Assessment
2019, Vol. 26(6) 1001–1013
© The Author(s) 2017
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1073191117719509
journals.sagepub.com/home/asm
SAGE

Differential Item Functioning of the Full and Brief Wisconsin Schizotypy Scales in Asian, White, Hispanic, and Multiethnic Samples and Between Sexes

David C. Cicero[1], Elizabeth A. Martin[2], and Alexander Krieg[1]

Abstract

The Wisconsin Schizotypy Scales, including their brief versions, are among the most commonly used self-report measures of schizotypy. Although they have been used extensively in many ethnic groups, few studies have examined their differential item functioning (DIF) across groups. The current study included 1,056 Asian, 408 White, 476 Multiethnic, and 372 Hispanic undergraduates. Unidimensional models of the brief Magical Ideation Scale and Perceptual Aberration Scales fit the data well. For both scales, global tests of measurement invariance provided mixed evidence, but few of the items displayed DIF across ethnicities or between sexes within a multiple indicator multiple causes model. For the full versions of the scales and the brief Revised Social Anhedonia Scale, multiple indicator multiple causes models within an exploratory structural equation modeling framework found that few of the items had DIF. These findings suggest that some of the items may have different psychometric properties across groups, but most items do not.

Keywords

Social Anhedonia Scale, Magical Ideation Scale, Perceptual Aberration Scale, measurement invariance, exploratory structural equation modeling, multiple indicators multiple causes

Schizotypy is a neurodevelopmental condition that includes traits or symptoms similar to symptoms of schizophrenia but in a diminished form (Claridge et al., 1996; Kwapil & Barrantes-Vidal, 2015; Lenzenweger, 2010; Meehl, 1990). In addition, schizotypy represents a risk for the future development of schizophrenia spectrum disorders (Chapman, Chapman, Kwapil, Eckblad, & Zinser, 1994; Kwapil, 1998). Research on schizotypy is important for several reasons. First, it may help understand risk for schizophrenia, which can be used for assessment and prevention efforts (Barrantes-Vidal, Grant, & Kwapil, 2015). Second, schizotypy symptoms can be used to model symptoms of schizophrenia without common confounds of patient research such as medication and more generalized deficits (Kwapil, Crump, & Pickup, 2002; Neale & Oltmanns, 1980). Third, many people with schizotypy experience clinically significant attenuated psychotic symptoms (Cicero, Martin, Becker, Docherty, & Kerns, 2014). Schizotypy comprises symptoms that are similar to Cluster A personality disorders (i.e., Schizotypal, Schizoid, and Paranoid personality disorder), and research on schizotypy may help understand these personality disorders (Tsuang, Stone, Tarbox, & Faraone, 2002).

Three widely used scales to measure schizotypy are the Wisconsin Schizotypy Scales (WSS), including the Magical Ideation Scale (MagicId; Eckbald & Chapman, 1983), Perceptual Aberration Scale (PerAb; Chapman, Chapman, & Raulin, 1978), and the Revised Social Anhedonia Scale (RSAS; Eckbald, Chapman, Chapman, & Mishlove, 1982). Developed over 30 years ago, these scales have considerable support for the reliability and validity of their scores in community (e.g., Blanchard, Collins, Aghevli, Leung, & Cohen, 2011; MacDonald, Pogue-Geile, Debski, & Manuck, 2001), undergraduate (e.g., Chapman et al., 1994; Kwapil, Barrantes-Vidal, & Silvia, 2008; Lenzenweger, 1991), and clinical samples (e.g., Blanchard, Horan, & Brown, 2001; Horan, Reise, Subotnik, Ventura, & Nuechterlein, 2008), and in both longitudinal (Chapman et al., 1994; Gooding, Tallent, & Matts, 2005; Kwapil, 1998) and cross-sectional

[1]University of Hawai'i at Manoa, Honolulu, HI, USA
[2]University of California, Irvine, CA, USA

Corresponding Author:

David C. Cicero, Department of Psychology, University of Hawai'i at Manoa, 2530 Dole Street, Sakamaki D-406, Honolulu, HI 96822-2294, USA.
Email: dcicero@hawaii.edu

studies (Cicero et al., 2014; Kerns, 2006; Lenzenweger, 1994; Martin, Cicero, & Kerns, 2012). For example, participants with high scores on the MagicId and PerAb have been shown to have higher levels of psychosis and psychotic-like experiences, as well as more relatives with schizophrenia spectrum disorders at a 10-year follow-up (Chapman et al., 1994). Moreover, people with high RSAS scores have also been shown to be at risk for schizophrenia spectrum disorders over time (Kwapil, 1998). In cross-sectional studies, scores on the WSS have been shown to be correlated with other measures of schizotypy, including interviews (e.g., Chapman & Chapman, 1980; Cicero et al., 2014; Kwapil, Chapman, & Chapman, 1999). Thus, the reliability and validity of WSS scores are well-established in many different populations.

In addition to the full versions of the WSS, researchers have developed brief versions of the scales (Winterstein, Silvia, et al., 2011). The brief and full versions of the scales are highly correlated with each other (*r*s = .88-.92) and have similar correlations with interview measures of global assessment of functioning, psychotic-like experiences, negative symptoms, Cluster A personality disorders, alcohol and drug impairment, mood symptoms, Big Five personality traits, and social functioning (Gross, Silvia, Barrantes-Vidal, & Kwapil, 2012). Moreover, the brief versions have been shown to have similar internal consistency as the full versions, despite being considerably shorter (e.g., Full RSAS = 40 items vs. Brief RSAS = 15 items). If the brief versions of the scales produce scores with similar reliability and validity, then researchers may choose to use the brief versions of the scales without sacrificing the usefulness of the scales.

Despite the considerable evidence for the validity and reliability of the full scales' scores in many different types of populations, some previous work has identified weaknesses in these scales (Reise, Horan, & Blanchard, 2011; Winterstein, Ackerman, Silvia, & Kwapil, 2011). The scales were developed before the prevalence of some modern statistical techniques such as item response theory (IRT). Recent differential item functioning (DIF) analyses have found that many items on full versions of the scales have DIF between African American and White participants and between men and women (Winterstein, Ackerman, et al., 2011). At the same time, many of the items do not have acceptable discrimination and difficulty parameters. One reason the brief versions of the scales were developed was to address these weaknesses. In the original development study, items that performed poorly in the single group IRT analyses and the multigroup DIF analyses were removed from the scale (Winterstein, Silvia, et al., 2011).

The finding that the full versions of the scales have DIF between White and African American participants suggests a need for DIF analyses across other ethnic groups as well. This will help determine whether the DIF is specific to these ethnic groups or is more widespread. Moreover, the brief versions of the scales were developed, in part, to eliminate the poorly performing items. Testing DIF in other ethnic groups will help determine the success of this effort.

In addition to known DIF between some ethnic groups, there are several reasons to examine the DIF of the WSS specifically in Asian, Hispanic, and Multiethnic populations. Researchers have found differences among ethnicities such that ethnic minorities tend to have higher schizotypy scores than do White participants (Chmielewski, Fernandes, Yee, & Miller, 1995; Kwapil et al., 2002). Research examining ethnic differences in scale scores has found that White college students tend to have lower PerAb, MagicId, and RSAS scores than Hispanic, Asian, and African American samples (Chmielewski et al., 1995). This is consistent with work using other schizotypy scales finding higher rates of schizotypy in ethnic minorities (Linscott, Marie, Arnott, & Clarke, 2006; Sharpley & Peters, 1999), as well as higher rates of schizophrenia (Morgan, Charalambides, Hutchinson, & Murray, 2010; Sharpley, Hutchinson, Murray, & McKenzie, 2001).

In addition, there may be specific cultural differences between White, Asian, Hispanic, and Multiethnic populations that leads to the groups interpreting the items in different ways, resulting in DIF (Bae & Brekke, 2002; Earl et al., 2015). It is possible that we may find DIF for items because of social desirability effects, or people's deliberate misrepresentation of themselves to manage their self-presentation (Paulhus, 1984). That is, if a certain characteristic or trait is seen as particularly socially "undesirable" for a given cultural group, they may be less likely to endorse such items.

