarily non-PSMS believers, in fact, many consider themselves spiritual and hold certain PSMS beliefs to be true (Glenn, 2013). On the same note, religious people are not necessarily emphatic believers in the paranormal. Since the FOC scale is a measure of PSMS beliefs that is not religiously bias, it can be used to study differences in the growing religious/non-religious divide. Moreover, it can be employed in research into the emerging psychology of religion.

In conclusion, the FOC scale can be used as a more pure measure of PSMS beliefs and can be used to help bridge the current disparate research into PSMS beliefs. Since the FOC scale is built around the CKCs theory, future research should begin to test some of the assumptions and propositions put forward by the theory. For example, the CKCs

framework assumes that CKCs develop over time and are remnants from childhood (Lindeman & Aarnio, 2007). Future research could test this by looking at differences in age or conducting a longitudinal study of PSMS believers.

APPENDIX A

INFORMED CONSENT

Informed Consent

California State University, Fullerton

Consent to Act as a Research Subject

Measuring Beliefs About the Nature of Reality

You are being asked to participate in a research study. Before you give your consent to volunteer, it is important that you read the following information and ask as many questions as necessary to be sure that you understand what you will be asked to do.

Investigators: The research project is being conducted by Fred J. Pasquarella, B. A., Graduate student at California State University, Fullerton under the supervision of Kathleen Preston, PhD., Assistant Professor at California State University, Fullerton in fulfillment of Master's of Arts in Psychology Thesis.

Purpose of the Study: The purpose of the study is to develop and validate a measure for beliefs about the nature of reality. You will be one several hundred participants recruited from the Research Pool at California State University, Fullerton.

Description of the Study: The study contains a series of questions that ask about beliefs and behaviors regarding reality, preferences toward cognitive effort, as well as some demographic information (such as age, gender, ethnicity, etc.). There are no correct or incorrect answers. Please answer honestly and to the best of your ability. All answers are anonymous, as you will not be asked to provide a name or other identifying information. The entire questionnaire will take approximately 30 minutes.

Risks or Discomforts: There are no health risks or discomforts associated with participating in this project. Because of the personal nature of the questions that will be asked during the questionnaire, you may reflect on unpleasant experiences. If you begin to feel uncomfortable, you may discontinue participation at any time, either temporarily or permanently, and it will not affect your relationship with the researcher or the institution.

Benefits of the Study: This study will help develop new measures and assess preexisting measures regarding how people view reality. The study may yield insights into how certain views of reality are developed and/or the factors related to these views. However, there is no guarantee that you will not receive any direct benefits from participating in this study.

Confidentiality: Confidentiality will be maintained to the extent allowed by law. Your name nor any personally identifying information will be retained by the researchers. Therefore, complete anonymity will be maintained during the entire course of your participation.

Data Storage: Data collected will be stored in a password-protected computer in which only the principal investigator and the faculty adviser will have access. No identifying information will be retained. Data will be retained indefinitely for future use. The data may be used for educational purposes or for later secondary analysis.

Incentives to Participate: You may be offered course credit for completion of this study as a Psychology 101 requirement based upon the length of the questionnaire. A one-hour credit will be assigned for completion of the questionnaire. The researchers will offer no other incentives to participate.

Voluntary Nature of Participation: Participation in this study is completely voluntary. Your choice of whether or not to participate will not influence your future relations with California State University,

Fullerton. If you decide to participate, you are free to withdraw your consent and to stop your participation at any time without penalty or loss of benefits to which you are entitled.

Questions about the Study: If you have any questions about this study, you may email the investigator Fred Pasquarella at fredpasquarella@csu.fullerton.edu or Kathleen Preston, PhD. at kpreston@fullerton.edu.

If you have questions about the rights of human research participants contact the CSUF IRB Office at (657) 278-7640 or irb@fullerton.edu.

Your consent below indicates that you have read the information in this document and have had a chance to ask any questions you may have about the study. Your consent also indicates that you agree to be in the study and have been told that you can change your mind and withdraw your consent at any time. You have been told that by agreeing to this consent form you are not giving up any of your legal rights.

