GROSS, GEORGINA M., Ph.D. Development and Psychometric Properties of the Multidimensional Schizotypy Scale: A New Measure for Assessing Positive, Negative, and Disorganized Schizotypy. (2017)
Directed by Dr. Thomas R. Kwapil. 116 pp.

This dissertation reports on the development of a new self-report questionnaire measure of schizotypy – the Multidimensional Schizotypy Scale (MSS). Schizotypy offers a useful and unifying construct for understanding schizophrenia-spectrum psychopathology. Questionnaire measures have been widely used to assess schizotypy and have greatly informed our understanding of the construct; however, available measures suffer from a number of limitations, including lack of a clear conceptual framework, outdated wording, unclear factor structure, and psychometric shortcomings. The MSS is based on current conceptual models and taps positive, negative, and disorganized conceptual dimensions of schizotypy. The derivation sample included 6,265 participants sampled from four universities and Amazon Mechanical Turk. A separate cross-validation sample of 1,000 participants from these sources was used to examine the psychometric properties of the final subscales. Scale development employed classical test theory, item response theory, and differential item function methods. The positive schizotypy and negative schizotypy subscales contain 26 items each, and the disorganized schizotypy subscale contains 25 items. The psychometric properties were almost identical in the derivation and validation samples. All three subscales demonstrated good to excellent reliability, high item-scale correlations, and good item and test curve characteristics.

DEVELOPMENT AND PSYCHOMETRIC PROPERTIES OF THE MULTIDIMENSIONAL SCHIZOTYPY SCALE: A NEW MEASURE FOR ASSESSING POSITIVE, NEGATIVE, AND DISORGANIZED SCHIZOTYPY

by

Georgina M. Gross

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2017

Approved by

_____
Committee Chair

APPROVAL PAGE

This dissertation written by GEORGINA M. GROSS has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____

Committee Members _____

_____

_____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

ACKNOWLEDGEMENTS

I would like to acknowledge my lab mates, Charlotte Chun and Sarah Sperry for assistance with item development and comments on the manuscript. Many thanks to Sarah Sperry and Drs. Mike Raulin and Chris Burgin for assistance with data collection. I thank my committee members for their guidance and expertise, and the university for the Bernard Dissertation Fellowship. Finally, I thank my mentor for his unwavering guidance and support.

TABLE OF CONTENTS

Page

LIST OF TABLES v

LIST OF FIGURES vii

CHAPTER

I. INTRODUCTION 1

Overview 1
Schizotypy and Schizophrenia 1
Psychometric Measurement of Schizotypy 6
Limitations of Current Scales 11
Goals of the Present Study 19

II. METHODS 21

Participants 21
Materials 24
Procedures 26

III. RESULTS 34

Item-Level Results 34
Subscale Results 38

IV. DISCUSSION 43

Strengths of the MSS Development 44
Strengths of the MSS Final Scale 48
Limitations 53
Future Directions 59

REFERENCES 65

APPENDIX A. TABLES AND FIGURES 89

LIST OF TABLES

Page

Table 1. Summary of Item Administrations 89

Table 2. Demographics for Derivation and Cross-Validation Samples 90

Table 3. Item-level CTT Statistics from the Derivation Sample for the MSS Positive Schizotypy Subscale 91

Table 4. Item-level CTT Statistics from the Derivation Sample for the MSS Negative Schizotypy Subscale 92

Table 5. Item-level CTT Statistics from the Derivation Sample for the MSS Disorganized Schizotypy Subscale 93

Table 6. Item Level CTT Statistics from Derivation Sample for Survey 3 Items Eliminated from Final Subscales 94

Table 7. Item-level CTT Statistics from the Cross-Validation Sample (N = 1,000) for the MSS Positive Schizotypy Subscale 96

Table 8. Item-level CTT Statistics from the Cross-Validation Sample (N = 1,000) for the MSS Negative Schizotypy Subscale 97

Table 9. Item-level CTT Statistics from the Cross-Validation Sample (N = 1,000) for the MSS Disorganized Schizotypy Subscale 98

Table 10. IRT Model Fit Statistics in the Derivation and Cross-Validation Samples 99

Table 11. Item-level IRT/DIF Statistics from the Derivation Sample for the MSS Positive Schizotypy Subscale 100

Table 12. Item-level IRT/DIF Statistics from the Derivation Sample for the MSS Negative Schizotypy Subscale 101

Table 13. Item-level IRT/DIF Statistics from the Derivation Sample for the MSS Disorganized Schizotypy Subscale 102

Table 14. Item-level IRT/DIF Statistics from the Cross-Validation (N = 1,000) Sample for the MSS Positive Schizotypy Subscale 103

Table 15. Item-level IRT/DIF Statistics from the Cross-Validation Sample (N = 1,000) for the MSS Negative Schizotypy Subscale 104

Table 16. Item-level IRT/DIF Statistics from the Cross-Validation Sample (N = 1,000) for the MSS Disorganized Schizotypy Subscale..... 105

Table 17. Item Level IRT/DIF Statistics from Derivation Sample for Survey 3 Items Eliminated from Final Subscales 106

Table 18. 2PL Item Fit Statistics for Positive, Negative, and Disorganized Subscales in Cross-Validation Sample..... 108

Table 19. Descriptive Statistics and Reliability of the Schizotypy Subscales 109

Table 20. Intercorrelations of the Subscales and Correlations with Neuroticism..... 110

Table 21. Exploratory Factor Analysis Eigenvalues..... 111

Table 22. Confirmatory Factor Analyses Fit Indices..... 112

LIST OF FIGURES

Page

Figure 1. Test Information and Standard Error Curves for the Three Subscales..... 113

Figure 2. Test Characteristic Curves for Men (Group 1) and Women (Group 2)..... 114

Figure 3. Test Characteristic Curves for White (Group 1) and Not White (Group 2) 115

Figure 4. Test Characteristic Curves for White (Group 1) and Black/AA (Group 2)..... 116

CHAPTER I

INTRODUCTION

Overview

This dissertation reports on the development of a new questionnaire measure of schizotypy – the Multidimensional Schizotypy Scale (MSS; Kwapil, Gross, Silvia, Raulin, & Barrantes-Vidal, 2017). Questionnaire measures have been widely used to assess schizotypy and have greatly informed our understanding of the construct. However, available measures suffer from a number of limitations, including lack of a clear conceptual framework, outdated wording, unclear factor structure, and psychometric shortcomings. The MSS was based on current conceptual models of the construct to tap positive, negative, and disorganized dimensions of schizotypy. The scale development was based on a large and diverse sample, and employed classical test theory (CTT), item response theory (IRT), and differential item functioning (DIF) methodologies.

Schizotypy and Schizophrenia

Schizophrenia is the most severe manifestation of a family of psychiatric disorders defined by abnormalities in one or more of the following domains: delusions, hallucinations, disorganized thinking (speech), grossly disorganized or abnormal motor behavior (including catatonia), and negative symptoms. Schizophrenia-spectrum disorders include psychotic illnesses such as schizoaffective and delusional disorder, as

well as nonpsychotic conditions such as schizotypal, schizoid, and paranoid personality disorder (Diagnostic and Statistical Manual of Mental Disorders, 5th edition, DSM-5; American Psychiatric Association, 2013). Schizophrenia has a lifetime prevalence of approximately 1% and the broader spectrum of disorders has a prevalence of approximately 3 to 5%, with slightly higher rates for men (Menezes, 2009). Treatments provide symptomatic relief for many people, but they are not curative and many patients experience a chronic and episodic course associated with marked impairment in functioning (Cornblatt et al., 1999). Based on the National Comorbidity Survey, Wu et al. (2005) estimated the overall economic impact of schizophrenia in 2002 in the United States to be $62.7 billion, and the disorder is associated with severe costs to patients, their families, and society. As such, schizophrenia presents a major public health problem, and research investigating its development and possible prevention is imperative.

Current models indicate that vulnerability for schizophrenia is expressed across a continuum of symptoms and impairment. This continuum is referred to as schizotypy and ranges from subclinical expression to the prodrome to schizophrenia-spectrum personality disorders to full-blown psychosis (Kwapil & Barrantes-Vidal, 2015; Lenzenweger, 2010). Schizotypy is a unifying construct that is useful for understanding mechanisms involved in the transition from predisposition to disorder. The study of at-risk individuals avoids confounds that often accompany schizophrenia such as medication, stigma, and institutionalization. These confounds make it difficult to

disentangle etiologically relevant factors from consequences of the disorders. Further, the schizotypy model should facilitate the identification of endophenotypes that are intermediate to the presentation of the full disorder (Gottesman & Gould, 2003) because these endophenotypes are expected to occur in at-risk individuals who do not have full-blown clinical disorders. Finally, the early identification of schizotypy and the longitudinal study of risk and protective mechanisms for the development of clinical symptoms should inform the development of prophylactic treatment interventions.

Ample evidence suggests that schizophrenia is characterized by marked heterogeneity in etiology, presentation, symptom development, and response to treatment (Andreasen & Carpenter, 1993; Mueser & Jeste, 2008). Further, DSM-5 diagnostic criteria for schizophrenia can be met with a variety of different symptom presentations (i.e., several combinations of distinct symptoms). In short, there is no prototypical schizophrenia patient or course of the illness, and the label of schizophrenia may not provide sufficient information for capturing its varied manifestations. This heterogeneity is also apparent across subclinical levels of the schizotypy continuum. For example, suspicions about the government, minimal interest in hobbies or spending time with other people, and mildly odd speech and eccentric behavior are distinct features that all fall under the broad label of schizotypy.

Dating back to Bleuler (1911/1950) and Kraepelin (1913/1919), the literature has attempted to account for this heterogeneity by proposing various symptom subtypes or dimensions of schizophrenia (Peralta & Cuesta, 2001). Historically these attempts have

involved classical Kraepelinian subtypes such as paranoid, disorganized (hebephrenic), undifferentiated, catatonic, and simple schizophrenia. However, subtype models have been replaced by dimensional approaches given that the schizophrenia subtypes have "poor reliability, low stability over time, and negligible prognostic value" (Tandon et al., 2013, p. 6). Commonly identified dimensions include positive (psychotic), negative (deficit), and cognitive and behavioral disorganization (Andreasen et al., 1994; Arndt, Alliger, & Andreasen, 1991; Lenzenweger, & Dworkin, 1996; Liddle, 1987). Other dimensions have been proposed such as social impairment (Lenzenweger, & Dworkin, 1996), paranoia (Vazquez-Barquero et al., 1996), and antisocial or nonconforming behavior (Harvey et al., 1996).

Consistent with the conceptualization of schizophrenia as the most severe expression of schizotypy, similar factors have been proposed to underlie both (Vollema & van den Bosch, 1995); however, currently there is no universally agreed upon factor structure. Early work using factor analysis of 11 prominent schizotypy scales suggested a general dimension comprised of hallucinatory predisposition, perceptual aberration, nonconformity, social fear, cognitive slippage and borderline traits, and a second dimension of anhedonia (Kelley & Coursey, 1992; Raine & Allbutt, 1989). More recent work has also suggested a two-factor structure of schizotypy with positive (magical ideation and perceptual aberrations) and negative (physical and social anhedonia) factors (e.g. Kwapil et al., 2008); however, these authors posited a third dimension of cognitive and behavioral disorganization that was not tapped by their schizotypy scales.

Others have put forth a three-factor model with cognitive/perceptual, interpersonal, and disorganization dimensions (e.g., Raine, 1991) and still others have suggested four or more factors underlying schizotypy, including unusual experiences, cognitive disorganization, introvertive anhedonia, and impulsive nonconformity/asociality (Mason, 1995), paranoia (e.g., Stefanis et al., 2004), and social withdrawal, unreality, eccentricity, and neuroticism factors (Gruzelier, 1996). Gross and Kwapil (2014) reviewed the literature on the factor structure of schizotypy, and argued for a three-factor model with positive, negative, and disorganized factors, consistent with widely suggested models of schizophrenia. However, further work is needed to develop assessments of this model and test its validity.

Despite these various multidimensional models, research in this area often ignores the heterogeneity of schizophrenia and/or schizotypy altogether. A quick review of the literature reveals that many studies simply divide participants into high or low schizotypy groups based on total scores on a measure. For example, Del Goleto, Kostova, & Blanchet (2016) separated participants into high and low schizotypy groups. They then examined group differences in context processing, mentalizing abilities, and interpersonal functioning; however, these associations may vary drastically dependent upon the dimension(s) of schizotypy represented in the sample. Other studies seem to base their model of schizotypy simply on the measures that they employ, rather than using theory to define multidimensional structure (e.g., Mason, Claridge, & Jackson, 1995). These practices have inhibited reaching a consensus regarding the actual

structure of schizotypy. Further, they result in a disjointed approach in which different research groups use different measures (that may not be comparable), thus limiting the generalizability/comparability of findings across studies to advance knowledge of the construct. Ideally, a model of schizotypy should be based on theory and then tested, and the psychometric approach seems to be a promising pathway to this end.

Psychometric Measurement of Schizotypy

Lenzenweger (2010) reviewed several commonly used approaches for identifying schizotypic psychopathology: clinical methods, consanguinity, and laboratory-psychometric approaches (also see Chapman, Chapman, & Kwapil, 1995; Lenzenweger, 1994). The psychometric approach, or the use of self-report questionnaires on which participants report on aspects of personality, psychopathology, and psychological functioning that are presumed to characterize schizotypic or pre-schizophrenic persons, affords several unique benefits. First, it allows for administration to large numbers of participants from clinical or community (general population) settings and for identifying people across a broad range of the schizotypy continuum. Psychometric assessment is relatively cost- and time-efficient, objective, and non-invasive, and has demonstrated convergent and predictive validity. Most importantly, the psychometric high-risk approach appears to be especially promising for capturing the heterogeneity of schizotypy. Specifically, it expands the study of schizophrenic psychopathology beyond patients and their relatives, and allows for the assessment of a broad range of clinical and subclinical presentations (Lenzenweger, 2010).

Although numerous scales have been developed to assess schizotypy (for reviews see Chapman et al., 1995; Kwapil & Chun, 2015; Mason, 2015; Mason et al., 1997), three scales have been used most prominently and each purportedly captures a slightly different factor structure of schizotypy. The Wisconsin Schizotypy Scales (WSS), made up of the Perceptual Aberration (Chapman, Chapman, & Raulin, 1978), Magical Ideation (Eckblad & Chapman, 1983), Physical Anhedonia (Chapman, Chapman, & Raulin, 1976), and Revised Social Anhedonia (Eckblad et al., 1982) Scales are widely used for assessing schizotypy. The scales contain 166 dichotomous (true/false) items that were based largely on Meehl's (1964) checklist of schizotypy symptoms and were developed using Jackson's (1970) guidelines for personality scale development using classical test theory (CTT). Internal consistency coefficients for the Magical Ideation, Perceptual Aberration, Physical Anhedonia, and Revised Social Anhedonia Scales are good and typically fall in the range of .80 to .90 (Vollema & van den Bosch, 1995). Numerous studies have demonstrated the validity of the WSS through associations with schizophrenia-spectrum symptoms and impaired functioning (e.g., Blanchard et al., 2011; Chapman, Chapman, Kwapil, Eckblad, & Zinser, 1994; Fernandes & Miller, 1995; Kwapil, Barrantes-Vidal, & Silvia, 2008; Park et al., 1995).

Recent studies indicate that a two-factor structure underlies the WSS, with positive and negative schizotypy dimensions (e.g., Brown, et al., 2008; Kwapil et al., 2008; Lewandowski et al., 2006). This factor structure has been replicated in cross-cultural studies (e.g., Chan et al., 2015; Fonseca-Pedrero et al., 2010; Kwapil, Ros-

Morente, Silvia, & Barrantes-Vidal, 2012). The two factors tend to be only modestly associated (e.g., *r* = .11 in 9,316 young adults screened in our lab). The positive and negative schizotypy dimensions are associated with differential patterns of symptoms and impairments in cross-sectional questionnaire (e.g., Lewandowski et al., 2006), interview (e.g., Kwapil et al., 2008; Barrantes-Vidal, Chun, Myin-Germeys, & Kwapil, 2013), laboratory (Kaczorowski, Barrantes-Vidal, & Kwapil, 2009), and experience sampling studies (e.g., Kwapil, Brown, Silvia, Myin-Germeys, & Barrantes-Vidal, 2012; Barrantes-Vidal et al., 2013). Both dimensions predicted the development of schizophrenia-spectrum disorders in a ten-year follow-up study, and positive schizotypy predicted the development of psychotic disorders (Kwapil et al., 2013).

The Schizotypal Personality Questionnaire (SPQ; Raine, 1991), comprised of 74 dichotomous (yes/no) items, assesses schizotypal personality disorder traits using a subscale for each of the nine diagnostic criteria from the Diagnostic and Statistical Manual of Mental Disorders, 3rd edition-revised (DSM-III-R; American Psychiatric Association, 1987). Raine (1991) reported high internal consistency reliability (.91) and test-retest reliability (.82), as well as good convergent, discriminant, and criterion validity. Finally, 55% of participants scoring in the top 10% of the sample had a clinical diagnosis of schizotypal personality disorder (Raine, 1991). Numerous studies support the validity of the SPQ through associations with clinical, functional, and cognitive deficits (e.g., Cohen, Callaway, Najolia, Larsen, & Strauss, 2012; Chen, Hsiao, & Lin, 1997; Park & McTigue, 1997; Raine, Benishay, Lencz, & Scarpa, 1997). Although the SPQ was

originally designed to assess schizotypal personality disorder, it is frequently used more broadly as a measure of schizotypy.

The factor structure of the SPQ has been widely investigated with most support for a three-factor model. Using confirmatory factor analysis (CFA), Raine et al. (1994) reported a three-factor model with cognitive-perceptual, interpersonal, and disorganized factors. Conceptually, the cognitive-perceptual factor maps closely onto positive schizotypy. The interpersonal factor is often used as a measure of negative schizotypy; however, Gross, Mellin, Silvia, Barrantes-Vidal, & Kwapil (2014) raised concerns about its ability to tap deficit schizotypy. Specifically, the interpersonal factor showed moderate correlations with neuroticism (measured by the NEO Five Factor Inventory [NEO-FFI], Costa & McCrae, 1992) equal to or greater than the correlations between the cognitive-perceptual factor and neuroticism. This is likely due in part to the inclusion of both social anxiety and suspiciousness, which are not conceptually a part of deficit schizotypy. Further, the WSS negative factor, but not the SPQ interpersonal factor, was correlated with low openness to experience (measured by the NEO-FFI). Theoretically, low openness to experience, characterized by a dearth of fantasy, emotions, values, and interests, closely relates to the schizoidal nature of negative schizotypy.

The three SPQ factors showed modest to high intercorrelations (cognitive-perceptual and disorganized: r = .71 and .75; disorganized and interpersonal: r = .44 and .60; and cognitive-perceptual and interpersonal: r = .20 and .37) in Raine et al.'s (1994)

two samples. The SPQ Manual (Raine, 2001) recommends these factors be computed using additive formulae summing the subscales. Studies have supported this three-factor model using exploratory (e.g., Fossati et al., 2003) and confirmatory (Chen et al., 1997; Rossi & Daneluzzo, 2002; Suhr & Spitznagel, 2001) factor analysis, and evidence for the validity of these factors has been demonstrated through studies of neurocognition (e.g., Daneluzzo et al., 1998), genetics (Raine & Baker, 1992), and clinical features (e.g., Axelrod, Grilo, Sanislow & McGlashan, 2001). However, other studies reported good fit for a three-factor model only after employing modification indices or model revisions (Bora, & Arabaci, 2009; Reynolds, Raine, Mellingen, Venables, & Mednick 2000; Wuthrich & Bates, 2006), which is problematic when claiming support for an a priori model.

Further, several studies have suggested alternative factor structures, including a four-factor model with a paranoid factor (Bora & Arabaci, 2009; Compton et al., 2009; Stefanis et al., 2004). Using CFA, Gross et al. (2014) reported poor fit for Raine's et al.'s (2014) three-factor model and found best fit for Stefanis et al.'s (2004) four-factor model. Using exploratory factor analysis (EFA) with the nine SPQ subscales, Gross et al. (2014) reported best fit for a two-factor model, with a primarily positive factor (high loadings for ideas of reference, perceptual experiences, odd behavior and speech, magical thinking, and suspiciousness) and a primarily interpersonal factor (high loadings for social anxiety, flattened affect, and no friends). These factors accounted for 43.5% and 16.1% of the variance, respectively, and correlated .41, $p < .001$. Finally, research

using item-level factor analyses has failed to support Raine's three-factor model (e.g., Chmielewski & Watson, 2008; Cohen, Matthews, Najolia & Brown, 2010); therefore, the factor structure of the SPQ remains unclear.

A third widely used measure of schizotypy is the Oxford-Liverpool Inventory of Feelings & Experiences (O-LIFE; Mason et al., 1995). Due to criticisms of many of the established schizotypy scales, the authors compiled a large battery, involving a combination of seven schizotypy measures, referred to as the Combined Schizotypal Traits Questionnaire (CSTQ; Bentall, Claridge, & Slade, 1989). The CSTQ contains 420 items, which made it impractical. For this reason, Mason et al. (1995) used factor analysis to create the O-LIFE. This measure is comprised of four subscales with no more than 30 items each: unusual experiences, cognitive disorganization, introvertive anhedonia, and impulsive nonconformity. Each scale displayed good internal consistency (.77 to .89; Mason et al., 1995), as well as test-retest reliability (.77 to .93; Burch, Steel, & Hemsley, 1998). Construct validity has been demonstrated through associations with perception and attention, physiological responding, reasoning and learning tasks, hemispheric function, paranormal beliefs, heritability, and other constructs (see Mason & Claridge, 2006).

Limitations of Current Scales

Overall, the construct of schizotypy has been widely explored and there is evidence to support its validity as a spectrum of subclinical and clinical symptoms and impairment. The three measures discussed above, as well as many others, have been

extensively employed in the study of schizotypy and have exhibited varying degrees of psychometric success; however, there are key limitations that warrant the formulation of more current and evolved measures of schizotypy. First, several of the older scales fail to assess schizotypy as a multidimensional construct, for example the Psychoticism Scale (Eysenck & Eysenck, 1975) and Schizotypal Personality Scale (Claridge & Broks, 1984). Secondly those that do assess schizotypy multidimensionally often differ in the number and content of the factors, as described above.