Likewise, some participants from certain Asian groups may have spiritual beliefs and experiences that could be misinterpreted as psychotic-like if viewed from a Western frame of reference (e.g., Kim, 2006; Loewenthal, 2006). The MagicId was designed to measure beliefs in causality that are generally considered to be invalid by conventional standards. If these conventional standards differ between groups, DIF could result. For example, the MagicId Item 24 ("If reincarnation were true, it would explain some unusual experiences I have had") may be interpreted differently by someone who practices a religion that includes reincarnation such as Buddhism, as compared with someone who practices a religion that does not include reincarnation such as Christianity.

In addition to magical ideation and perceptual aberration, it is possible that there are different rates of social anhedonia or DIF on the RSAS related to cultural orientation toward independent versus interdependent self-concepts. Western cultures tend to emphasize independent self-construal, while Eastern cultures tend to emphasize interdependent self-construal (Markus & Kitayama, 1991). For example, Item 28 on the RSAS ("I'm much too independent to really get involved with other people") may be interpreted differently in cultures that value independence than in cultures that value interdependence. Along with self-construal, cultural norms related to emotional expression may influence responses to the questions on the RSAS directly related to emotional expressivity (e.g., Item 13, "My emotional responses seem very different from those of other people" and Item 21, "People are usually better off if they stay aloof from emotional involvements with most others" on the full RSAS). There is a large body of work showing that, in general, Asian individuals are less emotionally expressive than White or Hispanic individuals (e.g., Soto, Levenson, & Ebling, 2005).

Another reason to examine DIF is that nearly all research involving the WSS has included some ethnic minorities. One of the most common ways in which the WSS are used is to recruits participants with very high scores on the WSS as "positive schizotypy" and "negative schizotypy" groups. Positive schizotypy represents attenuated positive symptoms of schizophrenia including delusions and hallucinations and is typically measured with the MagicId and PerAbs. Negative schizotypy represents attenuated negative symptoms of schizophrenia such as anhedonia and apathy, and is typically measured with the RSAS. These groups are then followed over time or compared with appropriately matched comparison groups to determine deficits associated with schizotypy. This is referred to as the psychometric high-risk paradigm (Lenzenweger, 1994). Testing the DIF of WSS scores is important because if the scales have significant DIF, then scores in one group may not have the same latent level of schizotypy as scores in another group. As a result, participants may be incorrectly assigned to a schizotypy group or incorrectly left out of a schizotypy group due to differences in appropriate cut-scores between groups.

In addition to nonequivalent psychometric properties between groups, one weakness in the WSS is that they may not have a unidimensional factor structure as originally designed. Some previous work has found that one-factor models of the separate schizotypy scales, particularly the RSAS, do not fit the data well (Reise et al., 2011). Although the fit cannot be compared with the full scales directly, a brief version of the scales including only well-performing items could be more likely to meet the basic assumption of IRT analyses that the scales are unidimensional. This is especially important in the current research because unidimensionality is a fundamental assumption of IRT, which this study uses to examine DIF (Osterlind & Everson, 2009).

Along with DIF among ethnic groups, there is some evidence that the full versions of the WSS have DIF between sexes (Winterstein, Ackerman, et al., 2011). This work found that 8 of the 30 MagicId items, 2 of the 35 PerAb items, and 12 of the 40 RSAS items had DIF between sexes, without a clear pattern as to whether the items overestimated the latent level of schizotypy in men or women. Moreover, several studies have found difference in levels of positive schizotypy between sexes, with most finding higher reported rates in women (Fossati, Raine, Carretta, Leonardi, & Maffei, 2003; Karcher, Slutske, Kerns, Piasecki, & Martin, 2014; Mata, Mataix-Cols, & Peralta, 2005; Rawlings, Claridge, & Freeman, 2001). Since only one study, to our knowledge, has examined the DIF of the WSS between sexes, it would be useful to replicate these findings in the current sample.

The first goal of the current research was to examine whether the full and brief versions of the WSS met the IRT assumption of being unidimensional. The second goal of the current research is to examine the DIF of the WSS full and brief scores across White, Hispanic, Asian, and Multiethnic samples. The third goal is to examine the measurement invariance and DIF of the scales between men and women. The fourth and final goal of the current research was to examine whether the full and brief versions of the scales were strongly correlated with each other.

Method

Participants

Participants were 2,312 undergraduates from a large public Pacific University and a large public West Coast University who participated in exchange for course credit or extra credit. Participants had the option to complete an alternate assignment for course credit or extra credit. The study was approved by the University of Hawai'i at Manoa and the University of California, Irvine Institutional Review Boards. They were 45.5% Asian, 18.0% White, 23.4% Multiracial, and 13.1% Hispanic. Fifty-six African American, 28 Pacific Islander, and 3 Native Americans participated in the study but were not included in any analyses because their small sample sizes did not allow for IRT or DIF analyses. Eighty-two participants with one or more missing data point were excluded list wise. Age ranged from 17 to 62 years ($M = 20.59$, $SD = 3.75$). There was a significant difference across ethnic groups with respect to age, $F(3, 2{,}309) = 23.311$, $p < .001$, $\eta^2 = 0.03$. However, this effect is very small and is likely only statistically significant due to the large sample size. They were 50.8% female, 48.9% male, and 0.3% declined to specify their sex.

Measures

Participants completed the WSS. The MagicId (Eckbald & Chapman, 1983) is a 30-item true–false questionnaire designed to measure "beliefs in forms of causation that by conventional standards are invalid" (Eckbald & Chapman, 1983, p. 215). In the current research, the full and brief MagicIds had Cronbach's αs of .88 and .79, respectively.

The PerAb (Chapman et al., 1978) is a 35-item true–false scale that includes 28 items designed to measure schizophrenic-like distortions in perception of one's own body and 7 items for other perceptual distortions. In the current research, the full and brief PerAbs had Cronbach's αs of .89 and .88, respectively. The RSAS (Eckbald et al., 1982) is a 40-item true–false questionnaire designed to measure lack of relationships and lack of pleasure from relationships. In the current research, the full and brief RSASs had Cronbach's αs of .81 and .73, respectively. The RSAS has been found to predict future development of schizophrenia spectrum disorders (Gooding et al., 2005; Kwapil, 1998). The MagicId, PerAb, and RSAS have considerable support for their scores' reliability and validity in a number of different populations (for a review, see Edell, 1995). The brief versions of these scales have 15 items for each scale (Winterstein, Ackerman, et al., 2011; Winterstein, Silvia, et al., 2011). Although the WSS also include the Physical Anhedonia Scale (Chapman, Edell, & Chapman, 1980), we focused our analyses on the MagicId, PerAb, and RSAS because they are the best predictors of future development of schizophrenia spectrum disorders and are typically used to create groups in psychometric high-risk research (Chapman et al., 1994; Kwapil, 1998; Lenzenweger, 1994). To differentiate between the full and brief versions of the scales, the brief versions of the scales will be abbreviated as B-MagicId, B-PerAb, and B-RSAS.

Procedure

Participants completed the WSS online as part of a larger study. The entire study took approximately 60 minutes.

Data Analytic Strategy

Data analyses were conducted with M*plus* version 7.31 (Muthén & Muthén, 1998-2016). First, we tested whether the full versions and brief versions of the scales met the primary assumption of IRT analyses that the scale is unidimensional. A series of one-factor models was specified for each scale including all participants. We planned to examine the fit of both the full and brief versions of the scales and focus the DIF analyses on scales that met this basic assumption of IRT. Following Hu and Bentler (1998), we used the following cutoffs for good fit: (a) comparative fit index (CFI) > .95, (b) Tucker–Lewis index (TLI), and (c) root mean squared error of approximations (RMSEA) < .05.

