Psychometric properties of the Multidimensional Schizotypy Scale and Multidimensional Schizotypy Scale–Brief: Item and scale test–retest reliability and concordance of original and brief forms

By: Kathryn C. Kemp, Georgina M. Gross, and Thomas R. Kwapil

Kemp, K.C., Gross, G.M., & Kwapil, T.R. (2019). Psychometric properties of the Multidimensional Schizotypy Scale and Multidimensional Schizotypy Scale–Brief: Item and scale test–retest reliability and concordance of original and brief forms. *Journal of Personality Assessment,* https://doi.org/10.1080/00223891.2019.1591425

This is an Accepted Manuscript of an article published by Taylor & Francis in *Journal of Personality Assessment* on 23 April 2019, available online: http://www.tandfonline.com/10.1080/00223891.2019.1591425.

Abstract:

We conducted 2 studies examining the psychometric properties of the Multidimensional Schizotypy Scale (MSS) and the Multidimensional Schizotypy Scale–Brief (MSS–B). These studies offered the first examination of the scales' test–retest reliability (dependability), impact of testing interval on test–retest reliability, and concordance of the 2 versions. The MSS and MSS–B contain positive, negative, and disorganized schizotypy subscales that reflect current models of schizotypy. In Study 1, MTurk participants ($n = 245$) completed the scales at 2 assessments across a 7-week time interval. Test–retest reliabilities (intraclass correlations) ranged from .84 to .90 for the MSS subscales and from .77 to .85 for the MSS–B subscales, and were unaffected by the interval length between administrations. The association between the same subscales of the MSS and MSS–B across the 2 assessments ranged from .79 to .87, indicating good correspondence between the MSS and MSS–B. Scores on the MSS–B subscales in Study 1 were derived from the full-length MSS. Therefore, we replicated the findings using the actual MSS–B subscales in Study 2 in 3 samples (total $n = 339$). The results indicated that the MSS and MSS–B have good internal consistency and test–retest reliabilities (dependability), and that there is high concordance between the full-length and brief versions.

Keywords: Multidimensional Schizotypy Scale (MSS) | Multidimensional Schizotypy Scale (MSS) | schizophrenia

Article:

Schizophrenia is a severe mental illness characterized by psychosis, negative symptoms, and disruptions in thought and behavior (American Psychiatric Association, 2013). Current models suggest that schizophrenia is the most extreme manifestation of a continuum of clinical and subclinical impairment referred to as schizotypy (Kwapil & Barrantes-Vidal, 2012;

Lenzenweger, 2010; Meehl, 1990). Schizotypy offers a useful construct as it encompasses subclinical manifestations, the psychosis prodrome, and schizophrenia-spectrum disorders. The construct also provides a framework for investigating the etiology, development, and expression of schizophrenia-spectrum psychopathology (Kwapil & Barrantes-Vidal, 2015) without the confounds associated with clinical disorders (Lenzenweger, 2006). Schizotypy is heterogeneous, and this heterogeneity can be captured in a multidimensional structure that includes positive, negative, and disorganized dimensions (Kwapil & Barrantes-Vidal, 2015; Mason & Claridge, 2006; Vollema & van den Bosch, 1995). Positive schizotypy involves odd beliefs, unusual perceptual experiences, and suspiciousness. Negative schizotypy is characterized by diminished functioning such as affective flattening, anhedonia, avolition, anergia, and social withdrawal. Disorganized schizotypy is characterized by disturbances in thoughts, speech, and behavior.

Questionnaire measures have been widely used for assessing schizotypic characteristics and examining risk for developing schizophrenia-spectrum psychopathology (see reviews by Chapman, Chapman, & Kwapil, 1995; Kwapil & Chun, 2015; Kwapil, Gross, Silvia, & Barrantes-Vidal, 2013; Mason, 2015; Mason, Claridge, & Williams, 1997). The most widely used measures of schizotypy include the Wisconsin Schizotypy Scales (WSS) consisting of the Perceptual Aberration (Chapman, Chapman, & Raulin, 1978), Magical Ideation (Eckblad & Chapman, 1983), Physical Anhedonia (Chapman, Chapman, & Raulin, 1976), and Revised Social Anhedonia (Eckblad, Chapman, Chapman, & Mishlove, 1982) Scales, the Schizotypal Personality Questionnaire (SPQ; Raine, 1991), and the Oxford–Liverpool Inventory of Feelings and Experiences (O–LIFE; Mason, Claridge, & Jackson, 1995). In addition, brief versions have been derived for each of these measures including the WSS–B (Winterstein et al., 2011), SPQ–B; (Raine & Benishay, 1995), and the O–LIFE–SV (Mason, Linney, & Claridge, 2005). Although these and other questionnaire measures of schizotypy have been widely used, extant measures often suffer from a number of limitations, including factor structures that are inconsistent with current conceptual models of schizotypy, outdated or biased items, and psychometric limitations.