Please indicate if you understand what is being asked of you and that you give your consent to participate. Consent indicates that you have read the information in this document, that you agree to be in the study and have been told that you can change your mind and withdraw your consent at any time.

_____I am at least 18 years old, I understand my rights as a participant, and I agree to participate.

_____I do not wish to participate.

APPENDIX B

THE FULLERTON ONTOLOGICAL CONFUSION SCALE ORIGINAL ITEM POOL

The Fullerton Ontological Confusion (FOC) Scale (Items 1-20)

1	There is life after death.
2	Thunder is angry
3	Negative thoughts can cause toxins to form in the body
4	Ghosts or spirits can interact with the physical world.
5	Certain people can cast spells (e.g. do magic, curse someone, etc.).
6	It is possible for people to have lived in a previous life.
7	Flowers are happy about the summer
8	It is possible to move material objects with only one's thoughts.
9	Plants are happy to receive water
10	When I make a wish I believe something or someone can hear it.
11	The universe has a purpose for me.
12	It is possible for some people to experience future events before they happen.
13	I have a destiny that has already been decided.
14	My life is guided by a non-physical being or entity (e.g. God, angels, positive energy).
15	It is possible for people to communicate with the dead.
16	Praying for people can heal them.
17	Bad things happen to certain people because they attract negative energy.
18	Nothing in life happens without a higher purpose.
19	Human beings possess something non-physical (e.g. a soul) that makes them different from animals.
20	Looking up a person's zodiac sign is a good way to gather information about who they are.

Note. Responses are scored with a 5-point scale ranging from 1 = *Strongly Disbelieve* to 5 = *Strongly Believe.* R indicates reverse scoring.

21	Certain items (e.g. crosses, magnetic bracelets, amulets, etc.) can help protect people.
22	There is no inherent purpose to life. r
23	Good things happen to some people because of positive energy.
24	Energy can be good or bad.
25	Someone designed the world.
26	The position of the planets can influence the events in my life.
27	Certain natural events such as eclipses can be an omen, or signs of something bad to come.
28	One's consciousness continues to exist even if one's brain is destroyed.
29	Natural disasters, such as earthquakes or tsunamis, can occur as a form of retribution for certain human's actions or beliefs.
30	Weather can be affected by the emotions of individuals or groups
31	People can feel the presence of friends or family after they have died
32	Deep thought, meditation, or concentration can cure other peoples illnesses
33	A person can influence the physical world solely with their thoughts
34	Animals were created to serve a purpose
35	Positive thoughts can cleanse the body of toxins
36	The essence, or spirit, of a person can remain in objects her or she came into contact with when they were alive, such as, articles of clothing, books, etc.
37	Mountains were created for a purpose
38	Energy lives in nature
39	A rock regrets that it cannot move
40	Thinking about an event makes it more likely to happen

Note. Responses are scored with a 5-point scale ranging from 1 = *Strongly Disbelieve* to 5 = *Strongly Believe.* R indicates reverse scoring.

APPENDIX C

THE 26-ITEM FOC SCALE

The 26-Item FOC Scale

1	Negative thoughts can cause toxins to form in the body
2	Ghosts or spirits can interact with the physical world
3	Certain people can cast spells (e.g. do magic, curse someone, etc.)
4	It is possible for people to have lived in a previous life
5	It is possible to move material objects with only one's thoughts.
6	When I make a wish I believe something or someone can hear it
7	The universe has a purpose for me
8	It is possible for some people to experience future events before they happen
9	I have a destiny that has already been decided
10	It is possible for people to communicate with the dead
11	Bad things happen to certain people because they attract negative energy
12	Looking up a person's zodiac sign is a good way to gather information about who they are
13	Certain items (e.g. crosses, magnetic bracelets, amulets, etc.) can help protect people
14	Good things happen to some people because of positive energy
15	Energy can be good or bad
16	The position of the planets can influence the events in my life
17	Certain natural events such as eclipses can be an omen, or signs of something bad to come
18	One’s consciousness continues to exist even if one’s brain is destroyed
19	Natural disasters, such as earthquakes or tsunamis, can occur as a form of retribution for certain human’s actions or beliefs
20	Weather can be affected by the emotions of individuals or groups
21	People can feel the presence of friends or family after they have died
22	Deep thought, meditation, or concentration can cure other peoples illnesses
23	A person can influence the physical world solely with their thoughts
24	Positive thoughts can cleanse the body of toxins
25	The essence, or spirit, of a person can remain in objects her or she came into contact with when they were alive, such as, articles of clothing, books, etc.
26	Energy lives in nature

Note. Responses are scored with a 5-point scale ranging from 1 = *Strongly Disbelieve* to 5 = *Strongly Believe.* R indicates reverse scoring.