Second, global tests of measurement invariance were conducted within a multiple indicator multiple causes (MIMIC) model. We first specified a two-parameter IRT model. We chose a two-parameter model because it allows for the estimation of both the difficulty (β) and discrimination (α) parameters. The latent factor and all of the individual items were then regressed on dummy-coded ethnicity variables and the dummy-coded sex variable. We compared the fit of this model with the fit of a model in which all these regression weights were constrained to zero. If the constrained model fit significantly worse than the model in which the regression weights were freely estimated, then we would conclude that there is significant DIF and examine the items individually. Model fit was compared with a chi-square difference test and with Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and Sample-Size Adjusted BIC (SABIC).

In the second set of analyses, we examined MIMIC models for each of the scales to examine whether the difficulty parameter had DIF between groups. We tested for both uniform DIF (i.e., differences in difficulty parameter between groups) and nonuniform DIF (i.e., differences between groups that varies as a function of the level of the latent trait). To test for uniform DIF, we regressed the dummy-coded ethnicity and sex variables on the latent factor and on each of the items individually. Following Woods and Grimm (2011), we tested nonuniform DIF by creating ethnicity by latent trait interaction and sex by latent trait interactions with the XWITH command in M*plus*. We then regressed the interaction terms on each item in the same analysis in which the ethnicity and sex variables were regressed on the latent variable and the individual items. A significant result for each individual item suggests that the item has DIF between the group regressed on and the reference group. Since these scales were developed in mostly White samples, the White group was used as the reference group for all analyses. The reference group with respect to sex was female. Thus, the reported factor loadings and thresholds can be interpreted as the parameters for White females, and the beta weights represent the difference between the group and the reference group. To correct for multiple comparisons in the DIF analysis, we divided the alpha by four since there were three ethnicity dummy-coded variables and one sex dummy-coded variable. In addition to examining the statistical significance of these analyses, we calculated Educational Testing Service (ETS) Δ as a measure of effect size (Monahan, McHorney, Stump, & Perkins, 2007). ETS Δ is a linear transformation of β in which it is multiplied by −2.35. ETS Δs with absolute values between 0 and 1 are considered to have negligible DIF, items with ETS Δs with absolute values between 1 and 1.5 are considered to have slight to moderate DIF, and items with absolute values above 1.5 are considered to have moderate to large DIF. The recommended estimator for categorical variables in M*plus* is weighted least squares mean and variance adjusted, but this estimator cannot be used with the XWITH command to estimate interactions with latent variables as necessary for nonuniform DIF. Maximum likelihood estimation produced a saddle point, which prevents reliable parameter estimates. Thus, we used MLF estimation for all MIMIC analyses.

Table 1. Fit Statistics for a One-Factor Model of the Full and Brief Versions of the Scales in all Participants.

Model	χ^2	*df*	RMSEA	90% CI	TLI	CFI
Magical Ideation Scale						
Full scale	3723.60	405	0.059	[0.057, 0.060]	0.821	0.833
Brief scale	307.232	90	0.032	[0.028, 0.035]	0.971	0.975
Perceptual Aberration Scale						
Full scale	3404.11	560	0.042	[0.041, 0.044]	0.930	0.930
Brief scale	306.350	90	0.029	[0.025, 0.032]	0.986	0.988
Social Anhedonia Scale						
Full scale	17582.34	740	0.089	[0.088, 0.090]	0.421	0.450
Brief scale	3316.529	90	0.111	[0.107, 0.114]	0.651	0.701

Note. df = degrees of freedom; RMSEA = root mean squared error of approximation; TLI = Tucker–Lewis index; CFI = comparative fit index.

Table 2. Global Measurement Invariance Analyses for the Brief Magical Ideation and Perceptual Aberration Scale.

Model	χ^2	Parameters	AIC	BIC	SABIC	χ^2_{diff} (*df*)	*p*
Brief Magical Ideation Scale							
Freely estimated	31128.258	150	31428.26	32291.49	31814.92		
Fixed to zero	31403.082	30	31463.08	31635.73	31540.41	274.824 (120)	<.001
Brief Perceptual Aberration Scale							
Freely estimated	19105.864	150	19405.86	20296.52	19819.92		
Fixed to zero	19391.96	30	19451.96	1960.092	19534.77	286.096 (120)	<.001

Note. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; SABIC = Sample Size Adjusted Bayesian Information Criterion; *df* = degrees of freedom. Freely estimate models include parameter estimates for ethnicity and sex variables regression on the latent factor and all items. In the fixed to zero models, these parameters are constrained to zero.

In the event that a one-factor model did not fit the data well for either the full or the brief versions of a scale, we planned to conduct MIMIC analyses through an exploratory structural equation modeling (ESEM) framework. In this framework, the factor structure of the items is identified and the dummy-coded demographic variables are regressed on the latent factors and items in the same manner as the MIMIC models within a confirmatory factor analysis (CFA) framework as described above. Finally, we examined the zero-order correlations among the full and brief versions of the scales for each ethnic group.

Results

As can be seen in Table 1, the single group model did not fit the data well for any of the full versions of the scales. The models fit the data especially poorly for the RSAS. This finding for the RSAS is consistent with previous single group CFA research that has also found challenges in fitting an IRT model to the RSAS (Reise et al., 2011). Since the models did not approach good fit for any of the full versions of the scales, we did not test DIF in these scales with the MIMIC models. Instead, we moved on to the brief versions of the scales.

Next, we examined the fit of a single Group 2 PL IRT model for the brief scales. As shown in Table 1, the model for the both the B-PerAb and the B-MagicId fit the data well, but the B-RSAS did not. Since the B-PerAb and B-MagicId fit the data well, we conducted a test of the global measurement invariance of these two scales. As shown in Table 2, there was mixed evidence for whether the B-PerAb and B-MagicId had measurement invariance among groups. For both scales, the model in which the regression parameters of all the items on the ethnicity and sex variables were constrained to zero fit the data worse than the model with these parameters freely estimated according to the χ^2 difference test. However, the BIC and SABIC scores, which emphasize parsimony, indicate that the model with the parameters fixed to zero for both the B-MagicId and B-PerAb fits the data best, while the AIC suggests a better fit for the freely estimated models. Since the models did not unequivocally display measurement invariance, we used a MIMIC model to examine whether the individual items displayed DIF.

Next, we examined group differences in latent means in a model that assumes no DIF between groups (i.e., the regression weights constrained to zero). For the B-MagicId, Asian ($\beta = 0.17$, $SE = 0.03$, $p < .001$) and Multiethnic participants ($\beta = 0.10$, $SE = 0.03$, $p = .001$) had higher latent means than White participants, but there was not a significant effect at the $\alpha = .0125$ level for either Hispanic ($\beta = 0.07$, $SE = 0.03$, $p = .031$) or Male variables ($\beta = 0.04$,

Table 3. Uniform Differential Item Functioning for the Brief Magical Ideation Scale by Ethnicity and Sex.