In addition, some of the available measures do not appear to map onto current models of schizotypy. For example, they contain factors that have not received consistent support (e.g., the O-LIFE's impulsive-nonconformity subscale). In some cases, the scales were not designed to measure latent factors and the factor scores were derived in a post hoc fashion. As a result, some measures may not provide appropriate content coverage of the dimensions that they purport to assess. For example, the WSS positive schizotypy factor only assesses odd beliefs and a limited range of unusual perceptual experiences, but does not contain items assessing suspiciousness or paranoia. Likewise, the WSS negative schizotypy factor only assesses anhedonia and social disinterest, and does not contain items tapping alogia, anergia, avolition, or flattened affect. In addition, scales that purport to measure the same factor at times appear to be measuring different constructs. For example, Gross et al. (2014) reported differences in the makeup of the negative schizotypy factors within each measure that limit their comparability. Specifically, the WSS negative factor was only modestly

associated with neuroticism, whereas the SPQ negative factor showed a moderate correlation, the WSS negative factor was associated with decreased openness to experience, whereas the SPQ interpersonal factor was uncorrelated, and the SPQ interpersonal factor showed a large degree of overlap with positive schizotypy (in both scales), whereas the WSS negative factor did not.

As mentioned previously, these findings for the SPQ interpersonal factor are contrary to the theoretical conceptualization of negative or deficit schizotypy. Further, there are problems within the factor structures of each scale. For example, the WSS are limited to assessing positive and negative dimensions, despite ample evidence that schizotypy and schizophrenia are comprised of more than two dimensions (Peralta & Cuesta, 2001; Vollema & van den Bosch, 1995).

The SPQ and the O-LIFE offer the potential advantage of tapping a disorganized schizotypy factor. In the case of the SPQ, however, the large degree of overlap with its cognitive-perceptual factor, as well as all three factors' high correlation with neuroticism, suggest that these factors may not be distinct and may measure constructs other than those of interest (Gross et al., 2014). Further, there is widespread disagreement regarding this 3-factor structure of the SPQ, with studies reporting 2, 3, and 4+ factor models (e.g., Chmielewski & Watson, 2008; Gross et al., 2014; Stefanis et al., 2004). Finally, the 4-factor structure of the O-LIFE is limited in that the cognitive disorganization factor assesses difficulties with attention, concentration, and decision making, but also includes a number of items tapping social anxiety and mood symptoms

that appear independent of schizophrenia-spectrum cognitive dysfunction. Finally, impulsive nonconformity, described by Mason and Claridge (2006) as impulsive, anti-social, and eccentric behavior often indicative of poor self-control, does not appear to tap core features of schizotypy. Rather, it captures impulsive, borderline, and antisocial features (Bouvard & Cosma, 2008) and several studies have reported that impulsive nonconformity is not associated with and/or does not predict schizophrenia (e.g., Chapman, Chapman, & Kwapil, 1994; Cochrane, Petch, & Pickering, 2010). Finally, this factor has also been shown to be unstable across time, (e.g., Lin et al., 2013) and Mason and Claridge (2006) acknowledged that some researchers were using the O-LIFE without it.

Another major limitation is that measures are not built on current multidimensional models of schizotypy, and suffer from an unclear or poorly developed theoretical basis. The SPQ was based entirely on the DSM-III-R diagnostic features of schizotypal personality disorder, therefore it is not entirely comparable to scales built to assess schizotypy more broadly, and lacks an a priori proposed factor structure. Schizotypal personality disorder is subsumed within the schizotypy continuum, but the constructs are not synonymous (Kwapil & Barrantes-Vidal, 2012). The O-LIFE was constructed using factor analysis on items taken from several previously existing scales. Thus, it captured the factor structure inherent in those existing items, but was not built to tap an a priori model of schizotypy.

There are several approaches to test construction, including the external/empirical approach, which emphasizes prediction of an external criterion using an empirical keying procedure, and the internal approach, which uses the internal structure of the items to determine test construction, for example through the use of factor analysis (Hornick, James, & Jones, 1977; for a review of approaches see Goldberg, 1972). A third approach, termed the rational approach, emphasizes the development of items based on theoretically meaningful constructs to develop a measure with strong psychometric properties, followed by a process of construct validation (e.g., Jackson's [1970] rational scale development approach for personality measures). Each approach has merits and criticisms and may be appropriate for different applications; however, the use of factor analysis to develop scales measuring schizotypy (for example in the case of the O-LIFE and WSS) has resulted in scales that work reasonably well but include concepts that are not theoretically part of the schizotypy (e.g., social anxiety, impulsive nonconformity), and that do not fully cover positive, negative, and disorganized content domains. Further, the lack of clear a priori models of schizotypy often appears to result in researchers defining schizotypy (and its multidimensional structure) based on whatever psychometric measure they employed.

To bring clarity to the study and measurement of schizotypy, scales should be built on a clear a priori theory/model of a construct, instead of the current practice of allowing post-hoc factor analysis to define the factor structure of that construct. The latter method has resulted in three scales with different factor structures, and ongoing

ambiguity and disagreement regarding the structure of schizotypy. New scales should be based on a model that precedes item development, and then investigation of the validity of this model through processes of factor analysis and construct validation. We currently have a system that is backwards in that the scales tend to define the construct, rather than a clear model of the construct driving the development of the measures.

In addition to very different factor structures and a lack of clearly defined a priori models, current measures are often not up to date in terms of measurement theory. The current scales were generally developed using CTT, which has served test creators and users well for some time. However, marked limitations with CTT (Hambelton and Swaminathan [1985] describe six major shortcomings of CTT) contributed to the development of IRT (also known as latent trait theory) as a more powerful model for test development. Further, differential item functioning (DIF) is useful for identifying items on which respondents who are equal on the latent trait (in this case, schizotypy) respond differently based on different group membership (e.g., men vs. women). Given previous reports of differences in schizotypy scale scores associated with sex and race/ethnicity (e.g., Chmielewski, Fernandes, Yee, & Miller 1995; Kwapil, Crump, & Pickup, 2002), the creation of unbiased test questions is crucial in examining whether these are “true” trait differences or based on scales with DIF.

Graves and Weinstein (2014) found evidence for unidimensionality and fit to the Rasch model (1-parameter IRT model) for three of the Wisconsin Scales (Perceptual Aberration, Magical Ideation, and Revised Social Anhedonia); however, they had a

relatively small sample size, did not include the Physical Anhedonia Scale, and did not test for DIF. Furthermore, they did not employ a 2-parameter IRT model, which may be more appropriate for assessing schizotypy, as it takes into account both difficulty and discrimination of the items. Winterstein, Silvia, et al. (2011) used a 2-paramter IRT model and DIF to examine the psychometric properties of the four scales comprising the WSS. They found many good items within the four scales but revealed some items with low discrimination values and a number of items with high DIF. Reise, Horan, and Blanchard (2011) used a 2-parameter IRT model to investigate the latent structure of the Social Anhedonia Scale (one of the four Wisconsin Schizotypy Scales) and reported that neither unidimensional nor bifactor IRT models accurately fit the scale. Specifically, they reported multiple small content clusters within the scale, modest relations between those clusters indicating a general factor of only modest strength, and items that shared little variance with the majority of the items. This scale was designed to be a unidimensional measure of social anhedonia (a feature of negative schizotypy), and Reise et al. (2011) reported the internal consistency of the scale is likely inflated by multidimensionality and multiple items that ask the same question in slightly different ways. Thus, scales developed with modern measurement models are likely to afford significant improvements over those developed with CTT.

Little research has been done using modern measurement models to investigate the psychometric properties of the SPQ and O-LIFE, which were not developed using IRT or DIF. Fonseca-Pedrero et al. (2015) conducted IRT in an investigation of the

measurement invariance of the O-LIFE; however, they only reported information functions for the scale and did not report parameter estimates, fit information, or DIF analyses. Fonseca-Pedrero et al. (2014) reported that 11 of the SPQ items showed significant DIF for gender, but did not analyze DIF for other group membership. Earleywine (2006) used the shortened SPQ (SPQ-B) to investigate the difference on schizotypy scores between regular cannabis users (at least once a week) and non-users (tried it but not within the past year). The initial results supported previous findings that higher schizotypy is associated with cannabis use; however, an investigation of the SPQ-B items using DIF revealed that two items were significantly biased, and removal of these items eliminated the significant difference on schizotypy scores between the two groups. This latter study highlights the role that measurement error can play in drawing erroneous conclusions about constructs and the need for updated scales using modern measurement tools.

Another general limitation of existing scales is wording that is at times problematic, outdated, and culturally biased. Items on the SPQ switch between first-person statements ("I prefer to keep to myself") and second-person questions ("Do you believe in telepathy?"). The WSS item "I have noticed sounds on my records that are not there at other times" is most likely no longer relevant to adolescent and young adult participants. Likewise, "The first winter snowfall has often looked pretty to me" is not as applicable to people who do not experience snow. Such items are problematic for

creating a measure with cross-cultural validity and can be eliminated or reworded to be more appropriate.

Overall, current questionnaire measures of multidimensional schizotypy measure different factor structures, there are problems inherent in the factor structure of each measure, and they tend to lack conceptual richness and an a priori model of multidimensionality. Further, current measures do not incorporate information from modern measurement models such as IRT and DIF, thus ignoring tools for identifying potential measurement error. The current lack of conceptual and methodological clarity inhibits the utility of the construct of schizotypy and new measures should address these issues to advance knowledge in this area.

Goals of the Present Study

The goal of this study was the development of the Multidimensional Schizotypy Scale (MSS; Kwapil, Gross, Silvia, Raulin, & Barrantes-Vidal, 2017), a theoretically driven, psychometric measure of schizotypy with three unidimensional subscales. The measure was based on Gross and Kwapil's (2014) model with three conceptualized dimensions of schizotypy and followed Jackson's (1970) rational scale development approach for personality measures, as outlined and expanded by DeVellis (2012). It was expected that the use of CTT would result in subscales with excellent coverage of their respective domains, high internal consistency reliability, and relatively modest intercorrelations with the other subscales. Similarly, it was expected that the use of IRT and DIF would enhance the scales through the selection of items with high discrimination, a target

range of difficulty, and minimal bias related to gender or race/ethnicity. Based on established gender differences in schizotypy (Hafner, 2003; Kulkarni et al., 2012; Kwapil, Crump, & Pickup, 2002; Plocka & Rybakowski, 1992), including earlier onset of schizophrenia, more severe negative symptoms clinically and sub-clinically, worse response to antipsychotics, and more pronounced premorbid social impairment for men, it was expected that male participants would show higher scores on the negative schizotypy subscale of the MSS. Studies have also suggested higher schizotypy scores in racial/ethnic minority groups (e.g., Chmielewski, Fernandes, Yee, & Miller 1995; Kwapil, Crump, & Pickup, 2002); however, it is unclear if item bias impacted these findings. Therefore, predictions were not made regarding racial/ethnic differences.

CHAPTER II

METHODS

Participants

A total of 8,750 participants was recruited from five sources during a series of twelve administrations over a two-year period. Specifically, items were administered to participants at four universities from Spring 2015 to Fall 2016: University of North Carolina at Greensboro (UNCG), University of Illinois at Urbana-Champaign (UIUC), Tennessee Technological University (TTU), and Youngstown State University (YSU), including seven administrations for a total of 4,863 participants. In addition, five administrations were collected using Amazon Mechanical Turk (MTurk; 3,887 participants) from May to November 2016. See Table 1 for the timeline, universities, number of subjects assessed, number dropped, and demographic characteristics at each administration.

As MTurk has gained popularity for obtaining large samples in remarkably little time, much research has examined the quality of data obtained through this method. Although some studies have suggested that MTurk responders are less attentive (e.g., Goodman, Cryder, & Cheema, 2013, Study 2) these same authors reported several strengths of MTurk and concluded it is a valuable opportunity for data collection when used with attention checks. Further, ample studies have reported that MTurk

participants engage in problematic responding behaviors at comparable (Goodman, et al., 2013, Study 1; Necka et al., 2016; Paolacci, Chandler, & Ipeirotis, 2010) or lesser (Hauser & Schwarz, 2016; Klein et al., 2014) rates, compared to community or campus samples.

A total of 1,485 participants was eliminated from the analyses based on the following criteria: overall completion of less than half of the items (n = 362), invalid protocols based on infrequency measures (n = 947), and age 60 or above (n = 176). The maximum age criterion was driven by the following factors: a) the study of schizotypy typically targets participants near the age of onset for developing schizophrenia-spectrum disorders, b) we aimed to avoid cognitive decline related to age as opposed to disorganized schizotypy, and c) participants age 60 or older comprised only two percent of the total sample, therefore we lacked adequate representation of that age group. The application of these criteria resulted in a total of 7,265 usable participants (4,004 college students and 3,261 from MTurk).

Five hundred female and 500 male participants were randomly selected from the data from administrations six through twelve (the four Fall 2016 university assessments and the three fall 2016 MTurk assessments) to comprise a separate cross-validation sample. The decision to select this sample from the last seven administrations was made in order to use only participants for whom the data had never been examined and for which item selection decisions had not been made. Further, these administrations all contained the same set of items (survey 3) from which the final scales were selected. To

create a proportionate sample, SPSS was used to randomly draw 35.3% of the men (500/1416) and 18.5% of the women (500/2703) from each of the seven samples. Thus, the composition of the cross-validation sample perfectly matches the proportion of participants from the final seven assessments. The derivation sample included the remaining 6,265 participants. Note that the cross-validation sample was not included in any analyses for selecting the final items, but was strictly retained for assessing the psychometric properties of the new scale. As seen in Table 2, the derivation and cross-validation samples were comparable in age, race/ethnicity, and English as first language. The derivation sample contained 67% female participants, as compared to 50% for the cross-validation sample, as the latter was sampled to have an equal number of men and women.

Regarding a comparison of MTurk and college participations in the total pool of subjects, 2.5% of MTurk participants and 5.4% of college students were dropped based on missing data, 9.0% of MTurk participants and 12.3% of college students were dropped due to invalid protocols based on infrequency measures, and 4.5% of MTurk participants and none of the college students were dropped due to age (<59). Before dropping participants, MTurk participants tended to be older (M = 36.27 vs. 19.66; t[7435]=84.84, p<.001) and had a higher proportion of males (37.8% vs. 30.7%; X^2[1]=41.23, p <.001). MTurk participants were more likely to be White (78.9% vs. 64.3%; p < .001, Fisher's exact test) and college students had a higher proportion of Black/African American (AA) participants (15.9% vs. 7.2%; p < .001, Fisher's exact test).

Materials

Schizotypy items: Trait specification and item generation. Development of the scale began with review of existing schizotypy scales and preparation of detailed trait specifications describing positive, negative, and disorganized schizotypy. These descriptions were used to guide the creation of large pools of items. These primarily included new items, as well as items from other scales in original or modified form. All candidate items had true-false response options. Note that some items were reverse scored. Items were scored 1 for an answer endorsing schizotypy and 0 for an answer in the non-schizotypic direction. Dichotomous rather than polytomous items were used because schizotypic experiences are presumed to be relatively rare and Likert scales can allow for normalizing of non-schizotypic experiences. Further, people experiencing psychopathology may have more difficulty with polytomous ratings.

The items were reviewed for content and grammar by eight expert and six non-expert reviewers. The first two administrations included 81 positive, 79 negative, and 86 disorganized schizotypy items. The item pool was reduced to 53 positive, 53 negative, and 49 disorganized schizotypy items for the third administration. The final nine administrations contained 42 positive, 39 negative, and 37 disorganized schizotypy items. For the sake of space, only the 118 items from survey 3 were included in this document: see Tables 3 through 5 for the items included in the final scales and Table 6 for the items from survey 3 excluded from the final scale.

Other measures. In addition to the schizotypy items, participants in all administrations completed the Infrequency Scale (Chapman & Chapman, 1983). The Infrequency Scale is a 13-item (true/false) measure of infrequent responding used to screen out participants who respond in a random or "fake bad" manner. Similarly, all administrations included the Attentive Responding Scale (ARS; Maniaci & Rogge, 2014), which contains two subscales. The infrequency subscale includes six highly unlikely items (e.g., "I enjoy the music of Marlene Sandersfield" [a fictional performer]) and the inconsistency subscales consists of six item pairs with nearly identical content (e.g., "I enjoy relaxing in my free time" and "In my time off I like to relax"). The Infrequency Scale and the six ARS infrequency items were randomly interspersed throughout the survey. The ARS inconsistency item pairs were presented in opposing halves of the survey and absolute differences were summed across item pairs so that higher scores reflect more inconsistent responding. Maniaci and Rogge reported that the ARS effectively identified the majority of respondents with problematic data and yielded consistently higher power for detecting effects.

Due to concerns about previous schizotypy scales (especially subscales purporting to measure negative schizotypy) tapping high levels of neuroticism, the neuroticism subscale of the NEO Five-Factor Inventory, 3rd edition (NEO-FFI-3; McCrae & Costa, 2010) was included in all item administrations. The NEO-FFI-3 is a gold standard measure of the 5-factor model of personality and the neuroticism subscale contains 12 items assessing the tendency to experience negative affect (such as fear, sadness,

embarrassment, anger, guilt, and disgust). The items contain a 5-point Likert scale (strongly disagree, disagree, neither agree nor disagree, agree, strongly agree). Given that all other items contained binary (true/false) responses, the neuroticism items were administered together at the end of the survey. Finally, participants in the first three administrations completed the Social Desirability Scale (Crowne & Marlowe, 1960), a 33-item (true-false) measure of a respondent's tendency to present him/herself favorably. The Social Desirability Scale was dropped from the battery after the third administration (n = 2174) because none of the retained schizotypy items had significant positive correlations with social desirability.

Procedures

All participants completed the survey online using Qualtrics software. The project received IRB approval at each of the four institutions. Students were recruited electronically at their respective universities and received course credit for participation. MTurk participants were recruited via the MTurk website and were limited to people in the United States. MTurk participants received $1.00 for taking part in the study. The survey began with an image of the informed consent form followed by a question verifying participants age (at least 18 years old) and asking for consent to participate. Next, demographic questions (gender, age, race/ethnicity, and English as a second language) were included, followed by this prompt: "The present study inquires about a variety of beliefs, attitudes, and experiences that some adults report having. Please indicate whether each item is true or false about you. Note that there are no right or

wrong answers, just answer in the way that best describes what you are like. All of your answers are anonymous and confidential." The schizotypy, infrequency, ARS, and social desirability items were intermixed manually and divided into five blocks. These blocks were presented in random order followed by a sixth block consisting of the neuroticism items. Participants received a prompt when failing to answer a question but could continue whether or not they provided an answer.

Data analysis and item evaluation. CTT, IRT, and DIF statistics were used for the item selection process, as well as selecting items that provided adequate coverage of the three schizotypy domains. CTT statistics, generated using IBM SPSS Version 20 (IBM Corp., 2011), included mean endorsement frequency, correlation of the item with its schizotypy dimension, correlations with the other two schizotypy dimensions, and correlations with neuroticism.

Two-parameter logistic (2PL) IRT models, generated using IRTPRO Version 3 (Scientific Software International Inc., 2015) produced discrimination and difficulty parameters, as well as item response curves. Multidimensional IRT was not conducted, rather each subscale was examined independently. IRTPRO uses maximum likelihood estimation for item parameter estimation. IRTPRO parameter estimates for all models are always in the logistic metric. To be rendered comparable to normal ogive discrimination parameters, the IRTPRO estimates of the a parameters could be divided by 1.7 (Scientific Software International, Inc., 2015). IRT models the relationship between the latent trait (theta) level and the probability of a "correct" response and this

relationship is represented graphically by the item characteristic curves. Item difficulty (b parameter) represents the point on the ability scale at which the examinee has a fifty percent probability of answering a question correctly. The concept of correctness follows from IRT's basis in ability testing; however, in this case the underlying trait is schizotypy. Therefore, difficulty can be better understood as deviance; a "harder" item represents a less common personality characteristic, higher on the schizotypy continuum. Further, the b parameter represents the level of theta at which there is a fifty percent change the item will be endorsed "yes" in the deviant (schizotypal) direction. The discrimination (a) parameter is proportional to the slope at p = .50 (the point of inflection of the item characteristic curve). Items with higher discrimination values will better differentiate between people low and high on schizotypy.

Alternative IRT models to the 2PL include 1 parameter logistic (1PL) and 3 parameter logistic (3PL) models. The 1PL constrains the slope (discrimination) parameter to be equal across items, and estimates this common slope based on the item pool. For both 1PL and 2PL models, the c parameter, or the lower asymptote of the item characteristic curve is set to zero. In the 3PL model, this parameter is allowed to vary. In traditional ability testing, this parameter may represent low ability participants who answer items correctly due to chance guessing. For our purposes, this parameter models participants lower on trait schizotypy endorsing medium-difficult items in the deviant direction. To our knowledge, no one has used the 3PL model for measures of schizotypy. Further, 2PL models are most common in personality assessment, as

researchers have argued a guessing parameter is not relevant (e.g., Reise & Waller, 1990). However, others have suggested that guessing on cognitive ability items may be analogous to "fake bad" answers for personality measures and that assessing model fit for 3PL models in personality testing is needed (Chernyshenko, Stark, Chan, Drasgow, & Williams, 2001; Rouse, Finger, & Butcher, 1999). Therefore, to test for the appropriateness of the 2PL model, all three models were compared on three indices of model fit: -2loglikelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). Smaller values on each indicate better model-data fit.

Differential item functioning refers to bias, and DIF analyses aim to detect items or tests that differentially favor test takers based on their group membership (e.g., gender, race/ethnicity), at the same ability level. DIF is present when people at the same level of the underlying trait (schizotypy) have different probabilities of endorsing an item in the deviant direction. Bias is extremely problematic because different test scores do not accurately represent differences in the underlying trait. Many methods can be used to detect DIF, and McNamara & Roever (2006) classified these methods into four categories: 1) analyses based on item difficulty (e.g., transformed item difficulty index or delta plot, 2) nonparametric methods (contingency tables and chi-square methods), 3) item response theory (1, 2, and 3 PL models), and 4) other approaches (e.g., Rasch measurement, generalizability theory).

In this study, DIF was run using IRT (IRTPRO), which offers the advantage of matching participants on the examinee's estimated ability level, as opposed to on the

observed score (as in other methods; Karami, 2012). DIF through IRT examines differences in the item information curves for the reference and focal group through comparison of the item parameters in the two groups. This approach also offers the advantage of testing differential functioning for the discrimination parameter, as opposed to other approaches that assume discrimination is equal across groups (Karami, 2012). Item parameters were compared separately for each subscale across two variables using the 2PL model: sex (male vs. female), and race/ethnicity (White vs. not White and White vs. Black/AA). The race/ethnicity variable was dichotomized in these ways due to low percentages of participants who self-identified as Hispanic/Latino (6% derivation; 6% validation), Asian/Pacific Islander (7%; 9%), and Native American (1%; <1%).

To detect DIF, IRTPRO uses Wald tests and produces chi-square statistics indicating significant differences across groups. Due to the sensitivity of chi-square to sample size, IRTPRO often yields statistically significant DIF that is of little clinical/practical importance (Edelen, Stucky, & Chandra, 2013). Due to our large samples, a *p* value of .001 was used. In the 2PL model, IRTPRO produces three chi-square values: overall DIF for the item, DIF for item discrimination, and DIF for item difficulty. Item parameters are also generated for each of the groups separately; therefore, following a significant chi-square, the researcher can examine the parameter estimates for each of the two groups.