APPENDIX D

THE FULLERTON ONTOLOGICAL CONFUSION SCALE

The Fullerton Ontological Confusion Scale

The following are statements regarding your personal beliefs about the nature of reality. Please rate the extent to which you BELIEVE or DISBELIEVE each statement.

These statements are to be read literally not metaphorically. There are no right or wrong answers. Please answer as honestly as possible. Remember, your answers are anonymous.

(1) It is possible to move material objects with only one's thoughts.

☐ I believe this statement
☐ I DO NOT believe this statement

(2) When I make a wish I believe something or someone can hear it.

☐ I believe this statement

☐ I DO NOT believe this statement

(3) The universe has a purpose for me.

☐ I believe this statement

☐ I DO NOT believe this statement

(4) Looking up a person's zodiac sign is a good way to gather information about who they are.

☐ I believe this statement

☐ I DO NOT believe this statement

(5) Certain natural events such as eclipses can be an omen, or signs of something bad to come.

☐ I believe this statement

☐ I DO NOT believe this statement

The Fullerton Ontological Confusion Scale

The following are statements regarding your personal beliefs about the nature of reality. Please rate the extent to which you BELIEVE or DISBELIEVE each statement.

These statements are to be read literally not metaphorically. There are no right or wrong answers. Please answer as honestly as possible. Remember, your answers are anonymous.

	Disbelieve	**Unsure**	**Believe**
(6) Ghosts or spirits can interact with the physical world.	☐	☐	☐
(7) Certain people can cast spells (e.g. do magic, curse someone, etc.).	☐	☐	☐
(8) It is possible for people to communicate with the dead.	☐	☐	☐
(9) Certain items (e.g. crosses, magnetic bracelets, amulets, etc.) can help protect people.	☐	☐	☐
(10) Good things happen to some people because of positive energy.	☐	☐	☐
(11) Energy can be good or bad.	☐	☐	☐
(12) Weather can be affected by the emotions of individuals or groups.	☐	☐	☐
(13) People can feel the presence of friends or family after they have died.	☐	☐	☐
(14) Positive thoughts can cleanse the body of toxins.	☐	☐	☐
(15) Energy lives in nature.	☐	☐	☐

The Fullerton Ontological Confusion Scale

The following are statements regarding your personal beliefs about the nature of reality. Please rate the extent to which you BELIEVE or DISBELIEVE each statement.

These statements are to be read literally not metaphorically. There are no right or wrong answers. Please answer as honestly as possible. Remember, your answers are anonymous.

	Strongly Disbelieve	**Disbelieve**	**Believe**	**Strongly Believe**
(16) It is possible for some people to experience future events before they happen.	☐	☐	☐	☐
(17) Bad things happen to certain people because they attract negative energy.	☐	☐	☐	☐
(18) The position of the planets can influence the events in my life.	☐	☐	☐	☐
(19) Deep thought, meditation, or concentration can cure other people's illnesses.	☐	☐	☐	☐
(20) A person can influence the physical world solely with their thoughts.	☐	☐	☐	☐
(21) The essence, or spirit, of a person can remain in objects he or she came into contact with when they were alive, such as, articles of clothing, books, etc.	☐	☐	☐	☐

(22) It is possible for people to have lived in a previous life.