Multidimensional Schizotypy Scale

The Multidimensional Schizotypy Scale (MSS; Kwapil, Gross, Silvia, Raulin, & Barrantes-Vidal, 2018) and the Multidimensional Schizotypy Scale–Brief (MSS–B; Gross, Kwapil, Raulin, Silvia, & Barrantes-Vidal, 2018) were developed to assess current conceptualizations of positive, negative, and disorganized schizotypy and to address the limitations associated with existing measures of schizotypy. Scale construction followed the recommendations of DeVellis (2012) and item generation was based on detailed trait descriptions of the three schizotypy dimensions. Classical test theory, item-response theory (IRT), and differential item functioning (DIF) were employed to derive the 77-item MSS and the 38-item MSS–B with both measures containing positive, negative, and disorganized schizotypy subscales. The MSS and MSS–B have good psychometric properties, good item discrimination, and minimal item bias for gender and race or ethnicity in large derivation (n = 6,265) and cross-validation (n = 1,000) samples. Coefficient alpha reliabilities range from .88 to .94 for the MSS subscales (Kwapil, Gross, Silva et al., 2018) and .78 to .90 for the MSS–B subscales (Gross, Kwapil, Raulin, et al., 2018). Furthermore, initial studies supported the construct validity of the schizotypy subscales (Gross, Kwapil, Burgin,

et al., 2018; Kemp, Gross, Barrantes-Vidal, & Kwapil, 2018; Kwapil, Gross, Burgin, et al., 2018). Although the preliminary reliability and validity findings are promising, the test–retest reliabilities of the MSS and MSS–B subscales and items have not been reported. Furthermore, the concordance between the MSS and MSS–B subscales has not been fully demonstrated.

Goals of this study

The primary goal of this study was to examine the test–retest reliability (dependability) of the MSS and MSS–B subscales in a large sample across 3- to 7-week intervals. Current assessment models distinguish between dependability and stability when conceptualizing test–retest reliability (e.g., Chmielewski & Watson, 2009; Watson, 2004). Cattell, Eber, and Tatsuoka (1970) defined dependability as the correlation between repeated administrations of a test when the time course was brief enough that participants should not change in terms of the construct being measured. In contrast, stability assesses the extent to which a measure captures change across intervals in which true change in the construct is expected. This study focused on assessing the dependability of the MSS and MSS–B subscales. Positive, negative, and disorganized schizotypy are expected to be stable across our retest interval. Note that Watson (2004) stressed the importance of selecting theoretically meaningful retest intervals and Cattell et al. (1970) suggested that intervals of 2 months or less were optimal for assessing dependability and reliability. Our retest interval also should be relevant for researchers who use the inventories as screening measures for selecting participants for subsequent laboratory or clinical studies.

It was expected that all three subscales for each measure would exhibit high test–retest reliability. It was expected that negative schizotypy, which is characterized as trait-like and enduring compared to more episodic positive schizotypy experiences, would exhibit the highest test–retest reliability, although we expected high dependability for all three subscales. In addition to examining the dependability of the subscales, we also examined the dependability of the individual items. This study sought to replicate the psychometric properties and intercorrelations of the MSS and MSS–B. We also examined whether the length of the interval between Time 1 (T1) and Time 2 (T2) affected the dependability of the subscale scores. Finally, we examined the association of scores on the original MSS subscales with scores on the MSS–B subscales at the two assessments to examine the degree of concordance of the two versions.

Study 1

Methods

Participants

The study employed Amazon Mechanical Turk (MTurk) and was available to participants 18 years or older who reside in the United States. A total of 359 participants (*M* age = 36.6 years, *SD* = 11.1; 53.5% female; 79% White, 9% African American, 6% Asian or Pacific Islander, 5% Hispanic, 1% other; 98% English as first language) completed the T1 survey. Participants from T1 were dropped for invalid responding (*n* = 66 or 18%). Participants with invalid T1 data were not invited to complete the T2 retest survey. Of the 293 T1 individuals

invited to complete the retest, 245 (84%) completed T2 surveys. No participants had invalid protocols at T2. Participants who completed the T2 assessment were older than those lost to attrition, $t(291) = -2.55$, $p < .05$, but did not differ on sex composition, Fisher's exact test = .75; ethnicity, $\chi^2(4) = 1.23$, $p = .94$; or English as a first language, Fisher's exact test = 1.00.