A significant difference for the b parameters across groups suggests uniform DIF, or that the item is systematically (i.e., at all levels of theta) more difficult for members of one group than the other. A significant difference for the a (and possibly also the b) parameter across groups suggests non-uniform DIF, or that the shift in difficulty is not consistent across levels of theta (i.e., across high and low ability participants). For the sake of space, only the chi square values for the b parameter were included in this document, although values for both a and b parameters were examined when selecting items. Differential functioning of the test as a whole (or each of the three subscales in this case) can be examined in a similar fashion, by comparing the test characteristic curves across groups. To our knowledge IRTPRO does not include a significance test for this comparison; however, the test characteristic curves were generated separately for each group for all DIF analyses and visually examined.

Retention and final selection of items was based on the following factors. First, efforts were made to promote content validity by generating and retaining items that covered the full range of the constructs described in the trait specifications. In terms of CTT, preference was given to items that had low endorsement frequency (.05 to .35), high item-scale correlation with the items for that dimension, and relatively lower correlations with the other two schizotypy dimensions. Preference was given to negative schizotypy items with low correlations with neuroticism, and positive and disorganized schizotypy items with low to medium correlations with neuroticism. Items with low endorsement frequency were selected given the relative rarity of schizotypic

experiences in the general population and to maximize discrimination at the high end of the scale. In terms of IRT values, preference was given to items with high discrimination. Consistent with preference for low endorsement items, it was expected that items would have difficulty values of approximately 0.5 to 2.5. Items with markedly elevated DIF for sex or ethnicity were eliminated.

After selection of the final items, exploratory factor analysis (EFA) in Mplus Version 7 (Muthén & Muthén, 2010) was used to examine the dimensionality of each of the three subscales. Items were treated as categorical and parameters were estimated using unweighted least squares solutions. Unweighted least squares estimation was used because it is a widely used estimation procedure that does not require assumptions of multivariate normal distributions (Krinjen, 1994). Analyses were conducted separately for the three sets of items (positive, negative, and disorganized schizotypy). It was expected the resulting factors would be correlated, as all three fall under the conceptual umbrella of schizotypy. Therefore, oblique rotation was employed since orthogonal rotation forces factors to be perfectly uncorrelated. Correlated factors are likely a more accurate representation of reality (in this and most practical situations) making oblique rotation more appropriate (Browne, 2001). Yate's (1987) geomin is an oblique rotation supported for use in dichotomous data (Finch, 2011) and Mplus employs geomin as the default rotation. Mplus was also used to run a series of CFAs using maximum likelihood with robust standard errors. A series of one- (single schizotypy factor), two- (negative factor and combined positive and disorganized factor),

and three-factor (positive, negative, and disorganized factors) item-level models was run.

CHAPTER III

RESULTS

Item-Level Results

A total of 26 positive, 26 negative, and 25 disorganized schizotypy items were retained for the final subscales from the pool of 119 schizotypy items in survey 3. Tables 3 through 5 display the items for the final subscales and Table 6 shows the items eliminated from survey 3. Note in the tables that the number of subjects completing each item varied for two possible reasons. First, participants could choose not to complete an item when they took the scale. Second, twelve items were added after survey one and three items were added after survey two.

Of the final subscale items, POS24 – POS26, NEG24 – NEG26, and DIS23 – DIS25 were added after survey one, therefore they have fewer participants (none of the three items added after survey two was retained). Ten of the items were taken directly from other scales (five from the Magical Ideation, three from the Perceptual Aberration, and one each from the Revised Social Anhedonia and Physical Anhedonia Scales) and eight items were modified from other scales (three each from the Revised Social Anhedonia and Cognitive Slippage [Miers & Raulin, 1987] Scales, and two from the Schizotypal Personality Questionnaire). Based on five indices, the average reading grade level of the items was 8.2 (Readable.io).

Item-level CTT. Tables 3 through 5 present the CTT statistics for the positive, negative, and disorganized subscales (final items) for the derivation sample, and Tables 7 through 9 present the CTT statistics for positive, negative, and disorganized subscales (final items) in the cross-validation sample. Table 6 displays the CTT information for survey 3 items eliminated from the final scales. These statistics include mean item endorsement and point-biserial correlations with positive, negative, and disorganized schizotypy and neuroticism total scores. The range of endorsements for positive schizotypy was .06-.37 (M=.14, SD=.08) for derivation and .06-.37 (M=.14, SD=.08) for cross-validation, range of endorsements for negative schizotypy was .05-.33 (M=.14, SD=.07) for derivation and .05-.34 (M=.15, SD=.06) for cross-validation, and range of endorsements for disorganized was .09-.25 (M=.16, SD=.05) for derivation and .08-.25 (M=.16, SD=.05) for cross-validation. The range of point-biserial correlations for positive schizotypy was .43-.60 (M=.50, SD=.04) for derivation and .44-.61 (M=.51, SD=.04) for cross-validation, range for negative schizotypy was .34-.63 (M=.49, SD=.07) for derivation and .33-.68 (M=.51, SD=.09) for cross-validation, and range for disorganized schizotypy was .54-.71 (M=.64, SD=.05) for derivation and .53-.75 (M=.64, SD=.06) for cross-validation.

IRT Model Comparison. Table 10 shows the fit statistics for the 1PL, 2PL, and 3PL models in the derivation and cross-validation samples for each of the three subscales. As unidimensionality is a key assumption of IRT, each subscale was analyzed separately. In the derivation sample, neither the 1PL nor 3PL model showed better fit on any of the

three fit indices than the 2PL model. In the cross-validation sample, the 3PL did not show better fit than the 2PL model on any of the three fit indices. The BIC was slightly lower for each of the three subscales for the 1PL model as compared to the 2PL model (positive BIC difference = 100.91; negative BIC difference = 0.92; disorganized BIC difference = 28.10).

Item-level IRT. Tables 11 through 13 show the IRT parameters for the three subscales run in the derivation sample (including all survey 3 items) and Tables 14 through 16 contain the parameters for the subscales in the cross-validation data (final items only). Table 17 displays the IRT parameters for the items eliminated from survey 3.

The range (mean, SD) of the discrimination (a) values for positive schizotypy was 1.35-2.26 (M=1.80, SD=.25) for derivation and 1.20-2.43 (M=1.88, SD=.31 for cross validation, for negative schizotypy was 1.12-2.59 (M=1.71, SD=.40) for derivation and 1.15-3.22 (M=1.78, SD=.53) for cross-validation, and for disorganized schizotypy was 1.85-3.27 (M=2.50, SD=.42) for derivation and 1.82-3.95 (M=2.50, SD=.56) for cross validation. The range (mean, SD) of the difficulty (b) values for positive schizotypy was .44-2.07 (M=1.56, SD=.43) for derivation and .66-2.05 (M=1.54, SD=.41) for cross-validation, for negative schizotypy was 1.10-2.65 (M=1.67, SD=.47) for derivation and .61-3.44 (M=1.63, SD=.58) for cross validation, and for disorganized schizotypy was .82-1.74 (M=1.22, SD=.27) for derivation and .83-1.72 (M = 1.29, SD = .28) for cross-validation. Item-level fit statistics were run in the cross-validation sample (see Table 18).

IRTPRO computes the trace line diagnostic statistic S - X^2 suggested by Orlando and Thissen (2000). None of the positive, negative, or disorganized items showed misfit at the $p < .001$ level suggesting that for all items the trace lines have been fitted sufficiently well. This indicates that the model-expected proportions responding 0 and 1 match the observed data. Because these statistics assume perfect fit to the 2PL model, these results suggest excellent item-fit for the final subscale items.

Item-level DIF. Tables 11 through 13 show the chi-square values for DIF for the final items run in the derivation sample (including all survey 3 items) and Table 17 displays the DIF results for the items eliminated from survey 3. These chi-square values represent differences on the difficulty parameter for each item driven by group membership as opposed to schizotypy. Once a significant chi-square value has been identified, comparison of the individual b value for each group reveals the direction of the difficulty (i.e., for which group the item was harder and easier). Note that chi-square is not an effect size and cannot be interpreted in terms of small, medium, and large effects. Furthermore, the large sample sizes (especially in the derivation sample) make interpreting statistical significance of chi-square for evaluation purposes problematic. Nevertheless, several items were eliminated based on relatively "large" chi-square statistics. For example, the negative item "My emotions are much less intense than other people's" was significantly easier for (i.e., more likely to be endorsed by) male participants than female participants, regardless of the trait level (X^2=108.1, p<.001). Therefore, the item was eliminated as endorsements are not purely reflective of

schizotypy. Some items with significant chi-square values were kept in the subscales based on the sensitivity of chi-square to sample size, the content validity of the item (i.e., the necessity of the item to tap content coverage theoretically comprising that schizotypy dimension), and the item's other psychometric properties.

Tables 14 through 16 contain the DIF results run in the cross-validation sample for the final subscale items. None of the positive items showed significant DIF for gender, White vs. not White, or White vs. Black/AA. For negative schizotypy, two items showed significant DIF for sex ("I rarely feel strong emotions even in situations in which other people usually do," and "My emotions have almost always seemed flat regardless of what is going on around me"). Both items were easier for male participants. One item showed significant DIF for White vs. not White ("When I move to a new place, I feel a strong desire to make friends," easier for White) and one item showed significant DIF for White vs. Black/AA ("I enjoy meeting new people and making new friends," easier for White). Finally, none of the disorganized items showed significant DIF.

Subscale Results

Descriptive statistics. Table 19 displays the descriptive statistics in the derivation and cross-validation samples. As expected, the total score distributions are positively skewed, reflecting the assumption that schizotypic experiences are relatively rare. Note that the descriptive statistics were closely comparable in the two samples and the reliability of the scales exhibited little to no shrinkage in the cross-validation sample. The internal consistency coefficients (Cronbach's alpha) ranged from .88 to .94.

Cronbach's alpha presumes continuous variables and is biased downward both for binary items and when several items have low endorsement rates (Bandalos & Enders, 1996; Lissitz & Green, 1975; Liu, Wu, & Zumbo, 2010).

One method of correction is to estimate alpha from a CFA using binary indicators (see Drewes, 2000; Hancock & Mueller, 2001). A tau-equivalent CFA model (binary indicators specified as categorical, paths constrained to be equal, factor variance fixed to one) was run in Mplus and alpha was estimated by plugging the squared factor loading (average correlation between all items) and the number of items into the computational formula for alpha. All three subscales demonstrated good to excellent internal consistency reliability using either Cronbach's or binary alpha in both samples.

Table 20 shows the intercorrelations of the subscales, as well as correlations with neuroticism. Note that given the large sample size, alpha was set at .001 and results were considered in terms of effect sizes. Intercorrelations were consistent across samples with small associations between positive and negative, medium to large associations between positive and disorganized, and medium associations between negative and disorganized schizotypy subscales. Neuroticism had a small to medium association with negative schizotypy, a medium association with positive schizotypy and a large association with disorganized schizotypy. Partial correlations of each of the three schizotypy dimensions with neuroticism with the other two schizotypy dimensions partialed out of the analysis were computed in the cross-validation sample. The partial

correlation of neuroticism with positive schizotypy was .12, $p < .001$, with negative schizotypy was .08, not significant, and with disorganized schizotypy was .43, $p < .001$.

As expected, men had higher scores than women on negative schizotypy (e.g., Tandon, Nasrallah, & Keshavan, 2009), but men and women did not differ on positive or disorganized schizotypy in either sample. ANOVAs were computed comparing the racial/ethnic groups on scores on the three subscales. None of the analyses was significant: positive schizotypy, $F(5,994) = 1.50$; negative schizotypy, $F(5,994) = 2.62$; disorganized schizotypy, $F(5,994) = 1.26$.

IRT item fit and information curves. All the final items showed good fit to the 2PL model. Information and standard error functions provide insight into the range of the trait level at which each subscale provides the most information. Figure 1 shows the information and standard error curves for the final positive, negative, and disorganized subscales in the cross-validation dataset. The test information maxima (trait level where scale provides the highest information and lowest standard error) are 1.6 for positive, 1.6 for negative, and 1.2 for disorganized schizotypy. Consistent with the item endorsements and difficulty parameters, as well as the idea that schizotypic symptoms are relatively rare, all three subscales provide the greatest information at high ends of the trait continuum, with disorganized being slightly lower than both positive and negative.

Differential Test Functioning. Figures 2 through 4 display the item characteristic curves generated separately by group for each subscale in each of the three sets of DIF

analyses. Although IRTPRO does not provide a significance/magnitude test of overall differential test functioning, the curves appear comparable across groups for all analyses without any notable differences.

Factor analysis. The EFA eigenvalues are presented in Table 21 and item loadings were as follows for the derivation (positive: .55-.78, negative: .50-.83, disorganized: .71-.88) and cross-validation (positive: .59-.79, negative: .41-.85, disorganized: .71-.91) samples. Numerous rules/indices are used to determine the appropriate number of factors when assessing dimensionality and currently there is no agreed upon decision-making method. Slocum-Gori and Zumbo (2011) summarized and evaluated nine commonly used criteria, three of which are based on eigenvalues: 1) the eigenvalues-greater-than-one rule (also known as the Kaiser criterion; Kaiser, 1970), 2) the ratio-of-first-to-second-eigenvalues-greater-than-three rule, and 3) the ratio-of-first-to-second-eigenvalues-greater-than-four rule.

Slocum-Gori and Zumbo (2011) reported that for large (800 participants), skewed (skew = 2.50) samples with high communality (i.e., conditions similar to the present study), the eigenvalue rules performed well. Others (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999) have reported significant problems with the Kaiser criterion (including overestimation of factors) despite widespread use, and Hattie (1985) found that it overestimated the number of factors for unidimensional cases. As seen in Table 21, use of the Kaiser criterion suggests four, four, and two factors for the positive,

negative, and disorganized subscales, respectively. Use of both the ratio > 3 and ratio > 4 rules suggests one dominant factor underlying each of the subscales.

A series of one- (single schizotypy factor), two- (negative factor and combined positive and disorganized factor), and three-factor (positive, negative, and disorganized factors) item-level CFAs showed best fit for a three-factor model in both sample (see Table 22 for fit statistics).

CHAPTER IV

DISCUSSION

Schizotypy provides a useful construct for the conceptualization of both subclinical and clinical manifestations of the schizophrenia spectrum and for identification of relevant etiological factors. In his landmark 1962 address that introduced the construct of schizotypy, Meehl stated the "most important research need is development of high-validity indicators for compensated schizotypy" (p. 830). Thirty-three years later the Chapmans concluded, "attempts to use questionnaires to measure proneness to schizophrenia or to psychosis have achieved a measure of success that encourages vigorous pursuit of the best possible measure or set of measures" (Chapman et al., 1995, p. 101). Kwapil and Barrantes-Vidal (2015) stated "psychometric assessment provides a promising point of entry for assessing schizotypy" and advocated for the development of new measures built on current conceptual models.

Questionnaire measures of schizotypy have provided useful tools for assessing large numbers of non-clinical participants, as well as clinical patients. However, the assessment of schizotypy seems to be at an impasse as currently available measures suffer from limitations including lack of a clear conceptual framework, outdated wording, unclear factor structure, and psychometric shortcomings. Specifically, many of

these measures do not map onto current multidimensional conceptualizations of schizotypy. Scales that assess schizotypy multidimensionally often differ in the number and the content of the factors, and in some cases, scales that purport to measure the same factor appear to be measuring different constructs altogether. Many of the available measures suffer from psychometric limitations – in part because they were developed prior to the advent of modern measurement tools. Evaluation of these scales with IRT and DIF reveals some items that suffer from low discrimination and high differential item functioning for sex and ethnicity (e.g., Winterstein et al., 2011). In addition, some of the available scales employ wording that appears outdated or culturally biased.

The goal of the current study was to employ CTT, IRT, and DIF to produce a new measure of schizotypy that both builds on the success of existing measures and advances beyond their limitations. The scale developmental procedures for the MSS offered several strengths, consistent with the best practices described by DeVellis (2012). These included using comprehensive trait specifications to develop large pools of items, administration to large, diverse, multi-site samples, and employment of innovative measurement tools.

Strengths of the MSS Development

Theoretical basis. The heterogeneity of schizotypy is widely recognized, but it sometimes appears that researchers base their model of schizotypy on the scale that they select, rather than selecting or building scales based on an a priori model.

Consistent with the conceptualization of schizotypy as encompassing schizophrenia and related disorders, it follows that their underlying multidimensional structure should be similar; however, none of the currently available schizotypy measures were specifically designed a priori to capture this three-factor structure. The SPQ has a debated three-factor structure and the O-LIFE encompasses positive, negative, and disorganized factors; however, both structures were the result of post hoc factor analyses (including some determined using modification indices) and offer factors that do not necessarily map onto their factor names.

A related criticism is that some measures fail to assess key content areas of schizotypy. For example, the WSS were based on a priori theory; however, the scales only assess positive and negative schizotypy. Furthermore, the WSS dimensions were derived years after the scales were created – and these factor-analytically derived dimensions fail to provide full content coverage of positive and negative schizotypy. Specifically, the WSS positive dimension is limited to assessing body-image aberrations and odd beliefs, whereas the negative dimension (like the O-LIFE introvertive anhedonia factor) measures physical and social anhedonia.

The MSS schizotypy subscales were designed to draw from more broadly defined content areas and offer more complete coverage of the constructs of positive, negative, and disorganized schizotypy. This was accomplished by using trait specifications to guide item development and retention so that they tapped content domains comprising the three dimensions. The positive domain included dreams, supernatural, special powers,

passivity, thought transmission, ideas of reference, paranoia, and perceptual aberrations. The negative domain included flat affect, alogia, avolition, anhedonia, and social anhedonia. The disorganized domain included cognitive slowing, racing/loose associations, confusion, difficulty understanding and producing speech, disorganized thoughts, and disorganized behavior. The inclusion of suspiciousness and passivity experiences (key features of positive schizotypy) and avolition, alogia, and flattened affect (key features of negative schizotypy) represent content areas missed by previous measures such as the WSS.

Perhaps the most challenging construct to assess is cognitive-behavioral disorganization. First, disorganization may be due to other transient or enduring conditions (e.g., ADHD, depression, substance abuse). Second, the experience of disorganized schizotypy (especially pronounced disorganization) may impair one's ability to recognize and report these experiences. Finally, it is not clear that currently available measures of disorganized schizotypy tap deterioration or disruption of thought and behavior that characterize cognitive and behavioral disorganization. As Gross et al. (2014) noted, the SPQ disorganized factor seems to tap "oddness" that may be "due to volitional behaviors resulting from positive symptoms, rather than cognitive and behavioral disorganization" (p. 404). Likewise, a review of the O-LIFE cognitive disorganization items suggests that the items primarily tap social anxiety and low self-esteem. The MSS, like the Cognitive Slippage Scale, targets disruptions in thought,

organization, and communication, rather than perceptions that others view the respondent as odd or eccentric.

Sample characteristics. The study of schizotypy has traditionally used college samples, as they are convenient and within the window of typical age of onset of the expression of schizotypy and the development of schizophrenia; however, they may have characteristics (and protective factors) that limit the generalizability of findings to other young adults their age and other adults more broadly. Furthermore, many scales were developed with participants at only one institution. This study represents an improvement by including data from four universities and from MTurk in an attempt to introduce a broader range of participants into the sample. Consistent with some previous work reporting MTurk participants as more attentive than other samples (e.g., Hauser & Schwarz, 2016), MTurk participants were less likely to be eliminated based on missing data and invalid responding. MTurk participants tended to be older and contained a higher proportion of males than the college students.

Previous research has reported differences between MTurk and college samples, for example, MTurk users may be more likely to be unemployed- or underemployed, more likely to be living at home with parents, and less likely to be married. Further, MTurk participants may be higher in personality variables including neuroticism and introversion (for a review of MTurk vs. non-MTurk samples see Miller, Crowe, Weiss, Maples-Keller, & Lynam, 2017). In the cross-validation sample, MTurk participants showed higher negative schizotypy scores, with no differences for positive and

disorganized. This makes sense given our conceptualization of negative schizotypes as higher in introversion and less likely to seek out social interaction. Characteristics that may limit the generalizability of this sample should be considered, but overall, extending the scale development sample beyond college participants should increase the utility of the MSS for broader populations.

Strengths of the MSS Final Scale

Length, reliability, and item-scale correlations. The 77-item MSS is comparable in length to the SPQ (72 items) and shorter than the O-LIFE (104 items) and the WSS (166 items). The psychometric properties of the MSS subscales were closely comparable in the derivation and validation samples. The subscale internal consistency coefficients were good to excellent in both samples. Furthermore, the reliabilities were comparable or superior to other leading schizotypy measures (e.g., Chapman et al., 1982; Mason et al., 1995; Raine, 1991). The item-total scale correlations (correlation of items with their respective subscale total scores) were consistently large across both samples and slightly higher for the disorganized subscale.

The disorganized schizotypy items tended to have small to medium correlations with the positive and negative schizotypy subscales. This likely reflects the fact that disruptions in cognition and functional impairment are part of positive and negative schizotypy. For example, positive schizotypy involves disruptions in the form of thought. Furthermore, positive symptoms such as odd beliefs, unusual perceptual experiences, and suspiciousness may disrupt thought and behavior. Negative schizotypy involves

diminution of thought and emotion, which again is likely to disrupt the processing of information. Furthermore, negative schizotypy involves diminution of interest in the world and others, which may reduce communication and contribute to further oddity in behavior.

Expected pattern of subscale intercorrelations. Given that the three subscales are conceptualized to be part of the broader construct of schizotypy, we expected them to be correlated, although still distinct. The positive and negative MSS subscales were minimally correlated, and the disorganized subscale showed a medium association with both. This pattern of correlations is consistent with our multidimensional model of schizotypy. Conceptually, positive schizotypy is associated with behavioral and affective excesses whereas negative schizotypy is characterized by diminished affect and avolition (Kwapil et al., 2008). This minimal correlation between positive and negative is consistent with the findings from the WSS (Kwapil et al., 2008) and the O-LIFE (Mason and Claridge, 2006). As described in Gross et al. (2014) the large correlation between the SPQ cognitive-perceptual and interpersonal factors is inconsistent with our multidimensional model of schizotypy, likely driven by the lack of a "true" negative (deficit) factor in the SPQ. As discussed above, the medium associations of the disorganized subscale with the other two subscales are likely because disorganization measures thought disorder and functional impairment, both of which are inherent in positive and negative schizotypy. Overall, the three subscales should not be orthogonal,

given they all fall under the "umbrella" of schizotypy, and the MSS subscales show a pattern of intercorrelations consistent with our theoretical model.