Strongly Disbelieve	**Disbelieve**	**Unsure**	**Believe**	**Strongly Believe**
☐	☐	☐	☐	☐

The Fullerton Ontological Confusion Scale

SCORING GUIDE

Item	Model
1	2PLM
2	2PLM
3	2PLM
4	2PLM
5	2PLM
6	NRM
7	NRM
8	NRM
9	NRM
10	NRM
11	GPCM
12	NRM
13	NRM
14	NRM
15	GPCM
16	NRM
17	NRM
18	NRM
19	GPCM
20	NRM
21	NRM
22	GPCM

Note. 2PLM represents the Two-Parameter Logistic Model; NRM represents the Nominal Response Model (Bock, 1972); GPCM represents the Generalized Partial Credit Model (Muraki, 1992).

APPENDIX E

THE REVISED PARANORMAL BELIEF SCALE

The Revised Paranormal Belief Scale (R-PBS) from Tobacyk, 2004

1	The soul continues to exist though the body may die.
2	Some individuals are able to levitate (lift) objects through mental forces.
3	Black magic really exists.
4	Black cats can bring bad luck.
5	Your mind or soul can leave your body and travel (astral projection).
6	The abominable snowman of Tibet exists.
7	Astrology is a way to accurately predict the future.
8	There is a devil.
9	Psychokinesis, the movement of objects through psychic powers, does exist.
10	Witches do exist.
11	If you break a mirror, you will have bad luck.
12	During altered states, such as sleep or trances, the spirit can leave the body.
13	The Loch Ness monster of Scotland exists.
14	The horoscope accurately tells a person's future.
15	I believe in God.
16	A person's thoughts can influence the movement of a physical object.
17	Through the use of formulas and incantations, it is possible to cast spells on persons.
18	The number "13" is unlucky.
19	Reincarnation does occur.
20	There is life on other planets.
21	Some psychics can accurately predict the future.
22	There is a heaven and a hell.
23	Mind reading is not possible. (R)
24	There are actual cases of witchcraft.
25	It is possible to communicate with the dead.
26	Some people have an unexplained ability to predict the future.

Note. Responses are scored with a 7-point scale ranging from 1 = *Strongly Disagree* to 7 = *Strongly Agree*. The R-PBS is comprised of seven subscales; Traditional Religious Belief (1, 8, 15, 22), Psi (2, 9, 16, 23), Witchcraft (3, 10, 17, 24), Superstition (4, 11, 18), Spiritualism (5, 12, 19, 25), Extraordinary Life Forms (6, 13, 20), and Precognition (7, 14, 21, 26). R indicates reverse scoring.

APPENDIX F

10-ITEM RATIONAL-EXPERIENTIAL INVENTORY

The 10-Item Rational-Experiential Inventory (REI) from Epstein et al., 1996

1	I don't like to have to do a lot of thinking. (R)
2	I try to avoid situations that require thinking in depth about something. (R)
3	I prefer to do something that challenges my thinking abilities rather than something that requires little thought.
4	I prefer complex to simple problems.
5	Thinking hard and for a long time about something gives me little satisfaction.
6	I trust my initial feelings about people.
7	I believe in trusting my hunches.
8	My initial impressions of people are almost always right.
9	When it comes to trusting people, I can usually rely on my "gut feelings."
10	I can usually feel when a person is right or wrong even if I can't explain how I know

Note. Responses are scored with a 5-point scale ranging from 1 = *Completely False* to 5 = *Completely True*. The REI is comprised of two subscales Need for Cognition scale (items 1-5) and the Faith in Intuition (items 6-10). R indicates reverse scoring.

APPENDIX G

THE RELIGIOUS COMMITMENT INVENTORY

The Religious Commitment Inventory—10 from Worthington et al., 2003	
1	I often read books and magazines about my faith.
2	I make financial contributions to my religious organization.
3	I spend time trying to grow in understanding of my faith.
4	Religion is especially important to me because it answers many questions about the meaning of life.
5	My religious beliefs lie behind my whole approach to life.
6	I enjoy spending time with others of my religious affiliation.
7	Religious beliefs influence all my dealings in life.
8	It is important to me to spend periods of time in private religious thought and reflection.
9	I enjoy working in the activities of my religious organization.
10	I keep well informed about my local religious group and have some influence in its decisions.

Note. Responses are scored with a 5-point scale ranging from 1 = *Not at all* to 5 = *Totally*.