		IRT		Asian		Multiethnic		Hispanic		Male	
Item	Percent endorsed	λ (*SE*)	τ (*SE*)	β (*SE*)	ETS	β (*SE*)	ETS	β (*SE*)	ETS	β (*SE*)	ETS
MagicId				0.74 (0.26)*		0.40 (0.27)		0.86 (0.31)*		−0.04 (0.18)	
1	19.8	0.62 (0.13)	0.87 (0.11)	−1.02 (0.43)	2.40	−0.23 (0.46)	0.55	−0.84 (0.55)	1.97	0.48 (0.34)	−1.13
2	15.1	0.67 (0.11)	1.01 (0.10)	−0.87 (0.34)*	2.06	−0.30 (0.41)	0.71	−0.41 (0.38)	0.97	0.43 (0.30)	−1.00
3	13.2	0.99 (0.15)	1.45 (0.14)	0.003 (0.61)	−0.01	0.85 (0.62)	−1.99	−0.53 (0.73)	1.24	1.03 (0.42)	−2.41
4	12.2	0.68 (0.15)	1.31 (0.16)	0.17 (0.56)	−0.41	0.33 (0.61)	−0.79	−1.44 (1.03)	3.39	0.02 (0.40)	−0.05
5	26.4	0.66 (0.11)	0.65 (0.08)	−0.38 (0.25)	0.90	−0.12 (0.28)	0.28	−0.46 (0.230)	1.09	0.14 (0.19)	−0.34
6	20.7	0.77 (0.15)	1.03 (0.13)	0.07 (0.45)	−0.16	0.07 (0.49)	−0.17	0.11 (0.51)	−0.25	0.15 (0.31)	−0.36
7	25.5	0.56 (0.11)	0.63 (0.09)	−0.52 (0.29)	1.21	−0.21 (0.32)	0.49	−0.77 (0.40)	1.81	−0.02 (0.24)	0.06
8	23.0	0.56 (0.11)	0.92 (0.09)	−0.08 (0.28)	0.19	−0.31 (0.33)	0.74	−0.96 (0.44)*	2.27	0.71 (0.23)*	−1.66
9	32.2	0.48 (0.09)	0.66 (0.07)	0.10 (0.22)	−0.23	0.15 (0.25)	−0.35	−0.43 (0.29)	1.01	0.39 (0.17)	−0.92
10	42.3	0.50 (0.10)	0.25 (0.06)	−0.32 (0.20)	0.76	−0.08 (0.29)	0.19	−0.40 (0.230)	0.93	0.09 (0.17)	−0.20
11	28.1	0.08 (0.08)	0.50 (0.07)	0.19 (0.14)	−0.45	0.23 (0.16)	−0.55	−0.24 (0.21)	0.56	−0.43 (0.12)*	1.00
12	13.3	0.83 (0.12)	1.35 (0.12)	−0.47 (0.37)	1.11	−0.78 (0.52)	1.84	−0.99 (0.51)	2.33	2.17 (0.49)*	−5.09
13	17.8	0.76 (0.14)	1.24 (0.12)	0.49 (0.39)	−1.16	0.45 (0.44)	−1.05	−1.09 (0.69)	2.56	0.58 (0.28)	−1.36
14	24.6	0.30 (0.10)	0.62 (0.08)	−0.19 (0.21)	0.45	0.22 (0.22)	−0.52	−0.35 (0.29)	0.82	−0.48 (0.17)*	1.12
15	15.4	0.72 (0.03)	1.04 (0.04)	−0.24 (6.13)	0.57	−0.41 (0.76)	0.96	−0.54 (0.47)	−1.26	0.66 (0.55)	−1.54

Note. Percent endorsed = the percentage of the total sample endorsing the item; IRT = item response theory; MagicId = Magical Ideation Scale; *SE* = standard error; λ = factor loading; τ = thresholds; β = regression weight; ETS Δ = effect size. Reference group is White, Female.
**p* < .0125.

SE = 0.03, *p* = .175). For the B-PerAb, Asian participants had higher latent means than White participants (β = 0.22, *SE* = 0.06, *p* = .001) and male participants had higher latent means than female participants (β = 0.18, *SE* = 0.05, *p* < .001). There was not a significant effect for Multiethnic participants (β = 0.08, *SE* = 0.07, *p* = .245) or Hispanic participants (β = 0.02, *SE* = 0.08, *p* = .780). As can be seen in Table 3, in the freely estimated model with the ethnicity and sex variables regressed on the latent factor and each individual item, Asian and Hispanic participants had higher latent means than White participants, but there was not a statistically significant effect for either Multiethnic participants or male participants. As shown in Table 5, there were no statistically significant differences with respect to ethnicity or sex for latent PerAb values when DIF is freely estimated. This suggests that the DIF present in the items may be significant enough to change the interpretation of group differences.

With respect to the individual items, none of the B-MagicId items displayed uniform DIF with respect to ethnicity when compared with the White group (see Table 3). However, four of the items displayed uniform DIF with respect to sex. Of the significant items, two items had ETS Δs greater than 1.5 in absolute value, indicating moderate to severe DIF. In both items, male participants scored higher than would be expected based on their latent level of MagicId. In contrast, two items had ETS Δs between 1 and 1.5 indicating mild to moderate DIF. In both of these, male participants scored lower than would be expected based on their latent levels of MagicId. As can be seen in Table 4, one item had mild to moderate nonuniform DIF in comparing Asian with White participants, one item had statically significant but negligible nonuniform DIF in comparing Multiethnic participants to White participants, and one item had slight to moderate nonuniform DIF in comparing male with female participants.

As can be seen in Table 5, none of the B-PerAb items displayed uniform DIF with respect to ethnicity, but one item displayed statistically significant, but negligible uniform DIF between sexes. No items displayed statistically significant nonuniform DIF for either ethnicity or sex (see Table 6). Previous work has suggested that scales with over 20% of the items with DIF are problematic (Byrne, Shavelson, & Muthén, 1989). Thus, the finding that one item on the B-PerAb has DIF is not necessarily problematic for the entire scale of 15 items.

Since the 2 PL CFA of neither the full nor the brief versions of the RSAS fit the data well, we conducted an ESEM analysis to examine the DIF of the items of the B-RSAS. Prior to the ESEM, we conducted an item-level EFA on the B-RSAS. The slope of the scree plot approach zero at two factors, suggesting that two factors should be extracted in the ESEM. A two-factor model with the dummy-coded ethnicity and sex variables regressed on Factor 1, Factor 2, and each of the items fit the data well, χ^2 = 387.534 (76), *p* < .001, RMSEA = 0.038, 90% CI [0.034, 0.042], CFI = 0.970, TLI = 0.935. Like in the MIMIC models presented for the B-PerAb and B-MagicId, the dummy-coded ethnicity and sex variables were regressed on the latent factors and each individual item in the ESEM. As can be seen in Table 7, only one item was statistically significant at the *p* < .0125 level. This item showed statistically significant but negligible DIF between men and women.

Table 4. Nonuniform Differential Item Functioning for the Brief Magical Ideation Scale by Ethnicity and Sex.

	Asian		Multiethnic		Hispanic		Male	
Item	β (*SE*)	ETS Δ	β (*SE*)	ETS Δ	β (*SE*)	ETS Δ	β (*SE*)	ETS Δ
1	0.54 (0.25)	−1.27	0.39 (0.27)	−0.91	0.38 (0.27)	−0.89	−0.32 (0.13)	0.75
2	0.24 (0.19)	−0.55	−0.01 (0.21)	0.03	−0.08 (0.20)	0.18	−0.27 (0.14)	0.64
3	−0.20 (0.28)	0.48	−0.43 (0.29)	1.02	−0.16 (0.31)	0.37	−0.40 (0.17)	0.94
4	0.04 (0.25)	−0.09	−0.12 (0.28)	0.27	0.39 (0.40)	−0.92	0.08 (0.17)	−0.18
5	−0.04 (0.15)	0.10	−0.01 (0.18)	0.02	−0.17 (0.17)	0.39	−0.08 (0.10)	0.18
6	−0.04 (0.24)	0.08	−0.10 (0.26)	0.23	−0.23 (0.27)	0.55	−0.01 (0.14)	0.02
7	0.14 (0.15)	−0.33	0.12 (0.18)	−0.32	0.08 (0.20)	−0.19	0.03 (0.11)	−0.07
8	0.09 (0.15)	−0.20	0.08 (0.17)	−0.18	0.28 (0.21)	−0.65	−0.10 (0.11)	0.23
9	0.16 (0.13)	−0.38	0.10 (0.14)	−0.24	0.13 (0.16)	−0.30	−0.18 (0.09)	0.27
10	0.03 (0.12)	−0.06	0.36 (0.18)	−0.85	0.20 (0.17)	−0.47	−0.01 (0.09)	0.02
11	−0.05 (0.11)	0.11	0.01 (0.12)	−0.01	−0.01 (0.13)	0.01	0.14 (0.07)	−0.34
12	0.08 (0.18)	−0.19	0.28 (0.25)	−0.67	0.29 (0.25)	−0.68	−0.77 (0.23)*	1.82
13	−0.11 (0.21)	0.26	−0.15 (0.24)	0.34	0.22 (0.29)	−0.51	−0.17 (0.13)	0.40
14	0.09 (0.12)	−0.22	0.06 (0.14)	−0.14	0.08 (0.16)	−0.18	0.19 (0.09)	−0.45
15	0.20 (0.24)	−0.46	0.064 (0.24)	−0.15	−0.15 (0.22)	0.36	0.08 (0.18)	−0.18

Note. SE = standard error; percent endorsed = the percentage of the total sample endorsing the item; β = regression weight; ETS Δ = effect size. Reference group is White, Female.
**p* < .0125.