Association of MSS subscales with neuroticism. The MSS subscales showed expected patterns of associations with neuroticism. In both samples, positive and negative schizotypy showed medium correlations with neuroticism and disorganized schizotypy showed a large association. The partial correlations (correlation of each subscale with neuroticism with the other two MSS subscales partialed out) revealed a small significant correlation for positive, nonsignificant correlation with negative, and medium to large association with disorganized schizotypy. The unique association of positive, but not negative, schizotypy with neuroticism is consistent with both our multidimensional conceptualization of schizotypy and with findings from the Wisconsin Schizotypy Scales; further, it marks an improvement over the SPQ's negative factor that is highly correlated with neuroticism. Positive schizotypy is often associated with affective dysregulation (i.e., depression, anxiety, mania) whereas negative schizotypy is characterized by decreased experience of strong affect (Kwapil et al., 2008). Finally, the association with disorganization and neuroticism is consistent with findings for the SPQ cognitive disorganization factor, although the MSS offers the advantage of eliminating items directing assessing anxiety, which are included in the O-LIFE.

Benefits of scale development with IRT. To our knowledge, the MSS is the first measure of schizotypy designed using IRT methods. IRT is a model-based measurement method that offers several advantages over CTT, including less dependence on the

sample of respondents, the provision of an estimate of the probability an examinee will answer a question correctly (rather than the CTT focus on total scale scores), and the ability to estimate levels of information and error at different points on the trait level. Neither the 3PL nor the 1PL model showed better fit than the 2PL model in the derivation data. In the cross-validation data, the 3PL did not show better fit than the 2PL on any indices and on one of three fit indices (BIC), the 1PL showed slightly lower values than 2PL. These results support the use of the 2PL model in generating item parameters. The final items showed excellent item fit in the cross-validation sample. All three subscales are maximally discriminating at the high end of the traits, with disorganized being slightly lower than the positive and negative subscales. This is consistent with conceptualization of schizotypy as representative of only a minority of people and harder scales were intended to decrease over-reporting and "false positives." Items with high discrimination values differentiate best between people low and high on the trait level. We eliminated several items based on relatively low discrimination values, which is an improvement over existing scales. For example, Winterstein et al. (2011) identified several items with low discrimination values in the questionnaires comprising the WSS, and reported the Physical Anhedonia Scale was the lowest of the four scales in ability to differentiate.

Benefits of scale development with DIF. Similarly, the MSS is the first schizotypy scale developed incorporating DIF methods, which represents an improvement given the inadequacy of classical test procedures to detect item bias. Schizotypy scale score

differences have been reported based on gender and race/ethnicity, which begs the question of whether there is bias inherent in these measures. Evidence suggests there is, for example Winsterstein et al. (2011) reported that 48% of the Revised Social Anhedonia Scale and 60% of the Physical Anhedonia Scale items displayed DIF. None of the positive or disorganized MSS subscale items showed significant DIF for any of the three groups (sex, White vs. not White, and White vs. Black/AA) in the cross-validation sample.

The negative subscale contained four items with significant DIF (two for sex, one for White vs. not White, and one for White vs. Black/AA). The two items with DIF for race/ethnicity were both easier for White participants. Both items with sex DIF involved diminished experience of emotions ("rarely feel strong emotions" and "emotions have almost always seemed flat") and were more likely to be endorsed by male participants. Diminution in the experience of emotions is a key feature of negative schizotypy and schizophrenia, therefore we argue these items are important to include.

Overall, only two items with sex DIF and two with race/ethnicity DIF represents a marked improvement over previous measurement of negative schizotypy. Further, the exclusion of items with high DIF resulted in a lack of significant differences on any of the three subscales by race/ethnicity (in contrast with other measures). Similarly, male and female participants did not differ on positive and disorganized schizotypy, and showed the small and expected effect of men scoring higher than women on negative schizotypy. Therefore, the MSS offers a measure with presumably less bias in measuring

schizotypy in different groups of people. This is an advantage in investigating whether past findings represent true trait differences amongst groups of people, which should inform the understanding of risk factors for the development of schizophrenia.

Sex and racial/ethnic differences. As hypothesized, men showed slightly higher scores on negative schizotypy and there were no sex differences on the positive or disorganized subscales. This is consistent with a body of work showing sex differences in schizotypy (subclinical and full-blown schizophrenia), and particularly more pronounced negative features in men (Hafner, 2003; Kulkarni et al., 2012; Kwapil, Crump, & Pickup, 2002; Plocka & Rybakowski, 1992). These differences persisted on the negative subscale, despite the removal of significant bias (DIF) inherent in other measures of negative schizotypy. Differences were not observed on any of the three subscales related to racial/ethnic group membership. This contrasts with previous findings, and suggests that item bias played at least some role in those differences. Future studies should administer the MSS to large, diverse samples and researchers using other common schizotypy measures should examine the potential role of item bias in their findings.

Limitations

Dimensionality. This study used eigenvalues and factor loading from EFA to examine the unidimensionality of the individual MSS subscales. Slocum-Gori, Zumbo, Michalos, and Diener (2009) distinguished between unidimensionality (one and only one latent variable) and essential unidimensionality (one predominant latent variable and secondary minor latent variables) and cite Humphreys' (1962, 1982) assertion that the

measurement of psychological variables will inherently include numerous minor latent variables. Therefore, the assumption of essential unidimensionality is likely more appropriate for the MSS subscales; however, there are limitations in this study's assessment of unidimensionality. First, conclusions drawn from the three different eigenvalue rules were inconsistent: the ratio rules both suggested one prominent factor while the Kaiser stopping rule suggested four, four, and two factors for the positive, negative, and disorganized subscales, respectively. This is not unexpected given previous work showing the Kaiser rule to over-estimate the number of factors (Fabrigar et al., 1999; Hattie, 1985). These and other authors have raised strong criticisms regarding the use of eigenvalues in selecting factors, although others (Slocum-Gori & Zumbo, 2011) have supported their use, when employed in conjunction with other decision-making parameters.

This study is limited in that it did not include other decision making criteria, aside from the use of CFA. Given their findings that no one indicator (of the nine they tested) performed correctly across all the manipulated conditions, Slocum-Gori & Zumbo (2011) argued for the use of multiple indicators when making decisions. Another potential limitation is that the negative MSS subscale narrowly met criteria for one dominant factor using the ratio > 4 rule. This subscale showed eigenvalues of 3.0 and 3.1 for the second factor in the derivation and cross-validation samples, respectively. This raises the question of whether this subscale truly has one dominant and several "minor" factors or if the second factor represents something more. Further testing using

multiple methods is needed to investigate the dimensionality of the subscales, with particular attention paid to negative schizotypy.

A third limitation is the examination of each subscale independently. This was done because the IRT models used assume unidimensionality. This assumption was not directly tested in our CFA models. CFA models were run comparing one-, two-, and three-dimensional factor structures, with best fit for a three-factor model. The models with two and three latent variables allowed these to be correlated; however, a three-factor model with orthogonal factors may provide greater insight into the unidimensionality of the subscales depending on model fit. CFA could also be used to examine the fit for one-factor models of each subscale independently.

Differential Test Functioning (DTF). Very few items with significant DIF provide a promising start for suggesting a lack of bias inherent in the MSS subscales. One or more test items with significant DIF may result in the overall test scores being biased (DTF); however, items with DIF may also results in negligible overall DTF. Therefore, it is important to examine whether the overall test (or subtests, in this case) is biased (Chalmers, Counsell, & Flora, 2016). This document reports the test characteristic curves generated separately in the two groups for each of the three DIF analyses, and comparison of these curves is one indicator of DTF, or the lack thereof. More formal examinations of DTF can be conducted through tests of measurement invariance and Millsap (2011) summarizes these approaches. For example, Levine et al. (2003) used structural equation modeling to examine measurement invariance of factor structure

across age and racial/ethnic groups to examine differential functioning . Raju, van der Linden, and Fleer (1995) proposed a prominent framework for IRT-based measures of DTF. Chalmers, Counsell, and Flora (2016) identified several limitations to this framework, including inability to account for the sampling variability of the parameter estimates across groups, and proposed improved DTF statistics using dichotomous and polytomous IRT models. Future studies should investigate DTF for the MSS subscales across groups such as age, gender, and race/ethnicity to make a more definitive statement about the overall bias inherent in the measure. Slocum-Gori and Zumbo (2011) recommended the use of parallel analysis in conjunction with eigenvalues rules when working with large, skewed datasets. Overall, future studies should investigate the factor structure and DTF of the MSS to examine the assumptions of unidimensionality and lack of bias.

Test Information Curves. The MSS subscale test information curves (Figure 1) have implications regarding the recommended use of the measure. Information function is a measure of test quality, and Hambleton and Swaminathan (1985), originally outlined by Birnbaum, (1968), recommend identification and description of a target (desired) information function at the start of test construction, followed by a process of selecting items that result in this desired function. The target information function should vary dependent on the intended purpose of the test. For example, the information function for a scholarship examination should provide maximal information at higher trait levels for that ability, a test designed to differentiate people above and

below a specific cut point should have a tall, steep curve at the cut point, and a test designed to differentiate people at all trait levels should have a flatter curve with high information and low standard error at a broader range of theta.

As seen in Figure 1, each of the three MSS subscales has a somewhat steep information curve; however, there is currently no known cut point for schizotypy. In other words, there is not a previously identified (and validated) value of theta that differentiates between those of high and low risk. These curves display the low information/high error for people at low and high values of theta, which will result in unstable estimates of theta at these levels. Consistent with the estimated prevalence of schizotypy, most participants fell at lower theta levels; however, the ability of the subscales to discriminate amongst these people becomes smaller than the amount of error at a point around zero on each of the subscales.

The purpose of the scale is to identify schizotypy, which is relatively rare; therefore, it may be uninteresting and/or impossible to add items that increase information at very low levels of theta. However, it would be desirable to add items that increase the width of the curve to provide information at a wider span of theta. Given the lack of validity, the values of 1.2, 1.2, and 1.6 for theta do not have real world meaning and thinner curves around these points may result in missing people at other levels (i.e., a person with a theta value of .5 may be at risk but can't be reliably identified). Following Hambleton and Swaminathan (1985), an approximation of the target information function can be achieved by adding items that provide high

information in the “hard to fill” areas of the existing function. In this case, items with high discrimination values could be added to widen the information curve for each subscale. Examination of the individual item information functions for all the items (those included and excluded in the final scales) revealed that the lower and upper values of the curve peaks (the four out of fifteen values of theta where that item provides the most information) ranged from 0.0 - 2.80, 0.0 - 2.80, and 0.40 - 2.40 for the positive, negative, and disorganized subscales. Since these items were written based on theoretical conceptualizations of schizotypy, it follows that having wider curves with discriminatory power spanning these ranges would be advantageous. Each item was presumed to fall under the umbrella of schizotypy; therefore, allowing the underlying theory to inform the specifications of the desired test information curves may provide more accurate identification of people along the schizotypy continuum.

It should also be noted that IRTPRO assumes a normal distribution, which likely impacted parameter estimation. Other programs (e.g., flexMIRT; Vector Psychometric Group, 2017) include scoring options to minimize the effects of non-normality. An additional point concerns the practical use of the MSS. It is likely that most researchers will not use IRT and instead will simply use summed scores for each of the subscales. Proponents of IRT caution that because the items have a range of difficulty values, a person could appear higher due to answering easier items in the deviant direction. However, if IRT is not possible, there is a very high correlation between theta and summed scores (.94 for the positive, .95 for the negative, and .93 for the disorganized

subscale in our sample); therefore, either method may be appropriate. Finally, as mentioned previously, this study is limited in its lack of validation data. Sound psychometric properties do not speak to whether a test measures what it purports to measure, therefore a process of construct validation is necessary and discussed below.

Future Directions

The MSS was derived and cross validated in a sample aged 18 to 59 years old. We recommend caution in using the MSS with participants outside of this age range until psychometric properties are established for these ages. We expect that the scale would not be appropriate for people younger than age 15 years old, given the reading level of the scale, the content of the items, the fact that schizotypic signs typically first emerge in mid-adolescence, and the possibility of the items being endorsed because of childhood experiences unrelated to schizotypy (e.g., imaginary friends). Future studies should investigate the properties of the MSS in large, diverse samples, and should include participants across the schizotypy continuum (i.e., both clinical and non-clinical samples).

Construct Validation. Good scale development procedures and initial psychometric properties provide a promising start, but the next step is to focus on validation of the scale. In this study, the cross-validation sample was randomly selected from seven samples. This was advantageous as these participants were not included in the derivation analyses used to select items. However, participants in the cross-validation sample were still administered all items in survey 3, which introduces the

potential for higher correlations between scores in the two samples due to shared method variance. Therefore, true cross-validation samples are needed which include only the final MSS items, as well as other measures for convergent and divergent validity. Further, multiple methods of measurement are required to provide evidence for the construct validity of a measure (as outlined by Campbell & Fiske, 1959) as self-report measures (e.g., questionnaires, interview) probably share some variance because they are based on subject awareness and willingness and ability to report on their own experiences.

Lenzenweger (2010), Fonseca-Pedrero et al. (2008), and Kwapil and Barrantes-Vidal (2015) provide useful guidelines for construct validation of schizotypy that ultimately involve identification and differentiation of the processes underlying the positive, negative, and disorganized dimensions of schizotypy. The MSS subscales should undergo a process of construct validation investigating differential etiology, development, phenomenology, and associations with related constructs. Longitudinal studies have demonstrated that positive schizotypy is associated with the development of psychotic disorders, psychotic-like, schizotypal, paranoid, and mood symptoms, and substance use disorders (Chapman, Chapman, Kwapil, Eckblad, & Zinser, 1994; Barrantes-Vidal et al., 2013) and that negative schizotypy predicts negative and schizoid symptoms and social impairment (Barrantes-Vidal et al., 2013; Kwapil et al., 2013), as well as schizophrenia-spectrum diagnoses and symptom severity (Kwapil, 1998). Therefore, the positive MSS subscale should predict psychotic symptoms better than the

negative subscale, and should be associated with mood psychopathology and substances abuse; whereas, the negative subscale should predict negative/schizoid symptoms and social impairment due to social anhedonia, but should not be related to emotion dysregulation.

The positive and negative subscales should also show differential patterns of associations with normal personality. Consistent with the conceptualization of negative schizotypy as deficit or diminution of thought, emotion, enjoyment, and social engagement, the negative MSS subscale should be associated with introversion and decreased openness to experience (Kwapil et al., 2008). Consistent with the conceptualization of positive schizotypy as excess of thought, emotion, and sensory experiences, the positive MSS subscale should be associated with increased openness to experience and neuroticism (Kwapil et al., 2008). The MSS positive subscale should be correlated with creativity (Claridge & Blakey, 2009), whereas the negative should not. In daily life, positive schizotypy is associated with increased negative affect and higher social anxiety, whereas negative schizotypy is associated with a preference to be alone, and lower positive affect (Brown, Silvia, Myin-Germeys, & Kwapil, 2007). Conceptually, the MSS disorganized subscale should be associated with decreased conscientiousness and increased cognitive and behavioral disorganization. Experience-sampling methodology should be used to examine the differential daily life experiences of individuals identified by the MSS subscales to examine construct validity.

Consistent with the schizophrenia literature, the MSS negative subscale should convey risk for worse (or lack of) response to pharmacological treatment interventions and poorer overall prognosis (Carpenter & Koenig, 2008). Construct validation should also include studies involving neuroimaging, as evidence suggests different structural abnormalities associated with positive (e.g., temporal lobe abnormalities) and negative (e.g., tissue decrements in frontal lobe) symptoms of schizophrenia (Buckley, 2005). Similarly, schizotypy is associated with schizophrenic-like patterns of cognitive impairment (e.g., Gooding, Matts, & Rollmann 2006); however, studies who have investigated schizotypy multidimensionally have reported inconsistent findings (see Kane et al., 2016). For example, Kane et al. (2016) reported mild deficits on mind-wandering and response time for positive but not negative schizotypy. Clear predictions are difficult to make for the MSS subscales at this time; however, a multidimensional approach using sound measurement will likely elucidate this area.

Predictions for the disorganized subscale of the MSS are less clear, as fewer studies of schizotypy have included this factor (Kerns & Becker, 2008), despite formal thought disorder being a well-established component of schizophrenia (Andreasen, 1979). However, evidence suggests disorganized schizotypes show cognitive impairment similar to those with disorganized symptoms of schizophrenia (Kerns & Becker, 2008; Moritz, Andresen, Naber, Krausz & Probsthein, 1999). Further, at least two studies reported differential impairment amongst schizotypy factors: disorganized (but not positive or negative) schizotypy was associated with poor cognitive control task

performance (processes involved in carrying out goal-directed behavior in the face of conflict; Kerns, 2006; Moritz et al., 1999). Kerns & Becker (2008) reported communication disturbances, working memory deficits, and poor emotion processing associated with disorganized schizotypy (although they did not investigate positive or negative schizotypy). Further, disorganized schizotypy has been associated with greater emotional ambivalence (simultaneously conflicting emotions; Kerns, 2006; Kerns & Becker, 2008), which is consistent with the conceptualization of disorganized schizotypy as dysregulation. Therefore, it is expected that the disorganized subscale of the MSS should be associated with similar patterns of cognitive impairment as seen in schizophrenia.

Genetic research should more explicitly address the heterogeneity of both clinical and subclinical manifestations of schizotypy, and it is predicted that the MSS subscales may be associated with different patterns of genetic risk. Takahashi (2013) cited the etiological and symptom heterogeneity of schizophrenia as a major source of conflict in elucidating the genetic etiology of schizophrenia, and found differential associations with genetic markers when dividing participants with schizophrenia into five groups (excitement/hostility, negative symptoms, depression/anxiety, positive symptoms, and disorganization). Finally, rigorous translation methods (e.g., Van de Vijver & Hambleton, 1996) should be followed to provide valid translations into other languages and examine the cross-cultural validity of the MSS. The development of this new scale should provide the basis for an ongoing program of construct validation

research, including the above-mentioned topics and countless others, to examine the degree to which the MSS measures what it purports to be measuring.

Brief Scale. Finally, 77 items may prove too long for some researchers to include in their studies; therefore, we have used similar processes as outlined in this paper to derive the MSS-Brief (Gross, Kwapil, Silvia, Raulin & Barrantes-Vidal, 2017) from the full-length MSS. This version contains 38 items that maintain content coverage of the three conceptualized dimensions and have comparable psychometric properties. This measure should also undergo a process of construct validation to investigate its utility in measuring multidimensional schizotypy.

REFERENCES

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders: DSM-III-R*. Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. Washington, D.C: American Psychiatric Association.

Andreasen, N.C. (1979). Thought, language, and communication disorders: I. Clinical assessment, definition of terms, and evaluation of their reliability. *Archives of General Psychiatry, 36*, 1315-1321. doi:10.1001/archpsyc.1979.01780120045006

Andreasen, N.C., & Carpenter, W.J. (1993). Diagnosis and Classification of Schizophrenia. *Schizophrenia Bulletin, 19*, 199-214. doi: 10.1093/schbul/19.2.199

Andreasen, N.C., Nopoulos, P., Schultz, S., Miller, D., Gupta, S., Swayze, V., & Flaum, M. (1994). Positive and negative symptoms of schizophrenia: Past, present, and future. *Acta Psychiatrica Scandinavica Supplementum, 90*, 51-59. doi: 10.1111/j.1600-04

Arndt, S., Alliger, R.J., & Andreasen, N.C. (1991). The distinction of positive and negative symptoms: The failure of a two-dimensional model. *The British Journal of Psychiatry, 158*, 317-322. doi: 10.1192/bjp.158.3.317 47.1994.tb05891.x

Bandalos, D.L., & Enders, C.K. (1996). The effects of nonnormality and number of response categories on reliability. *Applied Measurement in Education, 9*, 151-160. doi: 10.1207/s15324818ame0902_4

Barrantes-Vidal, N., Chun, C.A., Myin-Germeys, I., & Kwapil, T.R. (2013). Psychometric schizotypy predicts psychotic-like, paranoid, and negative symptoms in daily life. *Journal of Abnormal Psychology, 122*, 1077-1087. doi: 10.1037/a0034793

Barrantes-Vidal, N., Gross, G.M., Sheinbaum, T., Mitjavila, M., Ballespí, S., & Kwapil, T.R. (2013). Positive and negative schizotypy are associated with prodromal and schizophrenia-spectrum symptoms. *Schizophrenia Research, 145*, 50-55. doi: 10.1016/j.schres.2013.01.007

Bentall, R.P., Claridge, G.S., & Slade, P.D. (1989). The multidimensional nature of schizotypal traits: A factor analytic study with normal subjects. *British Journal of Clinical Psychology, 28*, 363-375. doi: 10.1111/j.2044-8260.1989.tb00840.x

Birnbaum, A. (1968). Classification by ability levels. In F.M. Lord & M.R. Novick, *Statistical theories of mental test scores* (pp. 436-452). Reading, MA.: Addison-Wesley.

Blanchard, J.J., Collins, L.M., Aghevli, M., Leung, W.W., & Cohen, A.S. (2011). Social anhedonia and schizotypy in a community sample: The Maryland Longitudinal Study of Schizotypy. *Schizophrenia Bulletin, 37*, 587-602. doi: 10.1093/schbul/sbp107

Bleuler, E.P. (1950). *Dementia praecox or the group of schizophrenias* (J. Zinkin, Trans.). New York: International Universities Press. (Original work published in 1911).

Bouvard, M., & Cosma, P. (2008). An exploratory study of a personality disorders questionnaire. *L'encéphale: Revue De Psychiatrie Clinique Biologique Et Thérapeutique, 34*, 517-525. doi: 10.1016/j.encep.2007.08.006

Bora, E., & Arabaci, L.B. (2009). Confirmatory factor analysis of schizotypal personality traits in university students. *Türk Psikiyatri Dergisi, 20*, 1-6.

Browne, M.W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111-150. doi: 10.1207/S15327906MBR3601_05

Brown, L.H., Silvia, P.J., Myin-Germeys, I., Lewandowski, K.E., & Kwapil, T.R. (2008). The relationship of social anxiety and social anhedonia to psychometrically identified schizotypy. *Journal of Social and Clinical Psychology, 27*, 127-149. doi: 10.1521/jscp.2008.27.2.127

Buckley, P.F. (2005). Neuroimaging of schizophrenia: Structural abnormalities and pathophysiological implications. *Neuropsychiatric Disease and Treatment, 1*, 193-204.

Burch, G.J., Steel, C., & Hemsley, D.R. (1998). Oxford–Liverpool Inventory of Feelings and Experiences: Reliability in an experimental population. *British Journal of Clinical Psychology, 37*, 107-108. doi: 10.1111/j.2044-8260.1998.tb01284.x

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105. doi: 10.1037/h0046016

Carpenter, W.T., & Koenig, J.I. (2008). The evolution of drug development in schizophrenia: Past issues and future opportunities. *Neuropsychopharmacology, 33*, 2061-2079. doi: 10.1038/sj.npp.1301639

Chalmers, R.P., Counsell, A., & Flora, D.B. (2016). It might not make a big DIF: Improved differential test functioning statistics that account for sampling variability. *Educational and Psychological Measurement, 76*, 114-140. doi: 10.1177/0013164415584576

Chapman, L.J., Chapman, J.P., 1983. Infrequency scale for personality measures. Unpublished scale available from T.R. Kwapil, UIUC Department of Psychology, Champaign, IL 61820.