APPENDIX H

MARLOWE-CROWNE SOCIAL DESIRABILTY SCALE, SHORT FORM C

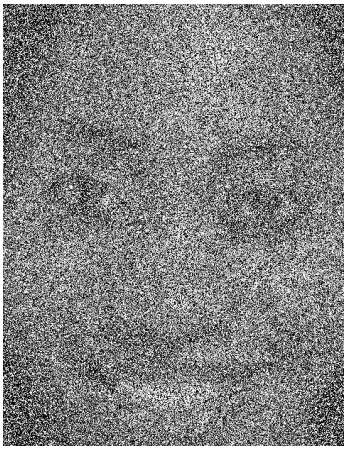
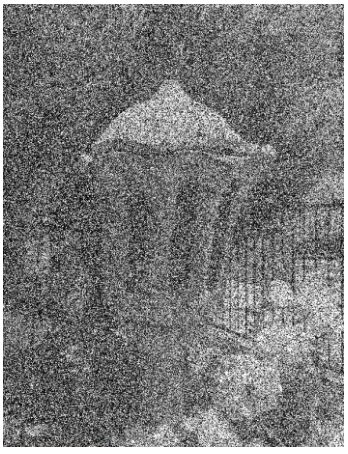
Short Form C of the Marlowe-Crowne Social Desirability Scale from Reynolds, 1982

1	It is sometimes hard for me to go on with my work if I am not encouraged. (R)
2	I sometimes feel resentful when I don't get my way (R)
3	On a few occasions, I have given up doing something because I thought too little of my ability (R)
4	There have been times when I felt like rebelling against people in authority even though I knew they were right. (R)
5	No matter who I'm talking to, I'm always a good listener.
6	There have been occasions when I took advantage of someone.
7	I'm always willing to admit it when I make a mistake.
8	I sometimes try to get even rather than forgive and forget. (R)
9	I am always courteous, even to people who are disagreeable.
10	I have never been irked when people expressed ideas very different from my own
11	There have been times when I was quite jealous of the good fortune of others (R)
12	I am sometimes irritated by people who ask me favors (R)
13	I have never deliberately said something that hurt someone's feelings

Note. Responses are scored with as 1= *True* and 0 = *False*. R indicates reverse scoring.

APPENDIX I

EXAMPLE STIMILI FOR THE ILLUSORY FACE PERCEPTION TASK

	Original Image	80% Degraded Image
Face Image		
	Original Image	80% Degraded Image
Non Face Image		

REFERENCES

Aarnio, K., & Lindeman, M. (2005). Paranormal beliefs, education, and thinking styles. *Personality and Individual Difference, 39,* 1227-1236. doi:10.1016/j.paid.2005.04.009

Aarnio, K., & Lindeman, M. (2007). Religious people and paranormal believers: Alike or different? *Journal of Individual Differences, 28*(1), 1-9. doi:10.1027/1614-0001.28.1.1

Addi, H. (2007) Signal Detection Theory (SDT). In N. Salkind (Ed.) *Encyclopedia of Measurement and Statistics* (pp. 887-890), Thousand Oaks, CA: Sage.

Amarnani, R. (2009). Two theories, one theta: A gentle introduction to item response theory as an alternative to classical test theory. *The International Journal of Education and Psychological Assessment, 3,* 104-109.

Auton, H. R., Pope, J., & Seeger, G. (2003). It isn't that strange: Paranormal belief and personality traits. *Social Behavior and Personality, 31*(7), 711-720. doi:10.2224/sbp.2003.31.7.711

Barber, J. (2014). Believing in a purpose of events: Cross-cultural evidence of confusions in core knowledge. *Applied Cognitive Psychology, 28,* 432-437. doi:10.1002/acp.3003

Beck, J., & Forstmeier, W. (2007). Superstition and belief as inevitable by-products of an adaptive learning strategy. Human Nature, 18, 35–46. doi:10.1007/BF02820845

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37,* 29-51.

Borsboom, D. (2006) The attack of the psychometricians. *Psychometrika, 71*(3), 425-440.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of Royal Statistical Society, 26,* 211-246.