Measures

Participants completed demographic questions and the full-length MSS at both T1 and T2 assessments. The MSS contains 77 true–false items that assess positive, negative, and disorganized schizotypy. The MSS–B contains 38 items drawn from the MSS. The scale items are presented in Kwapil, Gross, Silvia, et al. (2018) and Gross, Kwapil, Raulin, et al. (2018). Scores on the MSS–B subscales were derived from the full-length MSS. A 13-item infrequency questionnaire (Chapman & Chapman, 1983) was included to screen out invalid responders. Participants who endorsed more than two infrequency items were omitted from the analyses. The MSS and infrequency questionnaire were intermixed into five blocks presented in random order.

Procedures

Participants were recruited via MTurk, and the questionnaires were completed using the Qualtrics online survey system. The project received institutional review board approval from the University of Illinois at Urbana-Champaign, and participants provided informed consent prior to completing the surveys. Participants at T1 were notified that they would be contacted to complete another survey at T2, although they were not obligated to participate at T2. They were sent an invitation to complete the second survey 20 days after T1 completion, and the survey was left open for 1 month. The mean test–retest interval was 26.0 days ($SD = 7.2$ days, range = 20–49 days). Participants received $1 for completing the T1 survey and $2 for completing the T2 survey.

Results

Descriptive statistics

Descriptive statistics for the MSS and MSS–B at the T1 and T2 assessments are given in Table 1 and are comparable to findings from other large samples (Gross, Kwapil, Raulin, et al., 2018; Kwapil, Gross, Silvia, et al., 2018). The MSS and MSS–B subscales were developed to have positively skewed distributions by including items of low endorsement frequency that reflect the relatively rare nature of schizotypic experiences in the general population. The skew and kurtosis values in this sample are consistent with values from previous large samples. Note that parametric statistical tests were used in this study given that they are relatively robust to violations of normality when using large samples and highly reliable measures. However, the results were substantively unchanged when nonparametric analyses were employed. The alpha level was set at .05 for all analyses; however, we stress that results should be interpreted in light of effect sizes and that test–retest reliability should be interpreted in light of the standards of the field. For example, a test–retest reliability of .30 with a sample of 245 participants would be statistically significant, but would not offer much support for the dependability of the measure.

Table 1. Descriptive statistics for the Multidimensional Schizotypy Scale (MSS) and Brief Version (MSS–B) at test (Time 1) and retest (Time 2) assessments in Study 1.

Criterion	*M*	*SD*	Range	Coefficient alpha	Skew	*SE*	Kurtosis	*SE*
MSS Time 1								
Positive schizotypy	2.13	3.37	0–26	.87	3.02	.16	13.04	.31
Negative schizotypy	4.85	5.54	0–26	.91	1.58	.16	2.17	.31
Disorganized schizotypy	3.00	5.15	0–25	.94	2.18	.16	4.53	.31
MSS Time 2								
Positive schizotypy	2.24	3.80	0–24	.90	2.80	.16	9.37	.31
Negative schizotypy	4.54	5.58	0–26	.92	1.49	.16	1.68	.31
Disorganized schizotypy	2.78	5.12	0–25	.94	2.46	.16	6.06	.31
MSS–B Time 1								
Positive schizotypy	1.13	1.79	0–13	.75	2.80	.16	11.07	.31
Negative schizotypy	2.18	2.77	0–13	.84	1.67	.16	2.37	.31
Disorganized schizotypy	1.37	2.58	0–12	.89	2.35	.16	5.28	.31
MSS–B Time 2								
Positive schizotypy	1.16	2.04	0–13	.82	2.78	.16	9.56	.31
Negative schizotypy	2.10	2.76	0–13	.84	1.55	.16	1.97	.31
Disorganized schizotypy	1.30	2.61	0–12	.91	2.45	.16	5.76	.31

Note. $N = 245$.