Chapman, J.P., Chapman, L.J., & Kwapil, T.R. (1994). Does the Eysenck Psychoticism Scale predict psychosis? A ten year longitudinal study. *Personality and Individual Differences, 17*, 369-375. doi: 10.1016/0191-8869(94)90284-4

Chapman, J.P., Chapman, L.J., & Kwapil, T.R. (1995). Scales for the measurement of schizotypy. In A. Raine, T. Lencz, S.A. Mednick (Eds.), *Schizotypal personality* (pp. 79-106). New York, NY US: Cambridge University Press.

Chapman, L.J., Chapman, J.P., Kwapil, T.R., Eckblad, M., & Zinser, M.C. (1994). Putatively psychosis-prone subjects 10 years later. *Journal of Abnormal Psychology, 103*, 171-183. doi: 10.1037/0021-843X.103.2.171

Chapman, L.J., Chapman, J.P., & Raulin, M.L. (1978). Body-image aberration in schizophrenia. *Journal of Abnormal Psychology, 87*, 399-407. doi: 10.1037/0021-843X.87.4.399

Chan, R.K., Shi, H., Geng, F., Liu, W., Yan, C., Wang, Y., & Gooding, D.C. (2015). The Chapman psychosis-proneness scales: Consistency across culture and time. *Psychiatry Research, 228*, 143-149. doi: 10.1016/j.psychres.2015.04.031

Chen, W.J., Hsiao, C.K., & Lin, C.H. (1997). Schizotypy in community samples: The three-factor structure and correlation with sustained attention. *Journal of Abnormal Psychology, 106*, 649-654. doi: 10.1037/0021-843X.106.4.649

Chernyshenko, O.S., Stark, S., Chan, K., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research, 36*, 523-562. doi: 10.1207/S15327906MBR3604_03

Chmielewski, P.M., Fernandes, L.L., Yee, C.M., & Miller, G.A. (1995). Ethnicity and gender in scales of psychosis proneness and mood disorders. *Journal of Abnormal Psychology, 104*, 464-470. doi: 10.1037/0021-843X.104.3.464

Chmielewski, M. & Watson, D. (2008). The heterogeneous structure of schizotypal personality disorder: Item-level factors of the schizotypal personality

questionnaire and their associations with obsessive-compulsive disorder symptoms, dissociative tendencies, and normal personality. *Journal of Abnormal Psychology, 117*, 364-376. doi: 10.1037/0021-843X.117.2.364

Claridge, G., & Blakey, S. (2009). Schizotypy and affective temperament: Relationships with divergent thinking and creativity styles. *Personality and Individual Differences, 46*, 820-826. doi: 10.1016/j.paid.2009.01.015

Claridge, G., & Broks, P. (1984). Schizotypy and hemisphere function: I. Theoretical considerations and the measurement of schizotypy. *Personality and Individual Differences, 5*, 633-648. doi: 10.1016/0191-8869(84)90111-9

Cochrane, M., Petch, I., & Pickering, A.D. (2010). Do measures of schizotypal personality provide non-clinical analogues of schizophrenic symptomatology? *Psychiatry Research, 176*, 150-154. doi: 10.1016/j.psychres.2009.01.031

Cohen, A.S., Callaway, D.A., Najolia, G.M., Larsen, J.T., & Strauss, G.P. (2012). On 'risk' and reward: Investigating state anhedonia in psychometrically defined schizotypy and schizophrenia. *Journal of Abnormal Psychology, 121*, 407-415. doi: 10.1037/a0026155

Cohen, A.S., Matthews, R.A., Najolia, G.M., & Brown, L.A. (2010). Toward a more psychometrically sound brief measure of schizotypal traits: Introducing the SPQ-Brief Revised. *Journal of Personality Disorders, 24*, 516-537. doi: 10.1521/pedi.2010.24.4.516

Compton, M.T., Goulding, S.M., Bakeman, R., & McClure-Tone, E.B. (2009). Confirmation of a four-factor structure of the Schizotypal Personality Questionnaire among undergraduate students. *Schizophrenia Research, 111*, 46-52. doi: 10.1016/j.schres.2009.02.012

Cornblatt, B.A., Green, M.F., & Walker, E.F. (1999). Schizophrenia: Etiology and neurocognition. In T. Millon, P.H. Blaney, R.D. Davis (Eds.), *Oxford textbook of psychopathology* (pp. 277-310). New York, NY US: Oxford University Press.

Costa, P. T., & McCrae, R. R. (1992). *NEO-PI-R professional manual*. Odessa, FL: PAR.

Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354. doi: 10.1037/h0047358

Daneluzzo, E., Bustini, M., Stratta, P., Casacchia, M., & Rossi, A. (1998). Schizotypal Personality Questionnaire and Wisconsin Card Sorting Test in a population of DSM-III-R schizophrenic patients and control subjects. *Comprehensive Psychiatry, 39*, 143-148. doi: 10.1016/S0010-440X(98)90073-6

Del Goleto, S., Kostova, M., & Blanchet, A. (2016). Impaired context processing during irony comprehension in schizotypy: An ERPs study. *International Journal of Psychophysiology, 105*, 17-25. doi: 10.1016/j.ijpsycho.2016.04.009

DeVellis, R.F. (2012). Scale development: Theory and Applications (3rd ed.). Washington, DC: SAGE Publications.

Drewes, D.W. (2000). Beyond the Spearman–Brown: A structural approach to maximal reliability. *Psychological Methods, 5*, 214–227.

Earleywine, M. (2006). Schizotypy, marijuana, and differential item functioning. *Human Psychopharmacology: Clinical and Experimental, 21*, 455-461. doi: 10.1002/hup.802

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

Eckblad, M.L., Chapman, L.J., Chapman, J.P., & Mishlove, M. (1982). The Revised Social Anhedonia Scale. Unpublished test copies available from T.R. Kwapil, UNCG Department of Psychology, Greensboro, NC 27402.

Edelen, M.O., Stucky, B D., & Chandra, A. (2015). Quantifying 'problematic' DIF within an IRT framework: Application to a cancer stigma index. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 24*, 95-103. doi: 10.1007/s11136-013-0540-4

Eysenck, H.J. & Eysenck, S.B.G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder & Stoughton.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299. doi: 10.1037/1082-989X.4.3.272

Fernandes, L.O.L. & Miller, G.A. (1995). Compromised performance and abnormal psychophysiology associated with the Wisconsin psychosis-proneness scale. In G.A. Miller (Ed.), *The behavioral high-risk paradigm in psychopathology* (pp. 47-87). New York: Springer-Verlag.

Finch, W.H. (2011). A comparison of factor rotation methods for dichotomous data. *Journal of Modern Applied Statistical Methods, 10*, 549-570.

Fonseca-Pedrero, E., Fumero, A., Paino, M., de Miguel, A., Ortuño-Sierra, J., Lemos-Giráldez, S., & Muñiz, J. (2014). Schizotypal Personality Questionnaire: New sources of validity evidence in college students. *Psychiatry Research, 219*, 214-220. doi: 10.1016/j.psychres.2014.04.054

Fonseca-Pedrero, E., Ortuño-Sierra, J., Sierro, G., Daniel, C., Cella, M., Preti, A., & ... Mason, O.J. (2015). The measurement invariance of schizotypy in Europe. *European Psychiatry, 30*, 837-844. doi: 10.1016/j.eurpsy.2015.07.005

Fonseca-Pedrero, E., Paino, M., Lemos-Giráldez, S., Sierra-Baigrie, S., & Muñiz, J. (2010). Factor structure and measurement invariance of the Wisconsin Schizotypy Scales across gender and age. *The Spanish Journal of Psychology, 13*, 941-950. doi: 10.1017/S1138741600002584

Fossati, A., Raine, A., Carretta, I., Leonardi, B., & Maffei, C. (2003). The three-factor model of schizotypal personality: Invariance across age and gender. *Personality and Individual Differences, 35*, 1007-1019. doi: 10.1016/S0191-8869(02)00314-8

Goldberg, L.R. (1972). Parameters of personality inventory construction and utilization: A comparison of prediction strategies and tactics. *Multivariate Behavioral Research Monographs*, *72-2*, 59.

Gooding, D.C., Matts, C.W., & Rollmann, E.A. (2006). Sustained attention deficits in relation to psychometrically identified schizotypy: Evaluating a potential endophenotypic marker. *Schizophrenia Research, 82*, 27-37. doi: 10.1016/j.schres.2005.11.015

Goodman, J.K., Cryder, C.E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of mechanical Turk samples. *Journal of Behavioral Decision Making, 26*, 213-224. doi: 10.1002/bdm.1753

Gottesman, I.I., & Gould, T.D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *The American Journal of Psychiatry, 160*, 636-645. doi: 10.1176/appi.ajp.160.4.636

Graves, R.E., & Weinstein, S. (2004). A Rasch analysis of three of the Wisconsin Scales of Psychosis Proneness: Measurement of schizotypy. *Journal of Applied Measurement*, *5*, 160-171.

Gruzelier, J.H. (1996). The factorial structure of schizotypy: Part I. Affinities with syndromes of schizophrenia. *Schizophrenia Bulletin, 22*, 611-620. doi: 10.1093/schbul/22.4.611

Gross, G.M. & Kwapil, T.R. (2014). Mapping out the multidimensional structure of schizotypy. Unpublished manuscript.

Gross, G.M., Kwapil, T.R., Silvia, P.J., Raulin, M.L., Barrantes-Vidal, N. (2017). The Multidimensional Schizotypy Scale-Brief: Scale development and psychometric properties. Submitted for Publication.

Gross, G.M., Mellin, J., Silvia, P.J., Barrantes-Vidal, N., & Kwapil, T.R. (2014). Comparing the factor structure of the Wisconsin Schizotypy Scales and the Schizotypal Personality Questionnaire. *Personality Disorders: Theory, Research, And Treatment, 5*, 397-405. doi: 10.1037/per0000090

Häfner, H. (2003). Gender differences in schizophrenia. *Psychoneuroendocrinology, 28*(Suppl2), 17-54. doi: 10.1016/S0306-4530(02)00125-7

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff Pub.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Thousand Oaks, CA, US: Sage Publications, Inc.

Hancock, G.R., Mueller, R.O. (2001) Rethinking construct reliability within latent variable systems. In: Cudeck, R., du Toit, S., Sörbom, D. (Eds.), *Structural Equation Modeling: Present and Future*. Scientific Software International, Lincolnwood, pp. 195-216.

Harvey, C.A., Curson, D.A., Pantelis, C., Taylor, J., & Barnes, T.R. (1996). Four behavioural syndromes of schizophrenia. *The British Journal of Psychiatry, 168*, 562-570. doi: 10.1192/bjp.168.5.562

Hattie (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 20*, 1-14.

Hauser, D.J., & Schwarz, N. (2016). Attentive turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods, 48*, 400-407. doi: 10.3758/s13428-015-0578-z

Hornick, C.W., James, L.R., & Jones, A.P. (1977). Empirical item keying versus a rational approach to analyzing a psychological climate questionnaire. *Applied Psychological Measurement, 1*, 489-500. doi: 10.1177/014662167700100405

Humphreys, L.G. (1962). The organization of human abilities. *American Psychologist, 17*, 475-483. doi: 10.1037/h0041550

Humphreys, L.G. (1982). Systematic heterogeneity of items in tests of meaningful psychological attributes: A rejection of unidimensionality. University of Illinois (Unpublished).

IBM Corp. (Released 2011). IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.

Jackson, D.N. (1970). A sequential system for personality scale development. In Spielberger, C.D. (Ed.), *Current Topics in Clinical and Community Psychology*, (Vol. 2, pp. 61-96). Academic, San Diego.

Kaczorowski, J.A., Barrantes-Vidal, N., & Kwapil, T.R. (2009). Neurological soft signs in psychometrically identified schizotypy. *Schizophrenia Research, 115*, 293-302. doi: 10.1016/j.schres.2009.06.018

Kaiser, H.F. (1970). A second generation little jiffy. *Psychometrika, 35*, 401-415. doi: 10.1007/BF02291817

Kane, M.J., Meier, M.E., Smeekens, B.A., Gross, G.M., Chun, C.A., Silvia, P.J., & Kwapil, T.R. (2016). Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of Experimental Psychology: General, 145*, 1017-1048. doi: 10.1037/xge0000184

Karami, H. (2012). An introduction to differential item functioning. *The International Journal of Educational and Psychological Assessment, 11*, 59-76.

Kelley, M.P., & Coursey, R.D. (1992). Factor structure of Schizotypy Scales. *Personality and Individual Differences, 13*, 723-731. doi:10.1016/0191-8869(92)90243-I

Kerns, J.G. (2006). Schizotypy facets, cognitive control, and emotion. *Journal of Abnormal Psychology, 115*, 418-427. doi: 10.1037/0021-843X.115.3.418

Kerns, J. G., & Becker, T. M. (2008). Communication disturbances, working memory, and emotion in people with elevated disorganized schizotypy. *Schizophrenia Research, 100*, 172-180. doi: 10.1016/j.schres.2007.11.005

Klein, R.A., Ratliff, K.A., Vianello, M., Adams, R.J., Bahník, Š., Bernstein, M.J., & ... Nosek, B.A. (2014). Investigating variation in replicability: A 'many labs' replication project. *Social Psychology, 45*, 142-152. doi: 10.1027/1864-9335/a000178

Kraepelin, E. (1919). *Dementia praecox and paraphrenia*. Edinburgh, Scotland: Livingstone. (Original work published 1913).

Krinjen, W.P. (1994). Algorithms for unweighted least-squares factor analysis. *Computational Statistics and Data Analysis, 21*, 133-147.

Kulkarni, J., Gavrilidis, E., Worsley, R., & Hayes, E. (2012). Role of estrogen treatment in the management of schizophrenia. *CNS Drugs, 26*, 549-557. doi: 10.2165/11630660-000000000-00000

Kwapil, T.R. (1998). Social anhedonia as a predictor of the development of schizophrenia-spectrum disorders. *Journal of Abnormal Psychology, 107*, 558-565. doi: 10.1037/0021-843X.107.4.558

Kwapil, T.R., & Barrantes-Vidal, N. (2012). Schizotypal personality disorder: An integrative review. In T. A. Widiger, T. A. Widiger (Eds.), *The Oxford handbook of personality disorders* (pp. 437-477). New York, NY, US: Oxford University Press. doi: 10.1093/oxfordhb/9780199735013.013.0021

Kwapil, T.R., & Barrantes-Vidal, N. (2015). Schizotypy: Looking back and moving forward. *Schizophrenia Bulletin, 41*, suppl 2, S366-S373. doi: 10.1093/schbul/sbu186

Kwapil, T.R., Crump, R.A., & Pickup, D.R. (2002). Assessment of psychosis proneness in African-American college students. *Journal of Clinical Psychology, 58*, 1601-1614. doi: 10.1002/jclp.10078

Kwapil, T.R., Barrantes-Vidal, N., & Silvia, P.J. (2008). The dimensional structure of the Wisconsin Schizotypy Scales: Factor identification and construct validity. *Schizophrenia Bulletin, 34*, 444-457. doi: 10.1093/schbul/sbm098

Kwapil, T.R., Brown, L.H., Silvia, P.J., Myin-Germeys, I., & Barrantes-Vidal, N. (2012). The expression of positive and negative schizotypy in daily life: An experience sampling study. *Psychological Medicine, 42*, 2555-2566. doi: 10.1017/S0033291712000827

Kwapil, T.R., & Chun, C.A. (2015). The psychometric assessment of schizotypy. In O.J. Mason, G. Claridge, O.J. Mason, G. Claridge (Eds.), *Schizotypy: New dimensions* (pp. 7-32). New York, NY, US: Routledge/Taylor & Francis Group.

Kwapil, T.R., Gross, G.M., Silvia, P.J., & Barrantes-Vidal, N. (2013). Prediction of psychopathology and functional impairment by positive and negative schizotypy in the Chapmans' ten-year longitudinal study. *Journal of Abnormal Psychology, 122*, 807-815. doi: 10.1037/a0033759

Kwapil, T.R., Gross, G.M., Silvia, P.J., Raulin, M.L., & Barrantes-Vidal, N. (2017). Development and psychometric properties of the Multidimensional Schizotypy Scale: A new measure for assessing positive, negative, and disorganized Schizotypy. *Schizophrenia Research*, in press.

Kwapil, T.R., Ros-Morente, A., Silvia, P.J., & Barrantes-Vidal, N. (2012). Factor invariance of psychometric schizotypy in Spanish and American samples. *Journal of Psychopathology and Behavioral Assessment, 34*, 145-152. doi: 10.1007/s10862-011-9258-1

Lenzenweger, M.F. (1994). Psychometric high-risk paradigm, perceptual aberrations, and schizotypy: An update. *Schizophrenia Bulletin, 20*, 121-135. doi: 10.1093/schbul/20.1.121

Lenzenweger, M.F. (2010). *Schizotypy and schizophrenia: The view from experimental psychopathology*. New York, NY, US: Guilford Press.

Lenzenweger, M.F., & Dworkin, R.H. (1996). The dimensions of schizophrenia phenomenology: Not one or two, at least three, perhaps four. *The British Journal of Psychiatry, 168*, 432-440. doi: 10.1192/bjp.168.4.432

Lewandowski, K.E., Barrantes-Vidal, N., Nelson-Gray, R.O., Clancy, C., Kepley, H.O., & Kwapil, T.R. (2006). Anxiety and depression symptoms in psychometrically identified schizotypy. *Schizophrenia Research, 83*, 225-235. doi: 10.1016/j.schres.2005.11.024

Liddle, P.F. (1987). The symptoms of chronic schizophrenia: A re-examination of the positive-negative dichotomy. *The British Journal of Psychiatry, 151*, 145-151. doi: 10.1192/bjp.151.2.145

Lin, A., Wigman, J.W., Nelson, B., Wood, S.J., Vollebergh, W.M., van Os, J., & Yung, A.R. (2013). Follow-up factor structure of schizotypy and its clinical associations in a help-seeking sample meeting ultra-high risk for psychosis criteria at baseline. *Comprehensive Psychiatry, 54*, 173-180. doi: 10.1016/j.comppsych.2012.06.011

Lissitz & Green (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology, 60*, 10-13. doi: 10.1037/h0076268

Liu, Y., Wu, A.D., & Zumbo, B.D. (2010). The impact of outliers on Cronbach's coefficient alpha estimate of reliability: Ordinal/rating scale item responses. *Educational and Psychological Measurement, 70*, 5-21. doi: 10.1177/0013164409344548

Maniaci, M.R. & Rogge, R.D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality, 48*, 61-83. doi: 10.1016/j.jrp.2013.09.008

Mason, O., Claridge, G., & Jackson, M. (1995). New scales for the assessment of schizotypy. *Personality and Individual Differences, 18*, 7-13. doi: 10.1016/0191-8869(94)00132-C

Mason, O., & Claridge, G. (2006). The Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE): Further description and extended norms. *Schizophrenia Research, 82*, 203-211. doi: 10.1016/j.schres.2005.12.845

McCrae, R.R. & Costa, P.T. (2010). *NEO inventories for the NEO Personality Inventory-3, NEO-PI-3, NEO Five-Factor Inventory-3, NEO-FFI-3 Professional Manual*. PAR, Lutz, FL.

McNamara, T. & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA & Oxford: Blackwell.

Meehl, P.E. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist, 17*, 827-838. doi: 10.1037/h0041029

Meehl, P.E. (1964). *Manual for use with checklist of schizotypic signs*. Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Menezes, P.R. (2009). Incidence and outcome of schizophrenia across the globe. In W.F. Gattaz, G. Busatto. (Eds.), *Advances in Schizophrenia Research* (pp. 3-17). New York, NY US: Springer New York. doi: 10.1007/978-1-4419-0913-8

Miers, T.C., Raulin, M.L., 1987 Cognitive Slippage Scale. In Corcoran, K., Fischer, J. (Eds.), *Measures for Clinical Practice: A Source Book*. Free Press, New York, pp. 125-127.

Miller, J.D., Crowe, M., Weiss, B., Maples-Keller, J.L., & Lynam, D.R. (2017). Using online, crowdsourcing platforms for data collection in personality disorder research: The example of Amazon's Mechanical Turk. *Personality Disorders: Theory, Research, and Treatment, 8*, 26-34. doi: 10.1037/per0000191

Millsap, R.E. (2011). *Statistical approaches to measurement invariance*. New York, NY, US: Routledge/Taylor & Francis Group.

Moritz, S., Andresen, B., Naber, D., Krausz, M., & Probsthein, E. (1999). Neuropsychological correlates of schizotypal disorganisation. *Cognitive Neuropsychiatry, 4*, 343-349. doi: 10.1080/135468099395873

Mueser, K.T., & Jeste, D.V. (2008). *Clinical handbook of schizophrenia*. New York, NY, US: Guilford Press.

Muthén, L.K., & Muthén, B.O. (2010). Mplus 6.1 [Computer software]. Los Angeles, CA: Author.

Necka E.A., Cacioppo S., Norman G.J., & Cacioppo J.T. (2016) Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community

participants. *PLoS ONE 11*(6): e0157732. https://doi.org/10.1371/journal.pone.0157732

Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement, 24*, 50-64. doi: 10.1177/01466216000241003

Paolacci, G., Chandler, J., & Ipeirotis, P.G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*, 411-419.

Park, S., Holzman, P.S., & Lenzenweger, M.F. (1995). Individual differences in spatial working memory in relation to schizotypy. *Journal of Abnormal Psychology, 104*, 355-363. doi: 10.1037/0021-843X.104.2.355

Park, S., & McTigue, K. (1997). Working memory and the syndromes of schizotypal personality. *Schizophrenia Research, 26*, 213-220. doi: 10.1016/S0920-9964(97)00051-0

Peralta, V., & Cuesta, M.J. (2001). How many and which are the psychopathological dimensions in schizophrenia? Issues influencing their ascertainment. *Schizophrenia Research, 49*, 269-285. doi: 10.1016/S0920-9964(00)00071-2

Plocka, M. & Rybakowski, J. (1992) Gender differences in schizophrenia. *Psychiatria Polska, 26*, 327–336.

Raine, A. (1991). The SPQ: A scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophrenia Bulletin, 17*, 555-564. doi: 10.1093/schbul/17.4.555

Raine, A. (2001). *Manual for the Schizotypal Personality Questionnaire (SPQ and SPQ-B)*. Unpublished manual available from the author.

Raine, A., & Allbutt, J. (1989). Factors of schizoid personality. *British Journal of Clinical Psychology, 28*, 31-40. doi: 10.1111/j.2044-8260.1989.tb00809.x

Raine, A. & Baker, L. (1992). The Schizotypal Personality Questionnaire: Genetics, psychophysiology, neuropsychology, and gender differences. Western Psychological Association, Portland, Oregon, April 30 - May 3.