Broad, C. D. (1953). The relevance of psychical research to philosophy. In J. Ludwig (Ed.), *Philosophy and parapsychology* (pp. 43- 63). Buffalo, NY: Prometheus.

Browne, M. W. (2014). CEFApak: Comprehensive Exploratory Factor Analysis [Software]. Retrieved from http://faculty.psy.ohio-state.edu/browne/software.php.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21,* 230–258.

Buhrmester, M. Kwang, T., Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5. doi:10.1177/1745691610393980

Cai, L. (2013). flexMIRT version 2: Flexible multilevel multidimensional item analysis and test scoring [Computer software]. Chapel Hill, NC: Vector Psychometric Group.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39,* 752-766.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Chen, W. H., & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22,* 265–289.

Champely, S. (2012). pwr: Basic functions for power analysis [Software]. Retrieved from http://CRAN.R-project.org/package=pwr

Chi, M. T. H., Slotta, J. D., & de Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction, 4*, 27-43.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319.

Cornwall, M., Albrecht, S. L., Cunnignham, P. H., & Pitcher, B. L. (1986). The dimensions of religiosity: A conceptual model with an empirical test. *Review of Religious Research, 27*(3), 226-244.

Corso, R. A. (2014). Avoid black cats? Walk around ladders? Are Americans superstitious? Most Americans actually do not believe in any of the 12 common superstitions. The Harris Poll. Retrieved from http://www.harrisinteractive.com/NewsRoom/HarrisPolls/tabid/447/ctl/ReadCustom%20Default/mid/1508/ArticleId/1388/Default.aspx

Crowne, D. P., & Marlowe, D. A. (1960). A new scale of social desirability independent of pathology. *Journal of Consulting Psychology, 24,* 349-354.

de Ayala, R. J., & Sava-Bolesta, M. (1999). Item parameter recovery for the nominal response model. *Applied Psychological Measurement, 23*(3), 3-20. doi:10.1177/01466219922031130

de Ayala, R. J. (2009). The theory and practice of item response theory. New York, NY: Guilford Press.

de Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical Education, 44*, 109-117.

Dagnall, N., Parker, A., Munley, G. (2007). Paranormal belief and reasoning. *Personality and Individual Differences, 43*, 1406-1415. doi:10.1016/j.paid.2007.04.017

DeMars, C. E. (2003). Sample size and the recovery of nominal response model item parameters. *Applied Psychological Measurement, 27,* 275-288. doi:10.1177/0146621603027004003

Eckblad, M., & Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. *Journal of Consulting and Clinical Psychology*, *51*, 215-225. doi:10.1037/0022-006X.51.2.215

Embretson, S. E., & Reise, S. P. (2009). *Item Response Theory for Psychologists.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology, 71*(2), 390-405.

Gallagher, C., Kumar, V. K., & Pekala, R. J. (1994). The Anomalous Experiences Inventory: Reliability and validity. *Journal of Parapsychology, 58*, 402–428.

Glenn, H. (2013). Losing our religion: The growth of the 'Nones.' *National Public Radio: America.* Retrieved from http://www.npr.org/blogs/thetwoway/2013/01/14/169164840/losing-our-religion-the-growth-of-the-nones

Goulding, A., & Parker, A. (2001). Finding psi in the paranormal: Psychometric measures used in research in paranormal beliefs/experiences and in research on psi-ability. *European Journal of Parapsychology, 16*, 73–101.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practices, 12*(3), 38-47.

Hartman, S. E. (1999). Another view of the Paranormal Belief Scale. *Journal of Parapsychology*, *63*(2), 131-141.

Hergovich, A. & Arendasy, M. (2005). Critical thinking ability and belief in the paranormal. *Personality and Individual Differences, 38,* 1805-1812. doi: 10.1016/j.paid.2004.11.008

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin, 30*(2), 161-172. doi: 10.1177/0146167203259930

John, O. P., & Soto, C. J. (2007) The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, R. C. Fralea, & R. F. Krueger (Eds.). *Handbook of Research Methods in Personality Psychology* (pp. 461-494). New York: Guilford.