Paired-samples *t* tests indicated that the scores on the six schizotypy subscales did not differ significantly from T1 to T2; MSS positive, $t(244) = -.89$, $p = .37$; MSS negative, $t(244) = 1.94$, $p = .05$; MSS disorganized, $t(244) = 1.29$, $p = .20$; MSS–B positive, $t(244) = -.35$, $p = .73$; MSS–B negative, $t(244) = .79$, $p = .43$; and MSS–B disorganized, $t(244) = .65$, $p = .52$. Independent-samples *t* tests assessed whether there were any statistical differences between completers of both T1 and T2 and participants who were lost at T2 to attrition. There were 293 participants with valid data at T1, and 245 of these participated at T2. Participants who completed T2 and participants who were lost to attrition ($n = 48$) did not differ on any of the six schizotypy subscales at T1; MSS positive, $t(291) = .08$, $p = .94$; MSS negative, $t(291) = -.47$, $p = .64$; MSS disorganized, $t(291) = -.47$, $p = .64$; MSS–B positive, $t(291) = -.08$, $p = .94$; MSS–B negative, $t(291) = -.67$, $p = .50$; MSS–B disorganized, $t(291) = -.39$, $p = .69$. Intercorrelations of the MSS and MSS–B subscales were assessed at each time point and are presented in Supplementary Table S.1. The correlations were consistent at T1 and T2 and with previous findings (Gross, Kwapil, Raulin, et al., 2018; Kwapil, Gross, Silvia, et al., 2018).

Test–retest reliability

Test–retest reliabilities (dependability) of the MSS and MSS–B subscales were examined using both Pearson correlations and intraclass correlations (ICCs) with two-way mixed effects and absolute agreement for single measures (see Table 2). Note that Pearson and ICC values are nearly identical, suggesting both high agreement and comparable magnitude of the scores over time. Overall, the MSS and MSS–B subscales demonstrated high test–retest reliability across the two assessments. Scatterplots of the T1 and T2 scores for the MSS and MSS–B subscales are presented in Supplemental Figures S.1 and S.2.

Table 2. Test–retest reliability of the Multidimensional Schizotypy Scale and Multidimensional Schizotypy Scale–Brief in Study 1.

Criterion	Intraclass correlation	Pearson *r*
Multidimensional Schizotypy Scale		
Positive schizotypy	.84	.85
Negative schizotypy	.90	.90
Disorganized schizotypy	.85	.85
Multidimensional Schizotypy Scale–Brief		
Positive schizotypy	.77	.78
Negative schizotypy	.85	.85
Disorganized schizotypy	.82	.82

Note. All values are significant at $p < .05$ and represent large effect sizes.

Note that to confirm that the test–retest correlations were not inflated by large numbers of low-scoring participants, we recomputed the Pearson test–retest correlations for each subscale after dropping participants who scored 0 on the subscale at T1. The test–retest correlations were largely unchanged: MSS positive schizotypy $r = .81$, MSS negative schizotypy $r = .90$, MSS disorganized schizotypy $r = .83$, MSS–B positive schizotypy $r = .70$, MSS–B negative schizotypy $r = .83$, and MSS–B disorganized schizotypy $r = .78$. Thus, the test–retest reliability of the MSS and MSS–B did not appear to be artificially inflated by 0 scorers.

We also examined the Pearson correlations of the individual MSS and MSS–B items across the two assessments: MSS positive (mean item correlation = .50, range = .21–.74), MSS negative (mean item correlation = .58, range = .41–.74), MSS disorganized (mean item correlation = .53, range = .37–.72), MSS–B positive (mean item correlation = .49, range = .21–.74), MSS–B negative (mean item correlation = 0.54, range = .41–.65), MSS–B disorganized (mean item correlation = .53, range = .37–.72). The individual item correlation values are presented in Supplementary Table S.2.

Table 3. Interval as a potential moderator of the association between test and retest Multidimensional Schizotypy Scale (MSS) and Multidimensional Schizotypy Scale–Brief (MSS–B) subscales in Study 1.

	Step 1 (*df* = 1,243)			**Step 2 (*df* = 1,242)**			**Step 3 (*df* = 1,241)**			
	T1 Schizotypy subscale			**Interval**			**T1 Schizotypy × Interval**			
Criteria	**β**	**ΔR^2**	**f^2**	**β**	**ΔR^2**	**f^2**	**β**	**ΔR^2**	**f^2**	**Total R^2**
Time 2 MSS subscale										
Positive	***.850****	***.723***	***2.610***	.024	.001	.004	.002	.000	.000	.724
Negative	***.896****	***.803***	***4.076***	.043	.002	.005	−.013	.000	.005	.805
Disorganized	***.853****	***.728***	***2.676***	.027	.001	.004	.014	.000	.000	.729
Time 2 MSS–B subscale										
Positive	***.779****	***.606***	***1.538***	.046	.002	.005	.046	.002	.005	.610
Negative	***.846****	***.716***	***2.521***	.027	.001	.004	.034	.001	.004	.718
Disorganized	***.816****	***.666***	***1.994***	.011	.000	.000	−.032	.001	.003	.667

Note. Medium effect sizes (f^2) are shown in bold, and large effect sizes are shown in bold and italics. Each row represents a separate regression analysis in which the Time 1 schizotypy subscale entered at Step 1 corresponds with outcome Time 2 schizotypy subscale.
* $p < .001$.