Raine, A., Benishay, D., Lencz, T., & Scarpa, A. (1997). Abnormal orienting in schizotypal personality disorder. *Schizophrenia Bulletin, 23*, 75-82. doi: 10.1093/schbul/23.1.75

Raine, A., Reynolds, C., Lencz, T., Scerbo, A., Triphon, N., & Kim, D. (1994). Cognitive-perceptual, interpersonal, and disorganized features of schizotypal personality. *Schizophrenia Bulletin, 20*, 191-201. doi: 10.1093/schbul/20.1.191

Readable.io [online program]. (2017). https://readable.io/text/

Reise, S.P., Horan, W.P., & Blanchard, J.J. (2011). The challenges of fitting an item response theory model to the Social Anhedonia Scale. *Journal of Personality Assessment, 93*, 213-224. doi: 10.1080/00223891.2011.558868

Reise, S.P., & Waller, N.G. (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement, 14*, 45-58. doi: 10.1177/014662169001400105

Reynolds, C.A., Raine, A., Mellingen, K., Venables, P.H., & Mednick, S.A. (2000). Three-factor model of schizotypal personality: Invariance across culture, gender, religious affiliation, family adversity, and psychopathology. *Schizophrenia Bulletin, 26*, 603-618. doi: 10.1093/oxfordjournals.schbul.a033481

Rossi, A. & Daneluzzo, E. (2002). Schizotypal dimensions in normals and schizophrenic patients: A comparison with other clinical samples. *Schizophrenia Research, 54*, 67-75. doi: 10.1016/S0920-9964(01)00353-X

Rouse, S.V., Finger, M.S., & Butcher, J.N. (1999). Advances in clinical personality measurement: An item response theory analysis of the MMPI-2 PSY-5 scales. *Journal of Personality Assessment, 72*, 282-307. doi: 10.1207/S15327752JP720212

Scientific Software International, Inc. (2015). IRTPRO, Version 3. Skokie, IL: Scientific Software International, Inc.

Scientific Software International, Inc. (2015) IRTPRO: User's Guide. Skokie, IL: Scientific Software International, Inc.

Slocum-Gori, S.L., & Zumbo, B.D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social Indicators Research, 102*, 443-461. doi: 10.1007/s11205-010-9682-8

Slocum-Gori, S.L., Zumbo, B.D., Michalos, A.C., & Diener, E. (2009). A note on the dimensionality of quality of life scales: An illustration with the Satisfaction with

Life Scale (SWLS). *Social Indicators Research, 92*, 489-496. doi: 10.1007/s11205-008-9303-y

Stefanis, N.C., Smyrnis, N., Avramopoulos, D., Evdokimidis, l., Ntzoufras, l., & Stefanis, C.N. (2004). Factorial composition of self-rated schizotypal traits among young males undergoing military training. *Schizophrenia Bulletin, 30*, 335-350. doi: 10.1093/oxfordjournals.schbul.a007083

Suhr, J.A., & Spitznagel, M.B. (2001). Factor versus cluster models of schizotypal traits. I: A comparison of unselected and highly schizotypal samples. *Schizophrenia Research, 52*, 231-239. doi: 10.1016/S0920-9964(00)00170-5

Takahashi, S. (2013). Heterogeneity of schizophrenia: Genetic and symptomatic factors. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 162*, 648-652. doi: 10.1002/ajmg.b.32161

Tandon, R., Gaebel, W., Barch, D. M., Bustillo, J., Gur, R.E., Heckers, S., & ... Carpenter, W. (2013). Definition and description of schizophrenia in the DSM-5. *Schizophrenia Research, 150*, 3-10. doi: 10.1016/j.schres.2013.05.028

Tandon, R., Nasrallah, H.A., & Keshavan, M.S. (2009). Schizophrenia, 'just the facts' 4: Clinical features and conceptualization. *Schizophrenia Research, 110*, 1-23. doi: 10.1016/j.schres.2009.03.005

van de Vijver, F., & Hambleton, R.K. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1*, 89-99. doi: 10.1027/1016-9040.1.2.89

Vázquez-Barquero, J., Lastra, I., Cuesta Nuñez, M.J., Herrera Castanedo, S., & Dunn, G. (1996). Patterns of positive and negative symptoms in first episode of schizophrenia. *The British Journal of Psychiatry, 168*, 693-701. doi: 10.1192/bjp.168.6.693

Vollema, M.G., & van den Bosch, R.J. (1995). The multidimensionality of schizotypy. *Schizophrenia Bulletin, 21*, 19-31. doi: 10.1093/schbul/21.1.19

Weiss, D.J. (1995). Improving individual differences measurement with item response theory and computerized adaptive testing. In D.J. Lubinski, R.V. Dawis, D.J. Lubinski, R.V. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 49-79). Palo Alto, CA, US: Davies-Black Publishing.

Winterstein, B.P., Ackerman, T.A., Silvia, P.J., & Kwapil, T.R. (2011). Psychometric properties of the Wisconsin Schizotypy Scales in an undergraduate sample: Classical test theory, item response theory, and differential item functioning. *Journal of Psychopathology and Behavioral Assessment, 33*, 480-490. doi: 10.1007/s10862-011-9242-9

Wu, E.Q., Birnbaum, H.G., Shi, L., Ball, D.E., Kessler, R.C., Moulis, M., & Aggarwal, J. (2005). The economic burden of schizophrenia in the United States in 2002. *The Journal of Clinical Psychiatry, 66*, 1122-1129. doi: 10.4088/JCP.v66n0906

Wuthrich, V.M., & Bates, T.C. (2006). Confirmatory factor analysis of the three-factor structure of the Schizotypal Personality Questionnaire and Chapman Schizotypy

Scales. *Journal of Personality Assessment, 87*, 292-304. doi: 10.1207/s15327752jpa8703_10

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. Albany, NY, US: State University of New York Press.

APPENDIX A

TABLES AND FIGURES

Table 1. Summary of Item Administrations

Admin-istration	Survey Version[a]	Date	Sample[b]	Total sample size	Dropped missing	Dropped infrequency	Dropped > age 59	Usable sample size	% female	Age M (SD)
1	1	Spring 2015	UNCG	166	18	24	0	124	62.1	19.4 (1.8)
2	1	Fall 2015	UNCG/YSU	953	79	61	0	813	71.6	19.7 (4.2)
3	2	Spring 2016	UNCG/YSU	1055	88	124	0	843	68.4	20.5 (4.2)
4	3	May 2016	MTURK	391	28	37	21	305	56.7	35.9 (9.8)
5	3	June 2016	MTURK	1296	71	105	61	1059	63.9	34.4(10.5)
6	3	Fall 2016	UNCG	724	27	89	0	608	75.2	18.9 (2.3)
7	3	Fall 2016	YSU	641	15	90	0	536	70.8	20.0 (3.8)
8	3	Fall 2016	TTU	409	15	80	0	314	57.4	19.5 (4.2)
9	3	Fall 2016	UIUC	915	21	128	0	766	68.3	19.2 (1.5)
10	3	October 2016	MTURK	761	0	71	33	657	60.6	35.2 (9.1)
11	3	October 2016	MTURK	723	0	80	27	616	59.4	34.7(10.0)
12	3	November 2016	MTURK	716	0	58	34	624	64.3	34.4 (10.0)
Total				8750	362	947	176	7265	65.9	26.4 (10.4)

[a]survey 1: 246 schizotypy items, NEO-FFI, Social Desirability Scale, Infrequency Scale, Attentive Responding Scale (ARS); survey 2: 155 schizotypy items, NEO-FFI, Social Desirability Scale, Infrequency Scale, ARS; survey 3: 118 schizotypy items, NEO-FFI, Infrequency Scale, ARS

[b]sample: UNCG = University of North Carolina at Greensboro, YSU = Youngstown State University, TTU = Tennessee Technological University, UIUC = University of Illinois at Urbana-Champaign, MTurk = Amazon Mechanical Turk

Table 2. Demographics for Derivation and Cross-Validation Samples

	Derivation Sample n = 6265	Cross-Validation Sample n = 1000
Sex	1975 male, 4290 female	500 male, 500 female
Age in Years: Mean (SD)	26.4 (10.4)	26.7 (10.2)
Age in Years: Range	18-59	18-59
Race/Ethnicity		
White	4429 (71%)	695 (70%)
Black/African American	763 (12%)	114 (11%)
Hispanic/Latino	371 (6%)	63 (6%)
Asian/Pacific Islander	434 (7%)	88 (9%)
Native American	42 (1%)	4 (<1%)
Other	225 (4%)	36 (4%)
English as First Language	5890 (94%)	931 (93%)

Table 3. Item-level CTT Statistics from the Derivation Sample for the MSS Positive Schizotypy Subscale

Items	N	Mean	Point-biserial Correlations				
			SocD	Positive	Negative	Disorgan-ized	Neuroti-cism
1. I believe that dreams have magical properties.	6262	0.31	-0.05	0.54	0.01	0.18	0.15
2. I believe that ghosts or spirits can influence my life.	6264	0.27	-0.13	0.50	0.00	0.16	0.16
3. I believe that I could read other peoples' minds if I really tried.	6260	0.07	-0.04	0.43	0.08	0.16	0.07
4. I have had the momentary feeling that I might not be human.	6262	0.10	-0.16	0.48	0.17	0.30	0.22
5. Some people can make me aware of them just by thinking about me.	6261	0.13	-0.02	0.50	0.05	0.19	0.09
6. I have had the momentary feeling that someone's place has been taken by a look-alike.	6259	0.06	-0.08	0.44	0.13	0.25	0.13
7. I often wonder if everyone in the world is part of a secret experiment.	6261	0.14	-0.12	0.51	0.14	0.27	0.19
8. I have worried that people on other planets may be influencing what happens on Earth.	6261	0.09	-0.14	0.43	0.09	0.20	0.12
9. I occasionally have the feeling that my thoughts are not my own.	6262	0.09	-0.09	0.48	0.15	0.32	0.21
10. I have sometimes felt that strangers were reading my mind.	6260	0.09	-0.12	0.53	0.10	0.30	0.22
11. I have felt that there were messages for me in the way things were arranged, like furniture in a room.	6261	0.07	-0.08	0.51	0.05	0.22	0.13
12. Sometimes I feel that a television show or movie has a special message just for me.	6262	0.19	-0.13	0.49	0.02	0.23	0.16
13. I believe that there are secret signs in the world if you just know how to look for them.	6258	0.37	-0.16	0.60	0.05	0.24	0.19
14. I sometimes wonder if there is a small group of people who can control everyone else's behavior.	6261	0.07	-0.09	0.47	0.14	0.23	0.14
15. I occasionally worry that people I see on the street are spying on me.	6258	0.09	-0.13	0.46	0.15	0.31	0.23
16. I often worry that other people are out to get me.	6261	0.14	-0.26	0.50	0.17	0.41	0.36
17. I often think that I hear people talking only to discover that there was no one there.	6260	0.12	-0.19	0.55	0.15	0.36	0.24
18. Occasionally I have felt as though my body did not exist.	6259	0.10	-0.16	0.49	0.15	0.33	0.26
19. At times I have wondered if my body was really my own.	6259	0.09	-0.15	0.49	0.15	0.31	0.23
20. I have felt that something outside my body was a part of my body.	6258	0.06	-0.10	0.45	0.11	0.22	0.11
21. There are times when it feels like someone is touching me when no one is actually there.	6258	0.15	-0.20	0.55	0.11	0.29	0.23
22. Sometimes when I look at ordinary objects they seem strange or unreal.	6261	0.19	-0.19	0.58	0.16	0.37	0.24
23. There are times when I think I see another person, but there is actually no one there.	6258	0.16	-0.15	0.50	0.10	0.28	0.24
24. I have had experiences with seeing the future, ESP or a sixth sense.	5326	0.22	-0.09	0.49	0.08	0.15	0.14
25. I often worry that someone or something is controlling my behavior.	5328	0.06	-0.08	0.47	0.13	0.33	0.21
26. I often find hidden meanings or threats in things that people say or do.	5325	0.23	-0.19	0.55	0.14	0.36	0.30

Table 4. Item-level CTT Statistics from the Derivation Sample for the MSS Negative Schizotypy Subscale

Items	N	Mean	Point-biserial Correlations				
			SocD	Positive	Negative	Disorganized	Neuroticism
1. Throughout my life I have noticed that I rarely feel strong positive or negative emotions.	6262	0.14	0.04	0.07	0.42	0.14	0.01
2. I rarely feel strong emotions even in situations in which other people usually do.	6264	0.20	-0.01	0.11	0.45	0.18	0.02
3. Throughout my life there have been very few things that interest me.	6260	0.12	-0.04	0.14	0.52	0.31	0.23
4. My emotions have almost always seemed flat regardless of what is going on around me.	6261	0.14	-0.07	0.13	0.50	0.26	0.10
5. Generally I do not have many thoughts or emotions.	6261	0.05	0.00	0.08	0.36	0.17	0.04
6. I often look forward to upcoming events.	6262	0.07	-0.02	0.07	0.44	0.19	0.16
7. Throughout my life, very few things have been exciting or interesting to me.	6261	0.10	-0.02	0.14	0.55	0.31	0.22
8. I tend to have few interests.	6263	0.17	-0.08	0.11	0.50	0.29	0.24
9. I have always preferred to be disconnected from the world.	6263	0.16	-0.11	0.23	0.56	0.35	0.24
10. Having close friends is not as important as people say.	6262	0.12	-0.04	0.07	0.49	0.12	0.09
11. I have never really been interested in having close relationships.	6263	0.07	-0.05	0.07	0.51	0.16	0.09
12. In general, it is important for me to have close relationships with other people.	6260	0.13	-0.06	0.04	0.57	0.10	0.07
13. When I move to a new place, I feel a strong desire to make friends.	6259	0.33	-0.09	0.04	0.53	0.13	0.14
14. If given the choice, I would much rather be with another person than alone.	6257	0.26	-0.04	0.01	0.46	0.07	0.05
15. Although there are things I enjoy doing by myself, I usually have more fun when I do things with other people.	6259	0.15	-0.08	0.06	0.51	0.13	0.14
16. I enjoy meeting new people and making new friends.	6262	0.16	-0.12	0.04	0.57	0.18	0.19
17. It has never been important to me to be involved with other people.	6260	0.11	-0.01	0.10	0.51	0.16	0.08
18. Most of the time I feel a desire to be connected with other people.	6263	0.22	-0.01	0.01	0.59	0.11	0.08
19. Throughout my life, I have had little interest in dating or being in a romantic relationship.	6260	0.10	0.01	0.11	0.39	0.16	0.10
20. I generally am not interested in being emotionally close with others.	6262	0.18	-0.06	0.10	0.63	0.18	0.10
21. There are just not many things that I have ever really enjoyed doing.	6257	0.11	-0.06	0.17	0.51	0.33	0.25
22. I have little or no interest in sex or romantic relationships.	6260	0.08	0.01	0.09	0.38	0.15	0.12
23. I greatly enjoy traveling to new places.	6261	0.07	-0.08	0.04	0.34	0.10	0.11
24. Just being with other people can make me feel good.	5328	0.10	-0.05	0.04	0.48	0.13	0.11
25. Spending time with close friends and family is important to me.	5325	0.05	-0.06	0.07	0.49	0.15	0.10
26. Having a meal with other people is almost always better than eating alone.	5324	0.19	-0.15	0.07	0.48	0.13	0.12

Table 5. Item-level CTT Statistics from the Derivation Sample for the MSS Disorganized Schizotypy Subscale

Items	N	Mean	Point-biserial Correlations				
			SocD	Positive	Negative	Disorgan-ized	Neuroti-cism
1. Most of the time I find it is very difficult to get my thoughts in order.	6262	0.22	-0.19	0.31	0.23	0.71	0.43
2. No matter how hard I try, I can't organize my thoughts.	6263	0.11	-0.11	0.25	0.23	0.65	0.33
3. Even when I have time, it is almost impossible to organize my thoughts.	6264	0.16	-0.14	0.28	0.24	0.65	0.34
4. Most of the time my thoughts seem clear and organized.	6261	0.22	-0.16	0.26	0.18	0.63	0.39
5. My thoughts are so hazy and unclear that I wish that I could just reach up and put them into place.	6262	0.16	-0.18	0.32	0.23	0.66	0.37
6. My thoughts almost always seem fuzzy and hazy.	6261	0.10	-0.17	0.28	0.25	0.61	0.33
7. Things slip my mind so often that it's hard to get things done.	6263	0.25	-0.18	0.32	0.22	0.68	0.39
8. I have a hard time staying on topic while speaking.	6263	0.25	-0.16	0.30	0.19	0.59	0.35
9. My thoughts often feel so jumbled that I have difficulty doing anything.	6260	0.18	-0.18	0.32	0.24	0.70	0.40
10. My thoughts are almost always hard to follow.	6262	0.15	-0.17	0.34	0.28	0.71	0.37
11. I find that I am very often confused about what is going on around me.	6261	0.14	-0.15	0.33	0.22	0.62	0.35
12. I often find that when I talk to people I don't make any sense to them.	6259	0.15	-0.11	0.37	0.26	0.58	0.32
13. People find my conversations to be confusing or hard to follow.	6262	0.14	-0.09	0.31	0.27	0.56	0.30
14. I have trouble following conversations with others.	6260	0.10	-0.14	0.27	0.27	0.64	0.32
15. When people ask me a question, I often don't understand what they are asking.	6261	0.09	-0.13	0.28	0.22	0.54	0.27
16. It is usually easy for me to follow conversations.	6260	0.10	-0.06	0.24	0.25	0.54	0.28
17. My lack of organization often makes it hard to do the things that I am supposed to do.	6261	0.23	-0.22	0.26	0.20	0.62	0.35
18. My thoughts and behaviors are almost always disorganized.	6261	0.15	-0.17	0.31	0.22	0.69	0.36
19. I often feel so disconnected from the world that I am not able to do things.	6262	0.12	-0.19	0.34	0.33	0.56	0.38
20. My thoughts and behaviors feel random and unfocused.	6261	0.17	-0.16	0.37	0.26	0.71	0.39
21. I often have difficulty organizing what I am supposed to be doing.	6257	0.24	-0.21	0.28	0.22	0.68	0.38
22. When I try to do one thing, I often become confused and start doing something else.	6260	0.22	-0.17	0.36	0.21	0.67	0.37
23. I often feel so mixed up that I have difficulty functioning.	5325	0.15	-0.20	0.37	0.25	0.65	0.43
24. I often struggle to stay organized enough to complete simple tasks throughout the day.	5324	0.20	-0.16	0.30	0.21	0.66	0.39
25. I often have difficulty following what someone is saying to me.	5326	0.12	-0.14	0.31	0.24	0.64	0.32

Table 6. Item Level CTT Statistics from Derivation Sample for Survey 3 Items Eliminated from Final Subscales

			Point-biserial Correlations			
	N	M	Positive	Negative	Disorganized	Neuroticism
Positive Items						
EPOS1. I sometimes sense the presence of a force or person around me, even though I cannot see anyone.	6265	0.37	0.56	0.06	0.24	0.21
EPOS2. It is not possible to heal someone just by using your mind.	6261	0.26	0.29	-0.03	0.04	0.01
EPOS3. Objects such as crystals and good luck charms have special powers.	6262	0.16	0.48	0.01	0.16	0.14
EPOS3. At times I perform little rituals to keep bad things from happening.	6260	0.22	0.40	0.03	0.16	0.19
EPOS4. I believe that I can cause something to happen just by thinking about it.	6265	0.16	0.50	0.06	0.18	0.10
EPOS5. I sometimes have a feeling of gaining or losing energy when certain people look at me or touch me.	6261	0.24	0.47	0.10	0.30	0.23
EPOS6. No one has the power to predict the future.	6259	0.33	0.45	0.00	0.10	0.11
EPOS7. I sometimes worry that computers and electronic devices can control my actions.	6261	0.11	0.41	0.05	0.21	0.15
EPOS8. I believe that it is possible to control other people just by using my mind.	6263	0.06	0.38	0.08	0.14	0.04
EPOS9. I have occasionally had the feeling that a TV or radio broadcaster knew I was listening to him/her.	6261	0.08	0.43	0.07	0.20	0.11
EPOS10. I sometimes think that articles in newspapers, magazines, and the Internet contain messages for me.	6261	0.13	0.46	0.02	0.19	0.12
EPOS11. I have noticed sounds in my music that are not there at other times.	6259	0.20	0.50	0.10	0.30	0.20
EPOS12. I find that I often mistake objects or shadows for people, or noises for voices.	6261	0.23	0.54	0.13	0.40	0.30
EPOS13. I believe that dreams can predict the future.	4485	0.41	0.55	0.02	0.18	0.15
EPOS14.I often worry that strangers know things about me.	4485	0.18	0.47	0.17	0.41	0.34
EPOS15. Strangers always seem to be talking about me behind my back	4482	0.09	0.44	0.17	0.39	0.30
Negative Items	**N**	**M**	**Positive**	**Negative**	**Disorganized**	**Neuroticism**
ENEG1. My emotions are much less intense than other people's.	6262	0.27	-0.01	0.38	0.04	-0.11
ENEG2. I rarely get excited or upset about things (even when I probably should).	6262	0.20	0.09	0.42	0.14	-0.03
ENEG3. I typically have many interesting thoughts and ideas.	6263	0.09	-0.05	0.27	0.11	0.13
ENEG4. Throughout my life, I have rarely felt motivated to do anything.	6264	0.13	0.18	0.46	0.42	0.30
ENEG5. I have always found things that interest and motivate me.	6263	0.12	0.08	0.41	0.28	0.27
ENEG6. I'm much too independent to really get involved with other people.	6261	0.23	0.17	0.51	0.22	0.15
ENEG7. When things are bothering me, I like to talk to someone about it.	6263	0.23	0.05	0.46	0.12	0.07
ENEG8. I have always enjoyed looking at photographs of friends.	6260	0.13	0.00	0.40	0.10	0.05

ENEG9. The sounds of a concert or sporting event have never excited me.	6257	0.07	0.06	0.33	0.12	0.06
ENEG10. It is exciting to visit a big city.	6258	0.10	0.04	0.32	0.08	0.08
ENEG11. The beauty of sunsets is greatly overrated.	6260	0.14	0.08	0.29	0.14	0.09
ENEG12. Flowers aren't as beautiful as many people claim.	6260	0.13	0.06	0.32	0.14	0.09
ENEG.13. In general, I prefer to spend almost all of my time alone.	5325	0.27	0.16	0.62	0.29	0.26
Disorganized Items	**N**	**M**	**Positive**	**Negative**	**Disorganized**	**Neuroticism**
EDIS1. My thoughts are often confused.	6265	0.22	0.32	0.23	0.69	0.43
EDIS2. I struggle to think about things, even if it is a topic I like.	6262	0.12	0.28	0.29	0.55	0.29
EDIS3. Often my thoughts just seem to disappear.	6261	0.21	0.31	0.22	0.52	0.29
EDIS4. I have a hard time controlling my thoughts.	6258	0.26	0.34	0.22	0.61	0.44
EDIS5. I often struggle to make sense of what is going on around me.	6262	0.15	0.34	0.26	0.62	0.35
EDIS6. I often feel confused when I try to explain my ideas.	6263	0.25	0.34	0.26	0.65	0.41
EDIS7. My speech makes sense to me but not to other people.	6259	0.14	0.37	0.27	0.54	0.27
EDIS8. I often feel confused during conversations with more than one person.	6262	0.16	0.31	0.28	0.62	0.34
EDIS9. I often forget what I am doing while I am doing it.	6259	0.25	0.36	0.17	0.60	0.35
EDIS10. People often find my behavior strange or unusual.	6258	0.22	0.40	0.29	0.50	0.31
EDIS11. I am so preoccupied by daydreaming that it is hard to get things done.	5324	0.23	0.31	0.16	0.55	0.37
EDIS12. When I want to be, I am usually about as organized as other people.	5325	0.12	0.12	0.16	0.43	0.23

Note: M = mean score or endorsement frequency of the dichotomous items.