Table 5. Uniform Differential Item Functioning for the Brief Perceptual Aberration Scale by Ethnicity and Sex.

		IRT		Asian		Multiethnic		Hispanic		Male	
Item	Percent endorsed	λ (*SE*)	τ (*SE*)	β (*SE*)	ETS Δ	β (*SE*)	ETS Δ	β (*SE*)	ETS Δ	β (*SE*)	ETS Δ
PerAb				0.44 (0.23)		−0.26 (0.25)		0.67 (0.30)		0.50 (0.19)	
1	6.9	0.52 (0.14)	0.81 (0.02)	−1.83 (0.77)	4.31	−1.88 (1.18)	4.43	−1.80 (1.20)	4.23	−0.72 (0.66)	1.69
2	8.3	0.88 (0.18)	1.52 (0.26)	0.56 (1.15)	−1.31	−0.35 (1.49)	0.81	0.60 (1.42)	−1.41	−0.77 (0.81)	1.82
3	9.1	0.91 (0.18)	1.65 (0.24)	1.01 (1.03)	−2.37	0.62 (1.20)	−1.45	0.06 (1.29)	−0.14	−0.27 (0.55)	0.64
4	7.0	1.18 (0.19)	2.19 (0.25)	1.89 (1.58)	−4.44	3.18 (1.58)	−7.47	0.28 (1.91)	−0.66	0.86 (0.65)	−2.02
5	7.2	0.72 (0.16)	1.33 (0.22)	−1.28 (0.90)	3.00	−0.68 (1.14)	1.60	−0.64 (1.06)	1.50	−0.64 (0.71)	1.51
6	18.8	0.56 (0.10)	1.03 (0.09)	−0.35 (0.30)	0.83	−0.06 (0.30)	0.15	−0.58 (0.43)	1.36	0.55 (0.21)*	−1.29
7	7.2	0.92 (0.19)	1.60 (0.26)	0.53 (1.10)	−1.24	−0.55 (1.46)	1.29	0.89 (1.32)	−2.10	−0.39 (0.73)	0.92
8	7.2	1.23 (0.21)	2.15 (0.29)	2.04 (2.50)	−4.80	4.81 (2.44)	−11.30	3.54 (2.62)	−8.32	0.03 (0.91)	−0.06
9	7.4	0.87 (0.10)	1.58 (0.23)	−0.46 (0.90)	1.09	1.05 (0.89)	−2.46	−0.18 (1.24)	0.41	−0.49 (0.57)	1.14
10	9.9	1.02 (0.10)	1.86 (0.23)	2.26 (1.16)	−5.30	1.51 (1.33)	−3.55	0.33 (1.55)	−0.78	−0.02 (0.51)	0.05
11	10.9	0.74 (0.14)	1.50 (0.18)	0.46 (0.63)	−1.07	−0.40 (0.89)	0.93	0.81 (0.68)	−1.89	0.42 (0.42)	−0.98
12	6.9	0.88 (0.10)	1.66 (0.24)	0.18 (1.03)	−0.43	0.73 (1.11)	−1.72	1.62 (1.12)	−3.81	−0.84 (0.63)	1.98
13	14.0	0.90 (0.14)	1.44 (0.15)	0.83 (0.51)	−1.95	0.96 (0.58)	−2.24	0.14 (0.68)	−0.33	−0.16 (0.32)	0.38
14	7.6	0.98 (0.10)	1.83 (0.28)	−0.07 (1.39)	0.16	1.82 (1.44)	−4.28	0.39 (1.67)	−0.93	0.54 (0.90)	−1.28
15	12.7	0.81 (0.02)	1.36 (0.09)	−0.42 (0.64)	0.98	−1.56 (0.88)	3.68	0.32 (0.75)	−0.74	0.36 (0.48)	−0.84

Note. IRT = item response theory; percent endorsed = the percentage of the total sample endorsing the item; PerAb = Perceptual Aberration Scale;
λ = factor loading; τ = thresholds; β = regression weight; ETS Δ = effect size. Reference group is White, Female.
**p* < .0125.

As mentioned, the one-factor model did not fit the data well for any of the full versions of the scales. Thus, we tested DIF in these scales with MIMIC models within an ESEM framework as described above for the. First, we conducted EFAs for the full versions of all three scales. The slope of the scree plot approached zero at two factors for the MagicId and PerAb, but at three factors for the RSAS. The ESEM for the full MagicId MIMIC model fit the data well, $\chi^2 = 1758.88$ (526), $p < .001$, RMSEA = 0.029, 90% CI [0.028, 0.031], CFI = 0.960, TLI = 0.971. As can be seen in Supplemental Table 1, one item displayed statistically significant, but negligible DIF between Hispanic and White and between men and women. One additional item had statistically significant but negligible DIF between men and women.

Table 6. Nonuniform Differential Item Functioning for the Brief Perceptual Aberration Scale by Ethnicity and Sex.

	Asian		Multiethnic		Hispanic		Male	
Item	β (*SE*)	ETS Δ	β (*SE*)	ETS Δ	β (*SE*)	ETS Δ	β (*SE*)	ETS Δ
1	0.64 (0.44)	−1.50	0.97 (0.64)	−2.28	0.03 (0.43)	−0.06	0.15 (0.28)	−0.35
2	−0.24 (0.44)	0.57	0.45 (0.60)	−1.06	−0.58 (0.53)	1.37	0.09 (0.31)	−0.21
3	−0.27 (0.43)	0.63	0.30 (0.50)	−0.71	−0.35 (0.51)	0.82	−0.14 (0.23)	0.32
4	−0.63 (0.66)	1.48	−0.80 (0.68)	1.87	−0.65 (0.73)	1.52	−0.43 (0.28)	1.02
5	0.41 (0.37)	−0.95	0.52 (0.49)	−1.21	−0.23 (0.43)	0.54	0.02 (0.27)	−0.05
6	0.26 (0.18)	−0.60	0.05 (0.18)	−0.12	0.27 (0.24)	−0.64	−0.25 (0.13)	0.58
7	−0.33 (0.47)	0.78	0.34 (0.59)	−0.81	−0.89 (0.59)	2.09	−0.02 (0.28)	0.05
8	−0.84 (0.95)	1.96	−1.48 (0.97)	3.48	−1.82 (1.03)	4.27	−0.18 (0.34)	0.42
9	0.03 (0.39)	−0.08	−0.29 (0.41)	0.69	−0.50 (0.50)	1.18	0.03 (0.22)	−0.08
10	−0.72 (0.51)	1.70	−0.17 (0.53)	0.41	−0.20 (0.59)	0.46	−0.12 (0.21)	0.28
11	−0.08 (0.28)	0.18	0.48 (0.41)	−1.13	−0.43 (0.32)	1.01	−0.08 (0.18)	0.18
12	−0.14 (0.42)	0.33	0.07 (0.48)	−0.15	−1.03 (0.51)	2.43	0.17 (0.24)	−0.40
13	−0.37 (0.28)	0.87	−0.10 (0.30)	0.23	−0.38 (0.33)	0.90	−0.06 (0.15)	0.15
14	−0.05 (0.52)	0.11	−0.21 (0.58)	0.49	−0.55 (0.61)	1.29	−0.18 (0.33)	0.43
15	0.41 (0.38)	−0.97	0.47 (0.48)	−1.09	0.37 (0.50)	−0.86	−0.19 (0.28)	0.44

Note. Percent endorsed = the percentage of the total sample endorsing the item; β = regression weight; ETS Δ = effect size; *SE* = standard error. Reference group is White, Female.
**p* < .0125.

Table 7. Exploratory Structural Equation Modeling of the Brief Social Anhedonia Scale With Invariance Estimates by Ethnicity and Sex.