Effects of interval on test–retest reliability

The interval between T1 and T2 ranged from 20 to 49 days. We examined whether the interval moderated the association of the subscale scores at the two assessments by conducting a series of hierarchical regression analyses predicting T2 scores of each subscale. In each analysis, the analogous T1 schizotypy subscale score was entered at Step 1, interval duration was entered at Step 2, and the T1 Subscale × Interval interaction term was entered at Step 3 (Table 3). The standardized regression coefficient (β), change in R^2, and effect size f^2 were reported for each predictor in the regressions. According to Cohen (1992), f^2 values above .15 are medium and those above .35 are large effect sizes. In all six analyses, the T1 subscale scores significantly predicted T2 subscale scores (large effect), but neither the interval nor T1 × Interval interaction terms were significant. Thus, test–retest reliability (dependability) was high and consistent across a 3- to 7-week period.

Associations of long and short forms

When examining the comparability of original and brief forms of scales, Smith, McCarthy, and Anderson (2000) warned that reporting the correlation of the full-length scale and brief version derived from the same administration is problematic and overestimates the correlation of the two forms. Therefore, in this study, we examined the association of the original MSS and MSS–B versions of the subscales by correlating the original subscales at T1 with the brief subscales derived at T2, and vice versa. The correlations of the T1 MSS subscales with T2 MSS–B subscales were .82 for positive, .87 for negative, and .83 for disorganized schizotypy. The correlations of the T2 MSS subscales with T1 MSS–B subscales were .79 for positive, .85 for negative, and .83 for disorganized schizotypy (all statistically significant and large effect sizes). The correlation values for each subscale were almost identical across the two assessments.

Note that a limitation of Study 1 was that scores on the MSS–B were derived from the administration of the full-length MSS. The MSS–B subscales showed strong psychometric properties (comparable with the MSS). However, we cannot definitively draw conclusions about psychometric properties of the MSS–B independent of the administration of the full MSS.

Study 2

Study 2 was conducted to overcome the limitation of Study 1 that the MSS–B scores were derived from the MSS. Specifically, Study 2 involved the administration of the MSS–B to assess the subscale test–retest reliability and assess the association of the MSS–B subscales with the MSS subscales using comparable sampling procedures and timelines as in Study 1.

Methods

Participants

This study employed Amazon MTurk to assess three additional samples to examine the properties of the MSS–B when it was not derived from administration of the full MSS. Sample A

completed the MSS–B at both the test and retest assessments so that we could assess test–retest reliability for the MSS–B subscales. Sample B completed the MSS–B at T1 and the MSS at T2. Sample C completed the MSS at T1 and the MSS–B at T2. Samples B and C were collected to examine the association of the analogous subscales from the MSS and MSS–B.

A total of 314 Sample A participants (*M* age = 35.4 years, *SD* = 12.0; 44% female; 69% White, 11% African American, 5% Asian or Pacific Islander, 8% Hispanic, 7% other; 99% English as first language) completed the T1 survey. Participants from T1 were dropped for invalid responding (*n* = 108 or 34%). Note that the rate of invalid participants was higher in all three samples in Study 2 (conducted in Fall 2018) than in Study 1 (conducted in Fall 2017). This was likely due to the elevated rate of nonhuman "bot" participants in MTurk that arose in 2018 (e.g., Stokel-Walker, 2018). Participants with invalid T1 data were not invited to complete the T2 retest survey. Six additional participants were not invited to complete the T2 survey because they did not have valid contact information. Of the 200 T1 individuals invited to complete the retest, 159 (80%) completed T2 surveys. Six participants had invalid protocols at T2. Participants who completed the T2 assessment were older than those lost to attrition, $t(198) = -2.50$, $p = .01$, but did not differ on sex composition, Fisher's exact test = .29; ethnicity, $\chi^2(4) = 5.97$, $p = .20$; or English as a first language, Fisher's exact test = .50.