Table 7. Item-level CTT Statistics from the Cross-Validation Sample (N = 1,000) for the MSS Positive Schizotypy Subscale

		Point-biserial Correlations			
Items	**M**	**Positive**	**Negative**	**Disorganized**	**Neuroticism**
POS1	.33	.55	.00	.15	.13
POS2	.27	.54	.02	.17	.14
POS3	.09	.50	.03	.08	-.01
POS4	.11	.54	.10	.19	.17
POS5	.14	.53	.04	.13	.07
POS6	.07	.45	.04	.16	.10
POS7	.14	.61	.16	.26	.17
POS8	.08	.46	.09	.11	.10
POS9	.08	.48	.12	.31	.21
POS10	.10	.51	.06	.24	.17
POS11	.07	.46	.04	.12	.06
POS12	.23	.52	.01	.15	.15
POS13	.37	.57	.03	.17	.13
POS14	.09	.51	.10	.20	.13
POS15	.08	.50	.12	.30	.21
POS16	.13	.48	.22	.45	.39
POS17	.11	.44	.10	.29	.22
POS18	.11	.52	.12	.32	.24
POS19	.09	.55	.12	.29	.22
POS20	.07	.50	.08	.23	.09
POS21	.13	.54	.08	.27	.19
POS22	.19	.60	.14	.32	.24
POS23	.12	.47	.09	.21	.24
POS24	.23	.48	.02	.10	.09
POS25	.06	.49	.07	.30	.21
POS26	.23	.57	.19	.33	.25

Note: M = mean score or endorsement frequency of the dichotomous items.

Table 8. Item-level CTT Statistics from the Cross-Validation Sample (N = 1,000) for the MSS Negative Schizotypy Subscale

		Point-biserial Correlations			
Items	**M**	**Positive**	**Negative**	**Disorganized**	**Neuroticism**
NEG1	.16	.12	.45	.16	.04
NEG2	.21	.10	.45	.18	.00
NEG3	.13	.14	.54	.34	.24
NEG4	.17	.10	.51	.21	.07
NEG5	.05	.08	.26	.07	.01
NEG6	.08	.04	.48	.19	.18
NEG7	.11	.13	.56	.34	.26
NEG8	.19	.13	.51	.32	.26
NEG9	.18	.20	.59	.31	.20
NEG10	.13	.11	.53	.12	.08
NEG11	.07	.09	.59	.17	.09
NEG12	.16	.05	.62	.09	.07
NEG13	.34	.01	.56	.13	.13
NEG14	.22	.03	.51	.06	.05
NEG15	.16	.08	.55	.08	.16
NEG16	.16	.04	.59	.18	.24
NEG17	.14	.15	.55	.18	.06
NEG18	.21	.01	.65	.13	.10
NEG19	.10	.07	.43	.16	.09
NEG20	.19	.09	.68	.20	.11
NEG21	.13	.17	.54	.40	.30
NEG22	.08	.04	.40	.15	.12
NEG23	.08	.01	.33	.06	.11
NEG24	.09	.03	.46	.17	.12
NEG25	.05	.07	.43	.10	.04
NEG26	.20	.06	.51	.12	.13

Note: M = mean score or endorsement frequency of the dichotomous items.

Table 9. Item-level CTT Statistics from the Cross-Validation Sample (N = 1,000) for the MSS Disorganized Schizotypy Subscale

		Point-biserial Correlations			
Items	**M**	**Positive**	**Negative**	**Disorganized**	**Neuroticism**
DIS1	.20	.30	.19	.71	.39
DIS2	.10	.22	.18	.61	.35
DIS3	.18	.24	.21	.63	.34
DIS4	.22	.22	.15	.61	.37
DIS5	.15	.34	.19	.65	.36
DIS6	.08	.22	.19	.58	.26
DIS7	.25	.27	.18	.67	.38
DIS8	.22	.27	.22	.60	.33
DIS9	.17	.29	.25	.72	.42
DIS10	.15	.29	.24	.71	.33
DIS11	.14	.37	.22	.59	.36
DIS12	.14	.30	.26	.57	.29
DIS13	.13	.22	.30	.57	.25
DIS14	.10	.24	.32	.65	.29
DIS15	.08	.25	.23	.55	.22
DIS16	.11	.22	.30	.55	.26
DIS17	.22	.23	.19	.65	.37
DIS18	.14	.28	.24	.75	.37
DIS19	.12	.32	.33	.53	.38
DIS20	.15	.33	.24	.73	.38
DIS21	.22	.28	.15	.72	.39
DIS22	.20	.31	.11	.62	.35
DIS23	.14	.30	.23	.66	.39
DIS24	.18	.24	.17	.66	.36
DIS25	.11	.28	.27	.61	.27

Note: M = mean score or endorsement frequency of the dichotomous items.

Table 10. IRT Model Fit Statistics in the Derivation and Cross-Validation Samples

		Derivation Sample		
		-2loglikelihood:	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)
Positive				
	1PL	99779.08	99833.08	100015.13
	2PL	99334.19	99438.19	99788.82
	3PL	99431.49	99587.49	100113.43
Negative				
	1PL	102713.32	102767.32	102949.37
	2PL	101812.83	101916.83	102267.45
	3PL	101845.25	102001.25	102527.19
Disorganized				
	1PL	93275.44	93327.44	93502.75
	2PL	92575.29	92675.29	93012.43
	3PL	92633.55	92783.55	93289.25
		Cross-Validation Sample		
		-2loglikelihood:	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)
Positive				
	1PL	16375.17	16429.17	16561.68
	2PL	16303.38	16407.38	16662.59
	3PL	16426.73	16582.73	16965.53
Negative				
	1PL	16992.19	17046.19	17178.70
	2PL	16820.41	16924.41	17179.62
	3PL	16899.37	17055.37	17438.18
Disorganized				
	1PL	14779.35	14831.35	14958.95
	2PL	14641.66	14741.66	14987.05
	3PL	14739.54	14889.54	15257.63

Table 11. Item-level IRT/DIF Statistics from the Derivation Sample for the MSS Positive Schizotypy Subscale

	IRT Parameters				DIF Sex	DIF White vs. Not White	DIF White vs. Black/AA
Items	**a**	**s.e.**	**b**	**s.e.**	**X^2b**	**X^2b**	**X^2b**
POS1	1.57	0.05	0.73	0.03	20.3*	3.4	1.2
POS2	1.35	0.05	1.00	0.03	30.7*	8.0	10.3
POS3	1.69	0.08	2.07	0.06	0.5	7.4	3.0
POS4	1.84	0.08	1.78	0.05	11.6*	1.3	7.2
POS5	1.71	0.07	1.63	0.04	3.3	2.7	3.6
POS6	1.93	0.09	2.05	0.06	13.5*	1.6	1.3
POS7	1.78	0.07	1.53	0.04	18.5*	0.1	0.0
POS8	1.63	0.07	2.00	0.06	4.0	0.1	6.4
POS9	1.84	0.08	1.85	0.05	2.8	5.0	1.7
POS10	2.23	0.09	1.67	0.04	0.7	0.3	0.3
POS11	2.20	0.10	1.83	0.04	1.8	5.1	6.3
POS12	1.50	0.06	1.36	0.04	0.4	15.7*	39.7*
POS13	2.06	0.06	0.44	0.02	0.7	0.5	0.0
POS14	2.02	0.09	1.91	0.05	28.4*	8.4	5.9
POS15	1.76	0.08	1.91	0.05	0.4	0.7	3.1
POS16	1.68	0.07	1.54	0.04	5.9	18.2*	13.3*
POS17	2.18	0.08	1.48	0.03	4.5	2.5	2.1
POS18	1.85	0.08	1.72	0.04	0.4	15.2*	20.1*
POS19	1.91	0.08	1.78	0.04	1.6	0.1	0.0
POS20	1.92	0.09	2.06	0.06	29.4*	0.1	2.1
POS21	2.05	0.08	1.38	0.03	2.6	0.7	0.0
POS22	1.96	0.07	1.19	0.03	16.0*	1.2	2.0
POS23	1.66	0.06	1.44	0.04	3.1	7.2	2.1
POS24	1.38	0.05	1.23	0.04	7.2	18.1*	5.5
POS25	2.26	0.12	1.99	0.05	0.3	3.8	0.2
POS26	1.75	0.06	1.04	0.03	0.8	1.2	3.6

Differential Item Functioning Analyses: *****p < .001

Table 12. Item-level IRT/DIF Statistics from the Derivation Sample for the MSS Negative Schizotypy Subscale

	IRT Parameters				DIF Sex	DIF White vs. Not White	DIF White vs. Black/AA
Items	**a**	**s.e.**	**b**	**s.e.**	**X^2b**	**X^2b**	**X^2b**
NEG1	1.13	0.05	1.93	0.07	27.1*	1.6	0.8
NEG2	1.12	0.05	1.54	0.05	31.1*	4.7	6.4
NEG3	1.79	0.07	1.64	0.04	8.9	29.5*	10.3
NEG4	1.51	0.06	1.61	0.05	35.2*	1.4	1.1
NEG5	1.38	0.08	2.60	0.10	10.1	12.9*	7.2
NEG6	1.74	0.08	2.11	0.06	4.0	1.4	4.3
NEG7	2.05	0.08	1.65	0.04	4.2	21.5*	7.7
NEG8	1.46	0.06	1.47	0.04	14.8*	23.7*	7.7
NEG9	1.92	0.07	1.35	0.03	1.5	0.3	0.4
NEG10	1.67	0.07	1.67	0.04	9.3	4.4	0.3
NEG11	2.26	0.10	1.88	0.04	1.1	0.7	0.0
NEG12	2.10	0.08	1.48	0.03	0.4	1.7	1.1
NEG13	1.54	0.05	0.66	0.03	4.1	19.8*	4.6
NEG14	1.26	0.05	1.1	0.04	19.0*	12.4*	8.0
NEG15	1.75	0.07	1.44	0.04	8.2	18.8*	18.8*
NEG16	1.99	0.07	1.32	0.03	15.3*	49.5*	51.0
NEG17	1.82	0.07	1.68	0.04	3.0	13.2*	31.2*
NEG18	2.06	0.07	1.03	0.03	0.8	0.4	2.3
NEG19	1.28	0.06	2.15	0.07	8.1	14.6*	10.2
NEG20	2.45	0.09	1.13	0.02	0.1	1.5	2.1
NEG21	1.76	0.07	1.7	0.04	3.1	9.4	1.2
NEG22	1.32	0.06	2.32	0.08	51.5*	0.4	0.2
NEG23	1.23	0.07	2.65	0.11	5.6	11.1*	11.4*
NEG24	1.78	0.08	1.86	0.05	0.2	0.2	1.4
NEG25	2.59	0.14	2.05	0.05	6.3	0.3	0.3
NEG26	1.37	0.06	1.43	0.04	1.0	0.3	0.4

Differential Item Functioning Analyses: *****p < .001

Table 13. Item-level IRT/DIF Statistics from the Derivation Sample for the MSS Disorganized Schizotypy Subscale

	IRT Parameters				DIF Sex	DIF White vs. Not White	DIF White vs. Black/AA
Items	**a**	**s.e.**	**b**	**s.e.**	**X^2b**	**X^2b**	**X^2b**
DIS1	3.03	0.10	0.87	0.02	7.3	6.8	0.7
DIS2	2.96	0.12	1.42	0.03	1.8	1.5	0.3
DIS3	2.59	0.09	1.19	0.02	1.9	0.1	0.1
DIS4	2.10	0.07	0.98	0.02	7.8	10.3	5.2
DIS5	2.65	0.09	1.17	0.02	2.6	3.7	0.3
DIS6	2.65	0.10	1.54	0.03	0.3	0.5	0.4
DIS7	2.64	0.09	0.82	0.02	0.6	0.2	0.1
DIS8	1.85	0.06	0.93	0.03	23.6*	4.4	2.1
DIS9	2.96	0.10	1.05	0.02	1.1	0.1	0.6
DIS10	3.25	0.12	1.15	0.02	0.0	1.7	0.1
DIS11	2.35	0.08	1.34	0.03	0.2	4.2	0.1
DIS12	2.00	0.07	1.36	0.03	3.2	0.2	0.7
DIS13	1.90	0.07	1.46	0.03	20.3*	0.1	0.2
DIS14	2.85	0.11	1.49	0.03	0.8	0.7	6.9
DIS15	2.16	0.09	1.74	0.04	9.9	0.1	0.0
DIS16	1.96	0.08	1.70	0.04	1.3	0.0	3.4
DIS17	2.17	0.07	0.95	0.02	3.0	1.0	2.2
DIS18	3.03	0.11	1.18	0.02	0.3	11.4*	6.0
DIS19	2.06	0.08	1.54	0.04	5.9	0.3	1.4
DIS20	3.27	0.12	1.10	0.02	1.5	0.1	1.2
DIS21	2.75	0.09	0.83	0.02	0.9	3.9	4.2
DIS22	2.57	0.08	0.93	0.02	1.3	0.0	0.0
DIS23	2.61	0.10	1.25	0.03	0.0	2.1	0.7
DIS24	2.44	0.09	1.01	0.02	0.3	1.1	0.6
DIS25	2.64	0.11	1.42	0.03	1.4	0.2	0.5

Differential Item Functioning Analyses: *****$p < .001$

Table 14. Item-level IRT/DIF Statistics from the Cross-Validation (N = 1,000) Sample for the MSS Positive Schizotypy Subscale

	IRT Parameters				DIF Sex	DIF White vs. Not White	DIF White vs. Black/AA
Items	**a**	**s.e.**	**b**	**s.e.**	**X^2b**	**X^2b**	**X^2b**
POS1	1.56	0.13	0.66	0.06	6.7	0.3	0.2
POS2	1.38	0.13	0.97	0.08	8.9	1.2	0.2
POS3	1.98	0.22	1.83	0.12	0.1	0.0	0.3
POS4	2.14	0.23	1.60	0.09	1.4	0.0	0.4
POS5	1.75	0.17	1.52	0.10	4.8	0.0	0.2
POS6	1.91	0.23	2.05	0.15	0.4	0.1	0.0
POS7	2.36	0.23	1.33	0.07	0.5	2.8	4.6
POS8	1.87	0.22	1.92	0.13	1.5	0.1	0.1
POS9	1.98	0.23	1.86	0.12	0.8	0.9	0.0
POS10	1.91	0.20	1.70	0.11	7.5	0.1	0.0
POS11	1.98	0.24	1.99	0.14	0.6	4.0	6.8
POS12	1.51	0.14	1.15	0.08	0.1	0.8	0.7
POS13	1.64	0.14	0.51	0.06	0.1	0.2	2.1
POS14	2.03	0.22	1.75	0.11	2.5	1.4	0.2
POS15	2.14	0.24	1.81	0.11	1.4	0.0	0.6
POS16	1.61	0.17	1.63	0.11	0.5	0.3	1.7
POS17	1.49	0.17	1.87	0.14	0.0	2.2	4.6
POS18	2.08	0.22	1.61	0.10	0.1	1.1	0.6
POS19	2.43	0.27	1.63	0.09	0.0	0.0	2.0
POS20	2.24	0.27	1.88	0.12	5.7	3.0	3.8
POS21	1.92	0.19	1.50	0.09	0.6	1.5	0.3
POS22	2.12	0.2	1.17	0.07	1.4	1.4	1.3
POS23	1.56	0.17	1.77	0.13	2.0	2.6	1.4
POS24	1.20	0.12	1.27	0.11	0.4	10.5	0.9
POS25	2.23	0.27	1.93	0.12	0.9	0.5	0.0
POS26	1.76	0.16	1.04	0.07	1.2	5.5	2.5

Differential Item Functioning Analyses: *****p < .001

Table 15. Item-level IRT/DIF Statistics from the Cross-Validation Sample (N = 1,000) for the MSS Negative Schizotypy Subscale

	IRT Parameters				DIF Sex	DIF White vs. Not White	DIF White vs. Black/AA
	a	**s.e.**	**b**	**s.e.**	**X^2b**	**X^2b**	**X^2b**
NEG1	1.24	0.13	1.72	0.14	3.8	1.3	0.1
NEG2	1.15	0.12	1.43	0.12	13.2*	2.3	2.3
NEG3	1.76	0.18	1.57	0.10	0.9	3.8	4.6
NEG4	1.51	0.15	1.42	0.10	14.2*	4.0	6.9
NEG5	0.97	0.18	3.44	0.52	2.3	3.6	3.3
NEG6	1.85	0.21	1.95	0.13	1.4	0.9	0.2
NEG7	2.00	0.21	1.62	0.10	2.5	3.7	2.4
NEG8	1.45	0.14	1.38	0.10	3.4	2.6	1.3
NEG9	1.98	0.18	1.22	0.08	2.7	3.9	3.7
NEG10	1.82	0.18	1.58	0.10	2.2	0.6	0.0
NEG11	3.22	0.39	1.67	0.08	0.2	0.0	0.2
NEG12	2.26	0.21	1.25	0.07	0.0	0.0	0.1
NEG13	1.62	0.14	0.61	0.06	0.2	12.7*	2.3
NEG14	1.49	0.14	1.19	0.09	4.0	0.9	0.6
NEG15	1.81	0.17	1.36	0.09	2.4	1.0	1.3
NEG16	2.08	0.19	1.30	0.08	1.1	7.1	12.1*
NEG17	1.92	0.19	1.48	0.09	0.8	2.0	2.2
NEG18	2.42	0.22	1.02	0.06	0.0	0.0	2.0
NEG19	1.34	0.16	2.09	0.18	0.0	0.9	0.0
NEG20	2.95	0.28	1.04	0.05	0.8	0.0	0.1
NEG21	1.80	0.18	1.58	0.10	0.0	2.3	0.2
NEG22	1.39	0.18	2.30	0.20	1.9	0.3	0.7
NEG23	1.09	0.16	2.68	0.30	1.8	2.1	0.4
NEG24	1.70	0.19	1.96	0.14	0.2	0.1	0.1
NEG25	2.01	0.26	2.19	0.16	2.8	0.2	0.3
NEG26	1.47	0.14	1.29	0.09	0.2	0.3	0.4

Differential Item Functioning Analyses: *****p < .001

Table 16. Item-level IRT/DIF Statistics from the Cross-Validation Sample (N = 1,000) for the MSS Disorganized Schizotypy Subscale

	IRT Parameters				DIF Sex	DIF White vs. Not White	DIF White vs. Black/AA
	a	**s.e.**	**b**	**s.e.**	**X^2b**	**X^2b**	**X^2b**
DIS1	2.90	0.25	0.98	0.06	0.1	1.0	1.2
DIS2	2.56	0.26	1.55	0.08	0.2	0.1	0.0
DIS3	2.27	0.20	1.15	0.07	0.5	0.8	1.8
DIS4	1.93	0.17	1.03	0.07	1.3	0.0	0.4
DIS5	2.58	0.23	1.23	0.07	0.0	1.5	0.1
DIS6	2.50	0.27	1.72	0.09	0.0	0.0	0.0
DIS7	2.65	0.22	0.83	0.05	0.3	0.3	1.9
DIS8	1.87	0.16	1.03	0.07	2.5	0.5	0.6
DIS9	3.19	0.29	1.07	0.06	1.7	0.3	0.3
DIS10	3.02	0.28	1.19	0.06	1.0	0.1	1.0
DIS11	2.07	0.19	1.42	0.08	0.1	0.0	0.5
DIS12	1.91	0.18	1.49	0.09	0.2	0.1	0.5
DIS13	1.83	0.17	1.53	0.10	2.9	0.4	0.0
DIS14	2.82	0.29	1.50	0.08	0.9	0.1	0.7
DIS15	2.21	0.24	1.80	0.11	2.5	0.0	0.7
DIS16	1.96	0.20	1.68	0.10	1.3	0.5	0.5
DIS17	2.37	0.20	0.97	0.06	0.0	1.5	0.0
DIS18	3.95	0.40	1.19	0.06	0.5	1.2	0.0
DIS19	1.82	0.18	1.67	0.11	3.9	0.1	0.3
DIS20	3.41	0.33	1.19	0.06	0..0	0.0	0.0
DIS21	3.33	0.3	0.89	0.05	00	0.4	0.1
DIS22	2.13	0.18	1.10	0.07	3.1	0.0	0.7
DIS23	2.51	0.23	1.33	0.07	1.8	0.2	0.2
DIS24	2.44	0.21	1.14	0.06	0.2	0.0	0.6
DIS25	2.37	0.23	1.52	0.08	1.1	1.3	0.4

Differential Item Functioning Analyses: *****$p < .001$

Table 17. Item Level IRT/DIF Statistics from Derivation Sample for Survey 3 Items Eliminated from Final Subscales