		Factor loadings		Asian		Multiethnic		Hispanic		Male	
Item	Percent endorsed	F1 (*SE*)	F2 (*SE*)	B (*SE*)	ETS Δ	B (*SE*)	ETS Δ	B (*SE*)	ETS Δ	B (*SE*)	ETS Δ
F1				0.17 (0.14)		0.19 (0.15)		0.27 (0.17)		0.02 (0.10)	
F2				−0.01 (0.10)		0.17 (0.11)		−0.27 (0.14)		0.09 (0.07)	
1	6.9	0.57 (0.03)	0.24 (0.04)	−0.02 (0.10)	0.05	−0.08 (0.11)	0.20	−0.06 (0.13)	0.14	−0.09 (0.07)	0.22
2	8.3	0.61 (0.03)	0.06 (0.04)	0.04 (0.10)	−0.09	−0.04 (0.12)	0.10	0.02 (0.13)	−0.05	−0.03 (0.07)	0.07
3	9.1	0.61 (0.02)	0.01 (0.01)	0.20 (0.10)	−0.47	0.09 (0.11)	−0.20	0.10 (0.13)	−0.24	−0.07 (0.07)	0.17
4	7.0	−0.02 (0.03)	0.87 (0.02)	−0.01 (0.09)	0.00	0.05 (0.10)	−0.11	−0.08 (0.12)	0.18	0.02 (0.07)	−0.04
5	7.2	0.69 (0.02)	−0.05 (0.04)	0.20 (0.11)	−0.47	0.08 (0.12)	−0.19	0.20 (0.13)	−0.46	−0.08 (0.08)	0.19
6	18.8	0.64 (0.02)	−0.05 (0.03)	0.07 (0.10)	−0.17	−0.03 (0.11)	0.06	−0.02 (0.13)	0.05	−0.04 (0.07)	0.09
7	7.2	0.71 (0.02)	0.06 (0.04)	0.14 (0.11)	−0.32	−0.10 (0.12)	0.23	0.04 (0.14)	−0.10	−0.03 (0.08)	0.06
8	7.2	0.68 (0.02)	−0.07 (0.04)	0.25 (0.12)	−0.58	0.03 (0.12)	−0.06	0.04 (0.14)	−0.10	0.02 (0.08)	−0.05
9	7.4	−0.03 (0.03)	0.93 (0.02)	0.03 (0.09)	−0.07	0.01 (0.10)	−0.02	0.02 (0.14)	−0.04	−0.05 (0.07)	0.12
10	9.9	0.59 (0.02)	−0.13 (0.03)	0.34 (0.10)*	−0.79	0.17 (0.11)	−0.40	0.24 (0.13)	−0.55	0.02 (0.07)	−0.06
11	10.9	0.17 (0.03)	0.58 (0.03)	0.04 (0.08)	−0.09	−0.16 (0.09)	0.38	0.02 (0.11)	−0.05	−0.20 (0.06)*	0.48
12	6.9	0.16 (0.03)	0.75 (0.02)	−0.05 (0.08)	0.12	−0.08 (0.09)	0.19	−0.15 (0.12)	0.36	−0.03 (0.06)	0.08
13	14.0	0.01 (0.01)	0.72 (0.02)	0.09 (0.08)	−0.21	0.13 (0.09)	−0.31	0.26 (0.11)	−0.61	−0.15 (0.06)	0.35
14	7.6	0.01 (0.03)	0.80 (0.02)	−0.09 (0.09)	0.21	−0.14 (0.11)	0.32	−0.26 (0.15)	0.61	0.16 (0.07)	−0.37
15	12.7	0.65 (0.03)	0.18 (0.04)	0.55 (0.36)	−1.29	0.89 (0.42)	−2.04	1.62 (0.61)	−3.80	−0.44 (0.28)	−3.80

Note. Percent endorsed = the percentage of the total sample endorsing the item; F1 = Factor 1; F2 = Factor 2; *SE* = standard error; β = regression weight, ETS Δ = effect size. Reference group is White, Female.
**p* < .0125.

The ESEM for the full PerAb MIMIC model fit the data well, χ^2 = 1220.294 (376), $p < .001$, RMSEA = 0.031, 90% CI [0.029, 0.033], CFI = 0.956, TLI = 0.935. As can be seen in Supplemental Table 2, none of the individual items displayed DIF with respect to ethnicity or sex. The three-factor ESEM for the full RSAS MIMIC model also fit the data well, χ^2 = 1735.50 (663), $p < .001$, RMSEA = 0.024, 90% CI [0.023, 0.036], CFI = 0.965, TLI = 0.951. As can be seen in Supplemental Table 3, none of individual items had DIF in this model.

Finally, we examined the Pearson correlations between the full and brief scales. Overall, the full PerAb was correlated r = .92 with the brief version, the full MagicId was correlated r = .92 with the brief version, and the full RSAS was correlated r = .88 with the brief version. When conducted separately in each ethic group, the correlations were very similar across groups, ranging from .91 to .94 for PerAb, .90 to .92 for MagicId, and .86 to .89 for RSAS.

Discussion

The results of the current research present a nuanced picture of the measurement equivalence of WSS scores in White, Asian, Multiethnic, and Hispanic participants. The first goal of the current study was to test whether the full and brief versions of the WSS were unidimensional. Consistent with previous research, the one-factor models of the full versions of the scales did not fit the data well according to any of the fit statistics, and the fit was especially poor for the full version of the RSAS (Reise et al., 2011). Although the fit cannot be compared directly, the B-PerAb and B-MagicId exceeded conventional standards for good fit. The B-RSAS has closer to standards of good fit than the full version, but did not meet conventional standards. This discrepancy in fit between the full version of the scales and the brief versions suggests that the brief versions may be more appropriate to test DIF than the full versions. At the same time, CFA models tend to find better fit for simpler models, which may explain why item-level analyses with long scales such as the full versions of the these scales provide poor model fit. This finding may be a limitation of CFA rather than a limitation of the scales themselves.

The second goal of the current research was to examine the global measurement invariance and DIF on the scales in White, Asian, Hispanic, and Multiethnic participants. In the current research, for the B-PerAb and B-MagicId, a model in which all the regression parameters for the DIF analyses were fixed to zero fit significantly worse than the model in which these parameters were freely estimated according to the χ^2 difference tests, but fit better than the freely estimated model according to BIC. Since χ^2 difference tests tend to have high Type I error rates, the BIC results (i.e., that the models are invariant across these groups) may be more accurate. At the same time, the results comparing the latent means for groups were different in models that freely estimated these parameters than in models in which they were constrained to zero. This suggests the presence of DIF between groups, and differences in item difficulty parameters may be important for understanding differences between groups.

This finding is somewhat consistent with previous research that suggested the scales had DIF between ethnic groups, specifically between White and African American participants (Winterstein, Ackerman, et al., 2011). The DIF results of the current research are broadly consistent with results of previous work for the full (Winterstein, Ackerman, et al., 2011) and brief versions of the scales (Winterstein, Silvia, et al., 2011). Similar to Winterstein, Ackerman, et al. (2011), we found that the full versions PerAb and MagicId performed fairly well on DIF analyses within the ESEM framework. This is remarkable given that the DIF analyses were conducted with a different set of ethnic groups. The results of the current study provide further support for the psychometric properties of the brief versions of the scales. In developing the brief versions of the scales, Winterstein, Silvia, et al. (2011) removed items with DIF between sexes and ethnicities. This strategy of removing items with DIF appears successful in limiting DIF, even when including ethnic groups such as Hispanic, Asian, and Multiethnic, which were not part of the original sample.

The third goal of the current research was to examine the measurement invariance and DIF of these scales between men and women. The results suggest that DIF between sexes may only be an issue for the B-MagicId, in which four items (20% of the total items) displayed DIF with respect to sex. However, two of these items were biased toward males and two were biased against males. As a result, there may not be significant bias in the scale scores between men and women.