A total of 209 Sample B participants (*M* age = 34.1 years, *SD* = 11.0; 43% female; 67% White, 14% African American, 6% Asian or Pacific Islander, 5% Hispanic, 8% other; 98% English as first language) completed the T1 survey. A total of 80 (38%) of T1 participants were dropped due to invalid protocols and an additional two participants did not provide valid contact information. Of the 127 T1 individuals invited to complete the retest, 103 (81%) completed T2 surveys. Six participants had invalid protocols at T2. Participants who completed the T2 assessment were older than those lost to attrition, $t(125) = -2.07$, $p = .04$, but did not differ on sex composition, Fisher's exact test = .65; ethnicity, $\chi^2(4) = 5.08$, $p = .28$; or English as a first language, Fisher's exact test = 1.00.

A total of 205 Sample C participants (*M* age = 34.9 years, *SD* = 11.9; 40% female; 74% White, 9% African American, 3% Asian or Pacific Islander, 6% Hispanic, 7% other; 98% English as first language) completed the T1 survey. A total of 88 (43%) of T1 participants were dropped due to invalid protocols and one additional participant did not provide valid contact information. Of the 116 T1 individuals invited to complete the retest, 92 (79%) completed T2 surveys. Three participants had invalid protocols at T2. Participants who completed the T2 assessment did not differ from those lost to attrition on age, $t(114) = -0.93$, $p = .36$; sex composition, Fisher's exact test = .82; ethnicity, $\chi^2(4) = 2.06$, $p = .73$; or English as a first language, Fisher's exact test = .21.

Measures and procedures

Measures and procedures used in Study 2 were the same as in Study 1, except that participants completed the actual MSS–B, rather than having MSS–B scores derived from the full-length MSS. The mean test–retest interval was 21.4 days (*SD* = 3.3 days; range = 20–42 days) for Sample A, 21.2 days (*SD* = 3.0 days; range = 20–43 days) for Sample B, and 21.5 days (*SD* = 3.5 days; range = 20–37 days) for Sample C.

Results

Descriptive statistics

Descriptive statistics for the MSS and MSS–B for Samples A, B, and C at the T1 and T2 assessments are in Supplementary Table S.3 and are comparable to findings from Study 1. Paired-samples *t* tests indicated that the scores on the MSS–B schizotypy subscales did not differ significantly from T1 to T2 in Sample A for either the MSS–B positive, $t(152) = 1.13$, $p = .26$; or the MSS–B negative, $t(152) = 1.21$, $p = .23$, subscales. Scores were significantly lower at T2 for the MSS–B disorganized subscale, $t(152) = 2.67$, $p < .01$. Supplementary Table S.4 presents the comparisons of completers of both T1 and T2 and participants who were lost at T2 to attrition on T1 MSS or MSS–B subscale scores for each sample. Intercorrelations of the MSS and MSS–B subscales at each time point in each sample are presented in Supplementary Table S.5.

Test–retest reliability

Test–retest reliabilities of the MSS–B subscales were examined in Sample A using both Pearson correlations and ICCs with two-way mixed effects, and absolute agreement for single measures. Test–retest reliability was ICC = .71 and $r = .71$ for the positive schizotypy subscale, ICC = .85 and $r = .85$ for the negative schizotypy subscale, and ICC = .72 and $r = .71$ for the disorganized schizotypy subscale (note all values are statistically significant and represent large effect sizes). As in Study 1, the Pearson correlations and ICC values were closely comparable. We also examined the Pearson correlations of the individual MSS–B items across the two assessments: MSS–B Positive (mean item correlation = .46, range = .26–.69), MSS–B negative (mean item correlation = .54, range = .29–.71), and MSS–B disorganized (mean item correlation = .45, range = .32–.60). Note that due to the limited variability in interval length in all the Study 2 samples (87% completed the retest survey within the first 2 days and 96% within the first week of the retest), the effect of interval length on test–retest reliability was not examined.

Associations of long and short forms

The association of the MSS and MSS–B subscales across the retest interval was examined in Samples B and C. In Sample B, the correlations of the T1 MSS–B subscales with T2 MSS subscales were .76 for positive, .84 for negative, and .84 for disorganized schizotypy. In Sample C, the correlations of the T1 MSS subscales with T2 MSS–B subscales were .78 for positive, .78 for negative, and .73 for disorganized schizotypy (all statistically significant and large effect sizes).

Discussion

Schizotypy is a promising construct for understanding schizophrenia-spectrum psychopathology. Psychometric assessments are useful for examining correlates of schizotypy and identifying risk for developing spectrum disorders. However, existing measures of schizotypy suffer from a number of limitations. The MSS was developed to map onto current conceptual models of schizotypy and to address the shortcomings of existing measures. The MSS–B provides a brief

alternative with minimal loss of content coverage. Initial studies indicate that the MSS and MSS–B have good psychometric properties and promising validity.