Kahneman, D. (2011). *Thinking fast and slow*. New York, NY: Farrar, Straus and Giroux.

Lawrence, T. R. (1996). How many factors of paranormal belief are there? A critique of the Paranormal Belief Scale. *Journal of Parapsychology, 59*(1), 3-23. doi: 1996-32735-001

Lindeman, M. (2011). Biases in intuitive reasoning and belief in complementary and alternative medicine. *Psychology and Health, 26*(3*),* 371-382. doi:10.1080/08870440903440707

Lindeman, M., & Aarnio, K. (2007). Superstitious, magical, and paranormal beliefs: An integrative model. *Journal of Research in Personality, 41*, 731-744. doi:10.1016/j.jrp.2006.06.009

Lindeman, M., Cederstrom, S., Simola, P., Simula, A., Ollikainen, S., & Riekki, T. (2008). Sentences with core knowledge violations increase the size of N400 among paranormal believers. *Cortex, 44*, 1307-1315. doi10.1016/j.cortex.2007.07.010

Lindeman, M., & Saher, M. (2007). Vitalism, purpose, and superstition. *British Journal of Psychology, 98,* 33-44. doi: 10.1 348/000712606X 101808

Lindeman, M., & Svedholm, A. M. (2012). What's in a term? Paranormal, superstitious, magical and supernatural beliefs by any other name would mean the same. *Review of General Psychology, 16*(3), 241-255.doi:10.1037/a0027158

Little, R. J. A. (1988). A test of missing completely at random fro multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198-1202.

Macmillan, N. A., Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16,* 159-176.

Nemeroff, C., & Rozin, P. (2000). The makings of the magical mind: The nature and function of sympathetic magical thinking. In K. S. Rosengren, C. N. Johnson, & P. L. Harris (Eds.), *Imagining the impossible: Magical, scientific, and religious thinking in children* (pp. 1–34). New York: Cambridge University Press.

Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*. Thousand Oaks, CA: Sage Publications.

Paranormal. (n.d.). In Oxford English dictionary. Retrieved from http://www.oed.com.libproxy.fullerton.edu

Pew Research Center (2012). "Nones" on the rise. Pew Research Center: Religion & Public Life. Retrieved from http://www.pewforum.org/2012/10/09/nones-on-the-rise/

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research, 48,* 28-56. doi:10.1080/00273171.2012.710386

Preston, K. S. J. (2014a) Advanced Topics in IRT: Evaluating the effectiveness of each response option with the nominal response model. PowerPoint presentation at the 94th Annual Western Psychological Association. Retrieved from http://hssfaculty.fullerton.edu/psychology/kpreston/Plotting.txt

Preston, K. S. J. (2014b) Advanced Topics in IRT: Evaluating the effectiveness of each response option with the nominal response model. *PowerPoint presentation at the 94th Annual Western Psychological Association*. Retrieved from http://hssfaculty.fullerton.edu/psychology/kpreston/Wald.txt

Preston, K. S. J., & Reise, S. P. (2015). Detecting faulty within-item category functioning with the nominal response model. In S. P. Reise & D. Revicki (Eds.). *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment* (pp. 386-405). New York, NY: Routledge.

Preston, K. S. J., Reise, S. P., Cai, L., & Hays, R. D. (2011). Using the nominal response model to evaluate response category discrimination in the PROMIS Emotional Distress item pools. *Educational and Psychological Measurement, 71*, 523-550.

R Core Team. (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* Retrieved from http://www.R- project.org/.

Ramos, R. (2013) *iNoisify-Add noise to any image online*. iNoisify.com.

Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item response theory: Fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science, 14*(2), 95-101.

Reise, S. P., & Henson, J. M. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of Personality Assessment, 81*(2), 93-103.