	IRT Parameters				DIF Sex	DIF White vs. Not White	DIF White vs. Black/AA
Eliminated POS	**a**	**s.e.**	**b**	**s.e.**	**X^2b**	**X^2b**	**X^2b**
EPOS1	1.66	0.05	0.48	0.02	5.4	14.3*	5.6
EPOS2	0.61	0.03	1.81	0.10	0.8	7.8	5.0
EPOS3	1.49	0.06	1.49	0.04	78.1*	8.4	14.3*
EPOS4	0.99	0.04	1.54	0.06	1.6	3.4	11.0
EPOS5	1.57	0.06	1.45	0.04	2.3	30.0*	34.6*
EPOS6	1.28	0.05	1.16	0.04	1.0	7.6	9.1
EPOS7	1.14	0.04	0.78	0.03	39.8*	0.5	0.3
EPOS8	1.34	0.06	1.98	0.06	0.0	2.2	0.5
EPOS9	1.52	0.08	2.34	0.08	6.8	21.8*	13.4*
EPOS10	1.67	0.08	2.03	0.06	7.2	10.9*	19.2*
EPOS11	1.49	0.06	1.71	0.05	0.7	17.1*	23.8*
EPOS12	1.48	0.05	1.29	0.04	26.5*	0.8	1.8
EPOS13	1.67	0.06	1.05	0.03	4.6	6.2	7.7
EPOS14	1.67	0.06	0.26	0.03	20.6*	5.2	28.9*
EPOS15	1.36	0.06	1.42	0.05	0.9	1.2	5.6
EPOS16	1.57	0.08	1.92	0.07	0.3	3.5	1.7
Eliminated NEG	**a**	**s.e.**	**b**	**s.e.**	**X^2b**	**X^2b**	**X^2b**
ENEG1	0.82	0.04	1.4	0.06	108.1*	3.1	4.3
ENEG2	1.00.	0.04	1.66	0.06	101*	4.0	6.6
ENEG3	0.84	0.05	3.06	0.16	21.6*	5.9	22.9*
ENEG4	1.43	0.06	1.78	0.05	0.0	0.9	4.0
ENEG5	1.28	0.06	2.00	0.07	7.0	8.1	18.4*
ENEG6	1.48	0.05	1.15	0.03	21.7*	3.1	3.3
ENEG7	1.16	0.05	1.3	0.04	49.6*	0.2	0.4
ENEG8	1.17	0.05	1.98	0.07	113.9*	16.4*	12.9*
ENEG9	1.13	0.06	2.69	0.11	0.0	7.5	6.2
ENEG10	0.93	0.05	2.75	0.13	0.8	59.6*	40.8*
ENEG11	0.69	0.04	2.89	0.16	23.5*	10.0	4.6
ENEG12	0.81	0.05	2.59	0.13	93.4*	2.3	5.4
ENEG13	2.16	0.08	0.80	0.02	17.0*	1.9	0.2
Eliminated DIS	**a**	**s.e.**	**b**	**s.e.**	**X^2b**	**X^2b**	**X^2b**
EDIS1	2.72	0.09	0.91	0.02	10.7	1.9	1.8

EDIS2	2.01	0.08	1.55	0.04	0.1	16.4*	4.7
EDIS3	1.51	0.05	1.20	0.03	6.9	0.7	1.5
EDIS4	2.07	0.07	0.85	0.02	7.7	7.7	1.5
EDIS5	2.22	0.08	1.33	0.03	1.9	3.6	2.8
EDIS6	2.34	0.07	0.85	0.02	4.8	0.4	0.3
EDIS7	1.80	0.07	1.47	0.04	16.5*	42.5*	28.0*
EDIS8	2.15	0.08	1.26	0.03	2.6	0.5	2.7
EDIS9	1.95	0.06	0.89	0.02	24.6*	1.3	3.8
EDIS10	1.40	0.05	1.22	0.04	89.6*	0.8	0.1
EDIS11	1.68	0.06	1.03	0.03	1.6	0.3	0.6
EDIS12	1.36	0.06	1.92	0.06	3.7	16.4*	5.6

Differential Item Functioning Analyses: **p < .001

Table 18. 2PL Item Fit Statistics for Positive, Negative, and Disorganized Subscales in Cross-Validation Sample

	Positive			Negative			Disorganized		
	X^2	d.f.	P	X^2	d.f.	P	X^2	d.f.	P
1	22.40	15	0.097	17.09	19	0.585	19.32	17	0.310
2	23.56	17	0.131	20.90	19	0.344	20.40	21	0.498
3	28.67	18	0.052	14.38	19	0.762	15.47	20	0.749
4	15.89	18	0.601	21.17	18	0.270	20.32	20	0.440
5	26.12	18	0.097	7.21	18	0.988	20.98	20	0.400
6	16.61	19	0.617	12.57	19	0.860	16.47	22	0.792
7	20.43	17	0.252	19.53	19	0.425	12.72	17	0.756
8	36.87	18	0.005	18.67	18	0.414	23.30	20	0.274
9	19.19	18	0.382	10.07	17	0.901	12.34	17	0.780
10	13.13	18	0.785	14.26	19	0.769	20.34	19	0.376
11	21.92	19	0.288	10.76	16	0.825	29.84	22	0.122
12	21.90	17	0.188	7.19	17	0.981	20.71	22	0.540
13	14.57	15	0.484	10.53	15	0.786	19.12	21	0.579
14	24.41	18	0.142	17.13	18	0.516	17.01	20	0.653
15	16.58	17	0.485	8.23	18	0.975	13.72	22	0.911
16	20.81	18	0.289	11.32	17	0.840	23.83	22	0.358
17	25.31	19	0.150	14.66	18	0.686	24.90	19	0.163
18	17.39	18	0.498	6.39	15	0.972	29.20	17	0.033
19	20.82	17	0.234	14.89	19	0.730	11.05	22	0.974
20	23.15	17	0.144	6.00	15	0.980	11.84	18	0.856
21	19.64	17	0.292	16.47	17	0.492	15.41	16	0.496
22	25.30	16	0.065	21.94	20	0.345	14.24	20	0.819
23	26.88	18	0.081	16.83	18	0.536	19.23	20	0.508
24	24.65	18	0.135	22.89	19	0.242	17.28	20	0.636
25	13.54	18	0.759	12.20	19	0.877	8.760	20	0.986
26	13.20	15	0.588	12.06	18	0.845			

***p<.001**

Table 19. Descriptive Statistics and Reliability of the Schizotypy Subscales

Subscale	Items	Sample	Mean (SD)	Skew (SE)	Kurtosis (SE)	Alpha	Binary
Positive Schizotypy	26	Derivation	3.58 (4.41)	1.86 (.03)	3.75 (.06)	.89	.89
		Cross-Validation	3.71 (4.50)	1.70 (.08)	2.88 (.16)	.89	.89
Negative Schizotypy	26	Derivation	3.53 (4.36)	1.83 (.03)	3.45 (.06)	.88	.87
		Cross-Validation	3.78 (4.61)	1.66 (.08)	2.51 (.16)	.89	.88
Disorganized Schizotypy	25	Derivation	4.05 (5.81)	1.77 (.03)	2.44 (.06)	.94	.95
		Cross-Validation	3.88 (5.69)	1.83 (.08)	2.64 (.16)	.94	.94

Alpha = Coefficient alpha reliability
Binary = Binary alpha reliability

Table 20. Intercorrelations of the Subscales and Correlations with Neuroticism

Subscale	Positive Schizotypy	Negative Schizotypy	Disorganized Schizotypy	Sex	Neuroticism
Positive Schizotypy		.19*	**.48***	-.01	**.37***
Negative Schizotypy	.16*		**.34***	-.11*	**.24***
Disorganized Schizotypy	**.43***	**.34***		-.01	***.55****
Sex	.02	-.12*	.02		.15*
Neuroticism	**.32***	**.24***	***.55****	.18*	

*$p < .001$

Results for the derivation sample are listed above the diagonal and for the cross validation sample are listed below the diagonal.

Positive correlations with sex indicate higher scores in women.

Medium effect sizes are in bold, large effect sizes in bold and italics.

Table 21. Exploratory Factor Analysis Eigenvalues

	Factors					
Positive Schizotypy	**1**	**2**	**3**	**4**	**5-26**	
Derivation	13.0	1.7	1.2	1.1	<1.0	
Validation	12.7	2.0	1.4	1.1	<1.0	
Negative Schizotypy	**1**	**2**	**3**	**4**	**5-26**	
Derivation	12.2	3.0	1.5	1.3	<1.0	
	1	**2**	**3**	**4**	**5**	**6-26**
Validation	12.3	3.1	1.8	1.3	1.0	<1.0
Disorganized Schizotypy	**1**	**2**	**3-25**			
Derivation	16.7	1.4	<1.0			
Validation	16.4	1.5	<1.0			

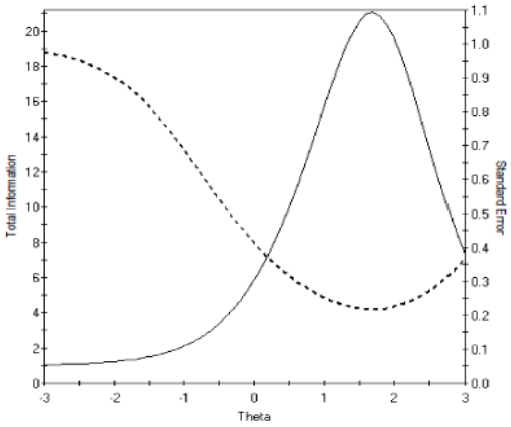
Table 22. Confirmatory Factor Analyses Fit Indices

	Akaike (AIC)	Bayesian (BIC)	Sample-Size Adjusted BIC
Derivation			
1-Factor [a]	316030.914	317069.295	316579.924
2-Factor [b]	302222.488	303267.612	302775.063
3-Factor [c]	293279.093	294337.702	293838.798
Validation			
1-Factor [a]	52007.270	52763.065	52273.952
2-Factor [b]	49527.385	50288.087	49795.798
3-Factor [c]	48001.692	48772.210	48273.569

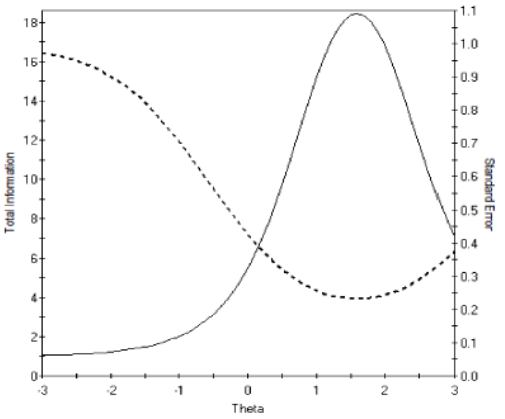
[a]1-Factor = single schizotypy factor
[b]2-Factor = negative schizotypy factor and combined positive and disorganized factor
[c]3-Factor = positive, negative, and disorganized factors

Figure 1. Test Information and Standard Error Curves for the Three Subscales

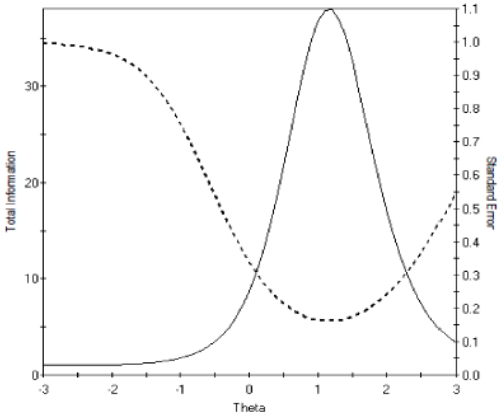
Positive Subscale



Negative Subscale



Disorganized Subscale

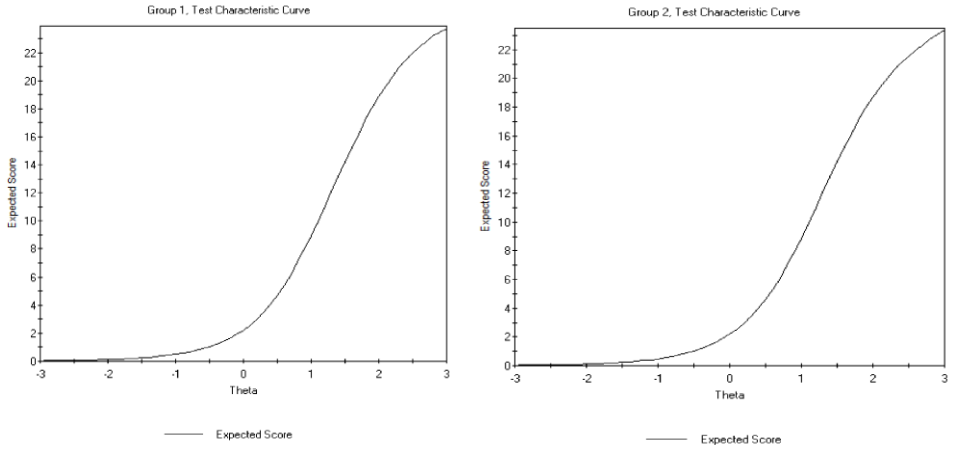


Test Information

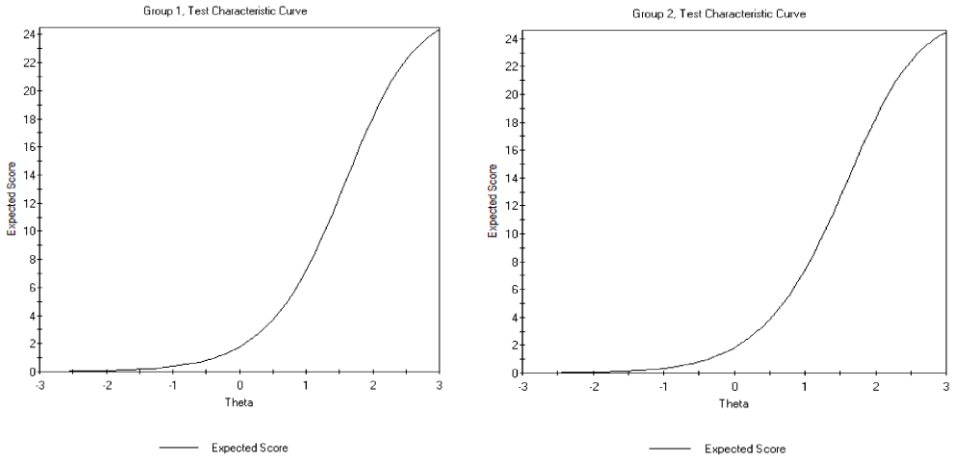
Standard Error

Figure 2. Test Characteristic Curves for Men (Group 1) and Women (Group 2)

Positive Schizotypy



Negative Schizotypy



Disorganized Schizotypy

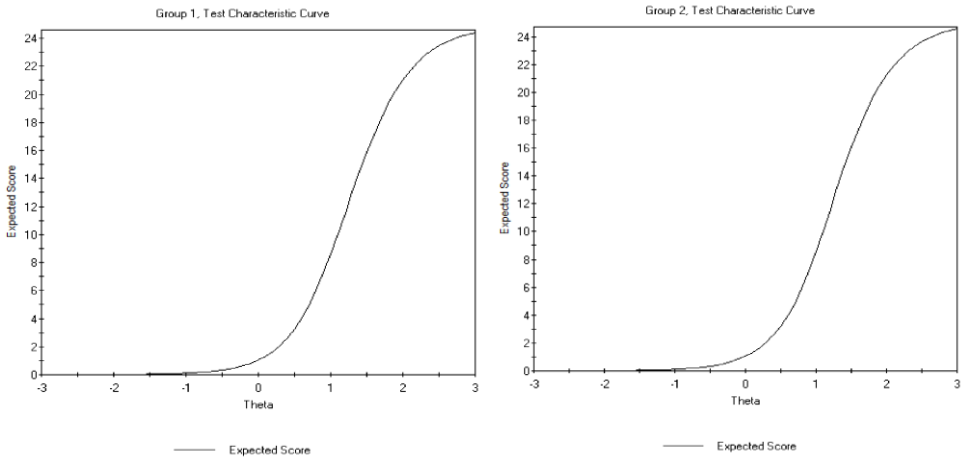
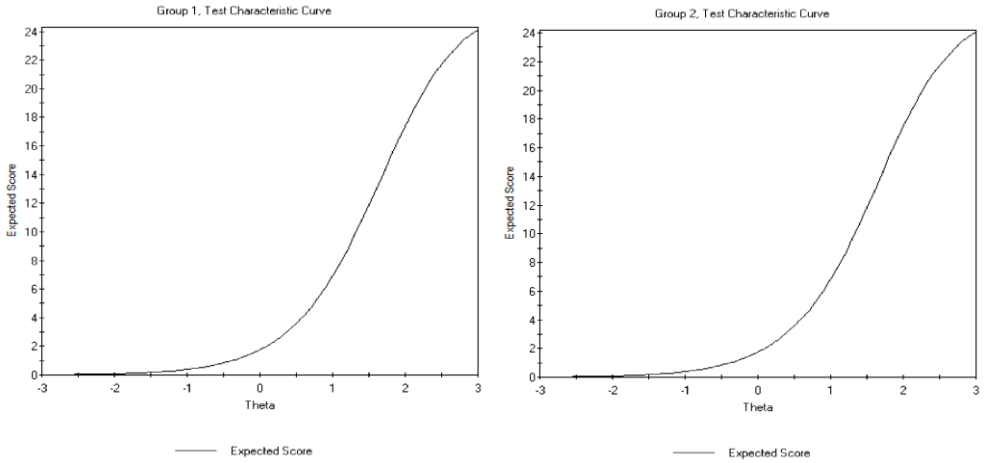
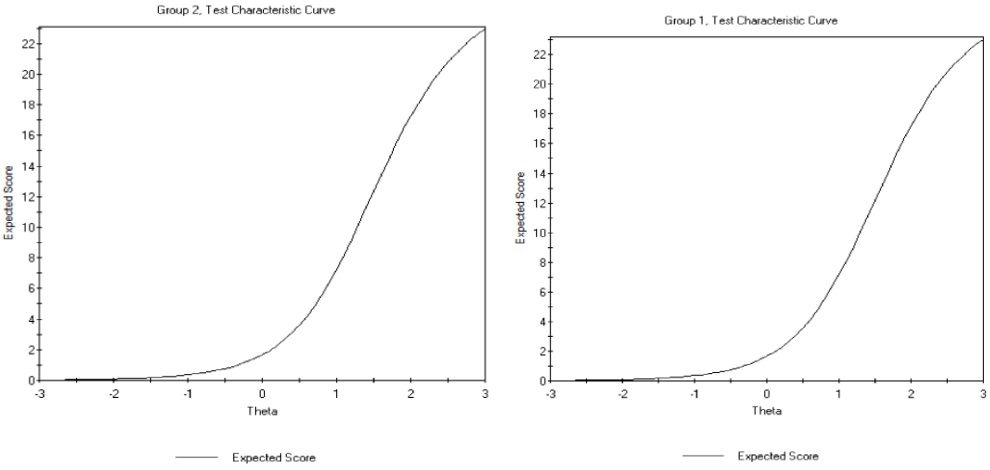


Figure 3. Test Characteristic Curves for White (Group 1) and Not White (Group 2)

Postive Schizotypy



Negative Schizotypy



Disorganized Schizotypy

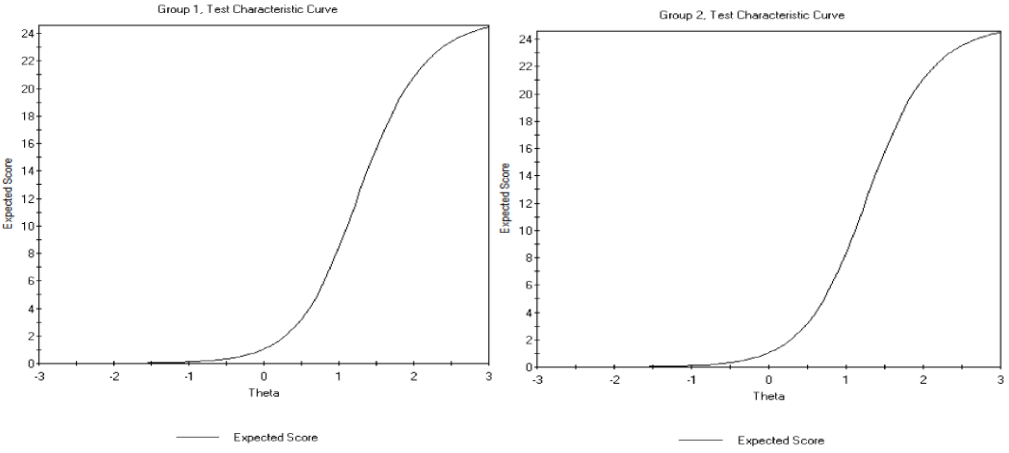
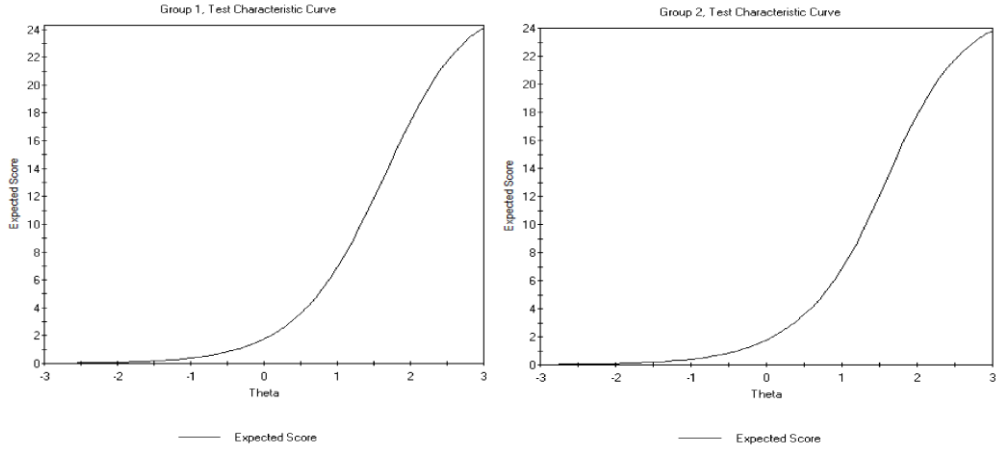
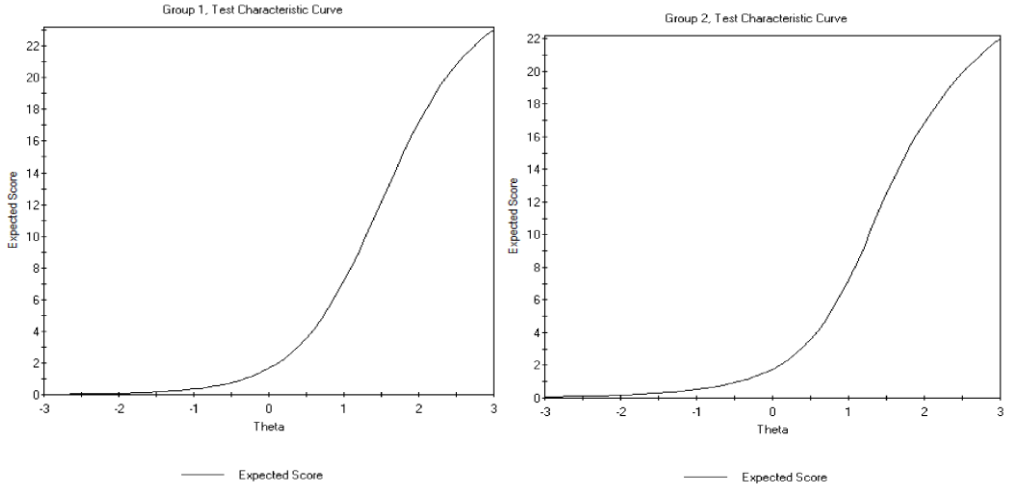


Figure 4. Test Characteristic Curves for White (Group 1) and Black/AA (Group 2)

Positive Schizotypy



Negative Schizotypy



Disorganized Schizotypy