Coefficient alpha values for the MSS subscales in the studies reported here were good to excellent, ranging from .87 to .95 in the two studies. Likewise, alpha values for the MSS–B subscales were solid, ranging from .75 to .91, and did not show reductions compared to the MSS beyond what was expected given the reduction in subscale length. The internal consistency reliability of the MSS and MSS–B subscales appears constant across multiple samples. These findings also provide further support for the two scales compared to other widely used measures of schizotypy and their brief versions. For example, coefficient alpha reliabilities for the MSS subscales are equal or superior to values for other established measures: WSS, range = .71 to .94 (Chapman, Chapman, & Miller, 1982; Vollema & van den Bosch, 1995); SPQ, total score range = .90 to .91, subscales range = .63 to .81(Raine, 1991); O–LIFE, subscales range = .77 to .89 (Mason et al., 1995). Likewise, reliabilities of the MSS–B subscales are consistent with or superior to established scales: WSS–B, range = .70 to .84 (Gross, Silvia, Barrantes-Vidal, & Kwapil, 2012); SPQ–B, total score range = .80 to .83, factor scores range = .72 to .78 (Raine & Benishay, 1995); O–LIFE–SV, range = .63 to .80 (Mason et al., 2005).

Although coefficient alpha provides information about internal consistency reliability, it does not provide information about the dependability of a measure across intervals in which the measured construct is expected to be stable. These studies reported test–retest reliability of the scales using ICC and Pearson *r*. ICC reflects both stability of scores and comparability of magnitude of scores, whereas Pearson *r* indicates degree of association. Both the MSS and MSS–B demonstrated strong test–retest reliability based on ICC and Pearson *r* values. Consistent with the notion that positive and disorganized schizotypy might involve more episodic experiences and negative schizotypy might be more trait-like, the test–retest reliability was slightly higher in the negative schizotypy subscale compared to the other two subscales. Note that the individual items of the MSS generally showed good reliability across the retest interval. The dependability of the MSS and MSS–B subscales was consistent with or superior to values for other established schizotypy questionnaires: WSS, subscales = .75 to .82 across 6 weeks, $n = 512$ (Chapman et al., 1982); WSS–B, factor scores = .78 to .84 across 10 weeks, $n = 106$ (Gross, Silvia, Barrantes-Vidal, & Kwapil, 2015); SPQ, total score = .82 across 2 months, $n = 25$ (Raine, 1991); SPQ–B, factor scores = .86 to .95, $n = 31$ (Raine & Benishay, 1995); O–LIFE, subscales = .77 to .93 across 3 to 6 months, $n = 30$ (Burch, Steel, & Hemsley, 1998); O–LIFE–SV, subscales = .69 to .87 across 1 month, $n = 102$ (Cella et al., 2013).

Not only did the MSS and MSS–B exhibit high test–retest reliability, but the reliability of the subscales was not affected by the interval between the assessments. These findings should be tempered by the fact that the interval between the two assessments was relatively narrow and most participants completed T2 within 3 to 4 weeks of T1. Nevertheless, the reliability of the subscales did not vary significantly across the time interval, which is consistent with the hypothesized trait-like nature of schizotypy. Note that these findings are relevant given that investigators often screen large numbers of participants with schizotypy questionnaires and then invite a subset of scorers to complete interview and laboratory measures weeks or months after the initial screening. Future studies of the dependability of the MSS and MSS–B subscales should also include benchmark measures in the same sample that can be used as a basis of

comparison for the dependability of the measures (Chmielewski & Watson, 2009; Watson, 2004).

To our knowledge, these are the first studies to compare the MSS and MSS–B concordance based on different administrations. Gross, Kwapil, Raulin, et al. (2018) reported high correlations of the MSS and MSS–B analogous subscales, but these were computed from the same administration and, as Smith et al. (2000) warned, undoubtedly overestimated the associations of these measures. Gross, Kwapil, Raulin, et al. (2018) reported that the analogous MSS and MSS–B subscales each correlated .95 or greater when employing this method. To address Smith et al.'s (2000) concerns of comparing a short form derived from its original counterpart in the same administration, we examined the association of the MSS and MSS–B at the two administrations. In these studies, the correlations ranged from .73 to .87. However, these associations might be slightly underestimated because they are confounded by stability over time.