Reise, S. P., & Waller, N. G. (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement, 14*(1), 45-58. doi:0146-6216/90/010045-14

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 38*(1), 119-125. doi:10.1002/1097-4679(198201)38:1

Riekki, T., Lindeman, M., Aleneff, M., Halme, A., & Nuortimo, A. (2013). Paranormal and religious believers are more prone to illusory face perception than skeptics and non-believers. *Applied Cognitive Psychology*, *27*, 150-155. doi:10.1002/acp.2874

Riekki, T., Lindeman, M., & Lipsanen, J. (2013). Conceptions about the mind-body problem and their relations to afterlife beliefs, paranormal beliefs, religiosity and ontological confusions. *Advances in Cognitive Psychology, 9*(3), 112-120. doi:10.5709/acp-0138-5

Shermer, M. (2008). Patternicity: Finding meaningful patterns in meaningless noise. *Scientific American.* Retrieved from http://www.scientificamerican.com/article/patternicity-finding-meaningful-patterns/

Smith, A. F., & Simmonds, J. G. (2006). Help-seeking and paranormal beliefs in adherents of mainstream religion, alternative religion, and no religion. *Counseling Psychology Quarterly, 19*(4), 331-341. doi:10.1080/09515070601095724

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25,* 173–180.

Supernatural. (n.d.). In Oxford English dictionary. Retrieved from http://www.oed.com.libproxy.fullerton.edu

Superstitious. (n.d.). In Oxford English dictionary. Retrieved from http://www.oed.com.libproxy.fullerton.edu

Svedholm, A. (2013). *The cognitive basis of paranormal, superstitious, magical, and supernatural beliefs: The roles of core knowledge, intuitive and reflective thinking, and cognitive inhibition* (Unpublished doctoral dissertation). University of Helsinki, Finland.

Svedholm, A. K., & Lindeman, M. (2013). The separate roles of the reflective mind and involuntary inhibitory control in gatekeeping paranormal beliefs and the underlying intuitive confusions. *British Journal of Psychology, 104*, 303-319. doi:10.1111/j.2044-8295.2012.02118.x

Svedholm, A. M., Lindeman, M., & Lipsanen, J. (2010). Believing in the purpose of events-Why does it occur, and is it supernatural? *Applied Cognitive Psychology, 24*, 252-265. doi:10.1002/acp.1560

Thalbourne, M. A. (1995) Further studies of the measurement and correlates of belief in the paranormal. *Journal of the American Society for Psychical Research, 89*, 233-247.

Thalbourne, M. A. (2001), Measures of the sheep-goat variable, transliminality, and their correlates. *Psychological Reports, 88*, 339-350.

Thalbourne, M. A., & Delin, P. S. (1993). A new instrument for measuring Sheep-Goat variable: Its psychometric properties and factor structure. *Journal of the Society for Psychical Research, 59*, 172–186.

Thissen, D., & Steinber, L. (1986). A taxonomy of item response models. *Psychometrika, 51*(4), 567-577.

Tobacyk, J. J. (2004). A revised paranormal belief scale. *The International Journal of Transpersonal Studies*, *24*, 94-98.

Tobacyk, J., & Milford, G. (1983). Belief in paranormal phenomena: Assessment instrument development and implications for personality functioning. *Journal of Personality and Social Psychology, 44*, 1029– 1037. doi:10.1037/0022-3514.44.5.102

United States Census Bureau (n.d.). *Regions and Divisions.* Retrieved from https://www.census.gov/econ/census/help/geography/regions_and_divisions.html

Ura, M., Preston, K. S. J., & Mearns, J. (2015). A Measure of Prejudice Against Accented English (MPAAE): Scale Development and Validation. *Journal of Language and Social Psychology 34*(5), 539-563. doi:10.1177/0261927X15571537

Vyse, S. A. (1997). *Believing in magic: The psychology of superstition*. New York, NY: Oxford University Press.

Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annuals of Mathematical Statistics, 16*(2), 117-186. doi:10.1214/aoms/1177731118

Whitley, B. E, & Kite, M. E. (2013). Developing a measurement strategy. In *Principles of research in behavioral science* (3rd ed.) (pp. 146-183). New York, NY: Routledge.

Worthington, E. L., Wade, N. G., Hight, T. L., Ripley, J. S., McCullough, M. E., Berry, J. W. . . , Bursley, K. H. (2003). The Religious Commitment Inventory-10: Development, refinement, and validation of a brief scale for research and counseling. *Journal of Counseling Psychology*, *50*(1), 84-96.