Overall, the results of the current research suggest that some of the items of these scales have DIF and lack clear measurement equivalence. As mentioned, schizotypy researchers generally treat these scales as either taxonic or continuous (Kwapil et al., 2008; Lenzenweger, 1994). Although resolving this issue is beyond the scope of this study, the current results have implications for both approaches. We tested the DIF of difficulty parameter of the scales, which reflects whether the summed scale scores represent the same latent level of the schizotypy symptoms among groups. When treating the variables as continuous, mean comparison of these ethnic groups should be interpreted with caution because the scores on the scales may not represent the same level of latent schizotypy in each group (Chen, 2008). At the same time, the limitations of χ^2 difference testing in comparing model fit, especially for measurement invariance, are well-known (Cheung & Rensvold, 2002). In particular, χ^2 difference tests are likely to underestimate measurement invariance (i.e., to suggest that models are not invariant, when in fact they are).

One major finding of the current research is that the full versions of all three scales and the B-RSAS performed well on the MIMIC analyses within an ESEM framework. This is in contrast to previous work that has found that many of the items display DIF when the scales are treated as unidimensional (e.g., Winterstein, Ackerman, et al., 2011). Importantly, these scales were designed to be unidimensional, and with the possible exception of the RSAS, multidimensional structures of the scales have not been reported (Cicero, Krieg, Becker, & Kerns, 2016). To our knowledge, these scales are always treated as unidimensional or random parcels are created for modeling purposes (e.g., Brown, Silvia, Myin-Germeys, Lewandowski, & Kwapil, 2008;

Kwapil et al., 2008). The finding that few items display DIF within an ESEM framework suggests that the DIF found in previous research may be related to the multidimensional nature of the scales.

The lack of clear measurement equivalence among these groups also suggests that future research could establish different norms when treating the variables as continuous variables. These norms could then be used to create different cut-scores for each ethnic group when treating the variables as categorical. As mentioned, the psychometric high-risk approach creates groups by determining cut-scores, typically meant to represent the top 2.5% of the population (e.g., Chapman et al., 1994; Kwapil, Miller, Zinser, Chapman, & Chapman, 1997; Lenzenweger, 1994). Although these groups were not meant to "carve nature at its joints" by discriminating schizotypes from healthy individuals, the presence of DIF may be problematic because it could lead to creating a high-risk group with people of varying latent levels of schizotypy that is related to their ethnicities. Since mean scores may not represent the same level of latent schizotypy in different ethnic groups, creating different norms and cut-scores for different groups based only on means for each group may not be appropriate. For example, future research could screen participants with the full or brief versions of the scales and then follow up with a structured interview. Based on the reports on the interview, receiver operator characteristic curves could be calculated and cut-scores could be developed to maximize sensitivity and specificity for each group separately. Previous research has used this strategy effectively for the assessment of schizotypy in African Americans (Kwapil et al., 2002).

The suggestion for the creation of different cut-scores for ethnic groups comes with several caveats. First, cut-scores may fail to replicate across samples, which suggests that these replication efforts are especially important for this work. Second, conceptualizations of psychopathology are clearly moving toward dimensional approaches and away from categorical diagnoses (e.g., Tackett, Silberschmidt, Krueger, & Sponheim, 2008). Thus, it may be preferable to treat these variables as dimensional, which does not require cut-scores. Finally, rather than attempting to create different cut-scores with these measures, researchers may instead focus on developing or identifying measures that do not have DIF.

The fourth goal of the current research was to compare and contrast the psychometric properties of the full and brief versions of the scales. The full and brief versions of the scales were correlated between $r = .88$ and $r = .92$ in the current research, which is consistent with previous research (Gross et al., 2012). These correlations indicate that 77% to 85% of the variance in the original scales can be explained by the brief versions of the scales. The brief versions of the scales also outperformed the full versions of the scales in the DIF analyses and provided a better unidimensional fit to the data when compared with recommended cutoffs for the fit indices. Given that little variance is lost by using the brief as opposed to the full versions of the scales, they have better psychometric properties, and are less burdensome to participants, researchers may choose to use the brief versions of the scales rather than the full versions. At the same time, the finding that the unidimensional structure fits better in the brief scales as compared with the full scales suggests that the brief scales may be more narrowly focused, which could affect their scores' relations with external correlates. Future research should continue to examine where the full and brief versions of the scales have the same magnitude of correlations with variables in their nomological network.

Another potential limitation of the current research is the use of college student participants. College students tend to have higher socioeconomic status and IQ than the general population. Thus, the results may not generalize to general population samples. At the same time, even if these results do not generalize beyond undergraduate samples, the current results are relevant to schizotypy research because the WSS are much more commonly used with college students than general population samples (Blanchard et al., 2011). Moreover, some evidence suggests that undergraduate have similar rates of psychopathology to people of the same age in the general population (Blanco et al., 2008), and have relatively high levels of attenuated psychotic symptoms (Cicero et al., 2014; Loewy, Bearden, Johnson, Raine, & Cannon, 2005). Future research could examine the DIF of these scales in samples drawn from the general population or clinical settings.

Another potential limitation of the current research is that the data were collected online and were not proctored by the experimenters. This could result in careless or invalid responding which could either contribute to measurement error or potentially even inflate correlations among scales leading to Type I errors (Huang, Liu, & Bowling, 2015). Future research on this topic could include special scales designed to detect careless or invalid responding, queries of diligence of responding, analysis of response times, and consistency of responses within individuals (Meade & Craig, 2012).

In summary, the results of the current research suggest that the one-factor models of the WSS only fit well for the B-MagicId and B-PerAb. Global tests of measurement invariance provided mixed results, but few of the items displayed DIF and many may cancel each other out. Within an ESEM framework, few of the items in the B-RSAS or the full versions of the scales displayed DIF across ethnic groups or between sexes. Future research could continue to examine the psychometric properties of these scales and develop specific cut-scores for psychometric high-risk designs in different ethnic groups.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplemental material

Supplemental material is available for this article online.

References

Bae, S. W., & Brekke, J. S. (2002). Characteristics of Korean-Americans with schizophrenia: A cross-ethnic comparison with African-Americans, Latinos, and Euro-Americans. *Schizophrenia Bulletin*, *28*, 703-717.

Barrantes-Vidal, N., Grant, P., & Kwapil, T. R. (2015). The role of schizotypy in the study of the etiology of schizophrenia spectrum disorders. *Schizophrenia Bulletin*, *41*(Suppl. 2), S408-S416. doi:10.1093/schbul/sbu191

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

Blanchard, J. J., Horan, W. P., & Brown, S. A. (2001). Diagnostic differences in social anhedonia: A longitudinal study of schizophrenia and major depressive disorder. *Journal of Abnormal Psychology*, *110*, 363-371.

Blanco, C., Okuda, M., Wright, C., Hasin, D. S., Grant, B. F., Liu, S. M., & Olfson, M. (2008). Mental health of college students and their non-college-attending peers: Results from the National Epidemiologic Study on Alcohol and Related Conditions. *Archives of General Psychiatry*, *65*, 1429-1437. doi:10.1001/archpsyc.65.12.1429

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 & Clinical Psychology*, *27*, 127-149. doi:10.1521/jscp.2008.27.2.127

Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466. doi:10.1037/0033-2909.105.3.456

Chapman, L. J., & Chapman, J. P. (1980). Scales for rating psychotic and psychotic-like experiences as continua. *Schizophrenia Bulletin*, *6*, 477-489.

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

Chapman, L. J., Edell, W. S., & Chapman, J. P. (1980). Physical anhedonia, perceptual aberration, and psychosis proneness. *Schizophrenia Bulletin*, *6*, 639-653.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*, 1005-1018. doi:10.1037/a0013193

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255. doi:10.1207/S15328007SEM0902_5

Chmielewski, M., Fernandes, L. O., 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

Cicero, D. C., Krieg, A., Becker, T. M., & Kerns, J. G. (2016). Evidence for the discriminant validity of the Revised Social Anhedonia Scale from social anxiety. *Assessment*, *23*, 544-556. doi:10.1177/1073191115590851

Cicero, D. C., Martin, E. A., Becker, T. M., Docherty, A. R., & Kerns, J. G. (2014). Correspondence between psychometric and clinical high risk for psychosis in an undergraduate population. *Psychological Assessment*, *26*, 901-915. doi:10.1037/a0036432

Claridge, G., McCreery, C., Mason, O., Bentall, R., Boyle, G., Slade, P., & Popplewell, D. (1996). The factor structure of "schizotypal" traits: A large replication study. *British Journal of Clinical Psychology*, *35*(Pt. 1), 103-115.

Earl, T. R., Fortuna, L. R., Gao, S., Williams, D. R., Neighbors, H., Takeuchi, D., & Alegria, M. (2015). An exploration of how psychotic-like symptoms are experienced, endorsed, and understood from the National Latino and Asian American Study and National Survey of American Life. *Ethnicity & Health*, *20*, 273-292. doi:10.1080/13557858.2014.921888

Eckbald, 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

Eckbald, M., Chapman, L. J., Chapman, J. P., & Mishlove, M. (1982). *The Revised Social Anhedonia Scale* (Unpublished test). University of Wisconsin, Madison, WI.

Edell, W. S. (1995). The psychometric measurement of schizotypy using the Wisconsin Scales of Psychosis-Proneness. In G. Miller (Ed.), *The behavioral high-risk paradigm in psychopathology* (pp. 1-46). New York, NY: Springer.

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

Gooding, D. C., Tallent, K. A., & Matts, C. W. (2005). Clinical status of at-risk individuals 5 years later: Further validation of the psychometric high-risk strategy. *Journal of Abnormal Psychology*, *114*, 170-175. doi:10.1037/0021-843x.114.1.170

Gross, G. M., Silvia, P. J., Barrantes-Vidal, N., & Kwapil, T. R. (2012). Psychometric properties and validity of short forms of the Wisconsin Schizotypy Scales in two large samples. *Schizophrenia Research*, *134*, 267-272. doi:10.1016/j.schres.2011.11.032

Horan, W. P., Reise, S. P., Subotnik, K. L., Ventura, J., & Nuechterlein, K. H. (2008). The validity of Psychosis Proneness Scales as vulnerability indicators in recent-onset schizophrenia patients. *Schizophrenia Research*, *100*, 224-236. doi:10.1016/j.schres.2007.12.469

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424-453. doi:10.1037//1082-989X.3.4.424

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*, 828-845. doi:10.1037/a0038510

Karcher, N. R., Slutske, W. S., Kerns, J. G., Piasecki, T. M., & Martin, N. G. (2014). Sex differences in magical ideation: A community-based twin study. *Personality Disorders: Theory, Research, and Treatment*, *5*, 212-219. doi:10.1037/per0000040

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

Kim, J. H. (2006). What's with the ghosts? Portrayals of spirituality in Asian American literature. *Spiritus: A Journal of Christian Spirituality*, *6*, 241-248.

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. (2015). Schizotypy: Looking back and moving forward. *Schizophrenia Bulletin*, *41*(Suppl. 2), S366-S373. doi:10.1093/schbul/sbu186

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., Chapman, L. J., & Chapman, J. (1999). Validity and usefulness of the Wisconsin manual for assessing psychotic-like experiences. *Schizophrenia Bulletin*, *25*, 363-375.

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., Miller, M. B., Zinser, M. C., Chapman, J., & Chapman, L. J. (1997). Magical ideation and social anhedonia as predictors of psychosis proneness: A partial replication. *Journal of Abnormal Psychology*, *106*, 491-495.

Lenzenweger, M. F. (1991). Confirming schizotypic personality configurations in hypothetically psychosis-prone university students. *Psychiatry Research*, *37*, 81-96.

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

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

Linscott, R. J., Marie, D., Arnott, K. L., & Clarke, B. L. (2006). Over-representation of Maori New Zealanders among adolescents in a schizotypy taxon. *Schizophrenia Research*, *84*, 289-296. doi:10.1016/j.schres.2006.02.006

Loewenthal, K. (2006). *Religion, culture, and mental health*. Cambridge, England: Cambridge University Press.

Loewy, R. L., Bearden, C. E., Johnson, J. K., Raine, A., & Cannon, T. D. (2005). The Prodromal Questionnaire (PQ): Preliminary validation of a self-report screening measure for prodromal and psychotic syndromes. *Schizophrenia Research*, *79*, 117-125.

MacDonald, A. W., Pogue-Geile, M. F., Debski, T. T., & Manuck, S. (2001). Genetic and environmental influences on schizotypy: A community-based twin study. *Schizophrenia Bulletin*, *27*, 47-58.

Markus, H. R., & Kitayama, S. (1991). Culture and self: Implications for cognition, emotion, and motivation. *Psychological Review*, *92*, 224-253.

Martin, E. A., Cicero, D. C., & Kerns, J. G. (2012). Social anhedonia, but not positive schizotypy, is associated with poor affective control. *Personality Disorders: Theory, Research, and Treatment*, *3*, 263-272. doi:10.1037/a0024488

Mata, I., Mataix-Cols, D., & Peralta, V. (2005). Schizotypal Personality Questionnaire-Brief: Factor structure and influence of sex and age in a nonclinical population. *Personality and Individual Differences*, *38*, 1183-1192. doi:10.1016/j.paid.2004.08.001

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*, 437-455. doi:10.1037/a0028085

Meehl, P. E. (1990). Toward an integrated theory of schizotaxia, schizotypy, and schizophrenia. *Journal of Personality Disorders*, *4*, 1-99.

Monahan, P. O., McHorney, C. A., Stump, T. E., & Perkins, A. J. (2007). Odds ratio, delta, ETS classification, and standardization measures of DIF magnitude for binary logistic regression. *Journal of Educational and Behavioral Statistics*, *32*, 92-109. doi:10.3102/1076998606298035

Morgan, C., Charalambides, M., Hutchinson, G., & Murray, R. M. (2010). Migration, ethnicity, and psychosis: Toward a sociodevelopmental model. *Schizophrenia Bulletin*, *36*, 655-664. doi:10.1093/schbul/sbq051

Muthén, L. K., & Muthén, B. O. (1998-2016). *Mplus: User's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Neale, J. M., & Oltmanns, T. F. (1980). *Schizophrenia*. New York, NY: Wiley.

Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning*. Thousand Oaks, CA: Sage.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598-609. doi:10.1037/0022-3514.46.3.598

Rawlings, D., Claridge, G., & Freeman, J. L. (2001). Principal components analysis of the Schizotypal Personality Scale (STA) and the Borderline Personality Scale (STB). *Personality and Individual Differences*, *31*, 409-419. doi:10.1016/S0191-8869(00)00146-X

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

Sharpley, M. S., Hutchinson, G., Murray, R. M., & McKenzie, K. (2001). Understanding the excess of psychosis among the African–Caribbean population in England. *British Journal of Psychiatry*, *178*, s60-s68.

Sharpley, M. S., & Peters, E. R. (1999). Ethnicity, class and schizotypy. *Social Psychiatry Psychiatric Epidemiology*, *34*, 507-512. doi:10.1007/s001270050168

Soto, J. A., Levenson, R. W., & Ebling, R. (2005). Cultures of moderation and expression: Emotional experience, behavior, and physiology in Chinese Americans and Mexican Americans. *Emotion*, *5*, 154-165. doi:10.1037/1528-3542.5.2.154

Tackett, J. L., Silberschmidt, A. L., Krueger, R. F., & Sponheim, S. R. (2008). A dimensional model of personality disorder: Incorporating *DSM* Cluster A characteristics. *Journal of Abnormal Psychology*, *117*, 454-459. doi:10.1037/0021-843X.117.2.454

Tsuang, M. T., Stone, W. S., Tarbox, S. I., & Faraone, S. V. (2002). An integration of schizophrenia with schizotypy: Identification of schizotaxia and implications for research on treatment and prevention. *Schizophrenia Research*, *54*, 169-175. doi:10.1016/S0920-9964(01)00364-4

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

Winterstein, B. P., Silvia, P. J., Kwapil, T. R., Kaufman, J. C., Reiter-Palmon, R., & Wigert, B. (2011). Brief assessment of schizotypy: Developing short forms of the Wisconsin Schizotypy Scales. *Personality and Individual Differences*, *51*, 920-924. doi:10.1016/j.paid.2011.07.027

Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement*, *35*, 339-361. doi:10.1177/0146621611405984