Scores on the MSS–B in Study 1 were derived from the full-length MSS and not from administration of the MSS–B. Therefore, the MSS–B was administered in Study 2 to examine the test–retest reliability of the scale and its association with the MSS. The findings in Study 2 confirmed the findings of Study 1 that the MSS–B subscales have good test–retest reliability across a 1-month period and that they have good concordance with the analogous MSS subscales. The findings from Study 1 and Study 2 were closely comparable, despite smaller samples in Study 2 and problems with MTurk that developed between the administration of the two studies. Study 1 data collection was completed in Fall 2017, whereas Study 2 was conducted in Fall 2018. In 2018, there were well-documented problems of bots in MTurk samples that resulted in marked increases in the number of invalid protocols. This is reflected in the fact that Study 1 excluded 18% of protocols as invalid, whereas in Study 2 the rate increased to 38%.

Participants in these studies were not selected based on the presence or absence of psychopathology, and history of psychopathology was not assessed. Nevertheless, participants scored across the full range of the MSS and MSS–B subscales. Future studies should examine whether scores on the MSS and MSS–B are equally stable in clinical and nonclinical samples.

In summary, these are the first studies to assess the test–retest reliability (dependability) of the MSS and MSS–B and the association of the long and short forms at two different assessments. In addition to the replicability of the scales' psychometric properties with previous samples, the MSS and MSS–B demonstrated high levels of dependability over 3 to 7 weeks. Furthermore, the reliability of the subscales does not appear to be affected by the length of the interval between assessments. The MSS and MSS–B appear to offer theoretically based and psychometrically sound assessments of positive, negative, and disorganized schizotypy.

Acknowledgments. This article has earned the Center for Open science badges for Open Data. The data and materials are openly accessible at https://osf.io/xjgyd/.

References

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

Burch, G.S., 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.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. . (1970). *Handbook for the sixteen personality factor questionnaire (16 PF)*. Champaign, IL: Institute for Personality and Ability Testing.

Cella, M., Serra, M., Lai, A., Mason, O. J., Sisti, D., Rocchi, M. B. L., … Petretto, D. R. . (2013). Schizotypal traits in adolescents: Links to family history of psychosis and psychological distress. *European Psychiatry*, *28*(4), 247–253. doi:10.1016/j.eurpsy.2012.04.002

Chapman, J. P., Chapman, L. J., & Kwapil, T. R. . (1995). Scales for the measurement of schizotypy. In A. Raine, T. Lencz, S. A. Mednick, A. Raine, T. Lencz, & S. A. Mednick. (Eds.), *Schizotypal personality* (pp. 79–106). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511759031.006

Chapman, L. J., & Chapman, J. P. . (1983). *Infrequency scale for personality measures*. Champaign, IL: UIUC Department Of Psychology. Unpublished scale available from T. R. Kwapil.

Chapman, L. J., Chapman, J. P., & Miller, E. N. . (1982). Reliabilities and intercorrelations of eight measures of proneness to psychosis. *Journal of Consulting and Clinical Psychology*, *50*(2), 187–195. doi:10.1037/0022-006X.50.2.187

Chapman, L. J., Chapman, J. P., & Raulin, M. L. . (1976). Scales for physical and social anhedonia. *Journal of Abnormal Psychology*, *85*(4), 374–382. doi:10.1037/0021-843X.85.4.374

Chapman, L. J., Chapman, J. P., & Raulin, M. L. . (1978). Body-image aberration in schizophrenia. *Journal of Abnormal Psychology*, *87*(4), 399–407. doi:10.1037/0021-843X.87.4.399

Chmielewski, M., & Watson, D. . (2009). What is being assessed and why it matters: The impact of transient error on trait research. *Journal of Personality and Social Psychology*, *97*(1), 186–202. doi:10.1037/a0015618

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

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

Eckblad, M., & Chapman, L. J. . (1983). Magical ideation as an indicator of schizotypy. *Journal of Consulting and Clinical Psychology*, *51*(2), 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*. Champaign, IL: UIUC Department of Psychology. Unpublished test copies available from T. R. Kwapil.

Gross, G. M., Kwapil, T. R., Burgin, C. J., Raulin, M. L., Silvia, P. J., Kemp, K. C., & Barrantes-Vidal, N. . (2018). Validation of the multidimensional schizotypy scale-brief in two large samples. *Journal of Psychopathology and Behavioral Assessment*, *40*(4), 669–677. doi:10.1007/s10862-018-9668-4

Gross, G. M., Kwapil, T. R., Raulin, M. L., Silvia, P. J., & Barrantes-Vidal, N. . (2018). The multidimensional schizotypy scale-brief: Scale development and psychometric properties. *Psychiatry Research*, *261*, 7–13. doi:10.1016/j.psychres.2017.12.033

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*(2-3), 267–272. doi:10.1016/j.schres.2011.11.032

Gross, G. M., Silvia, P. J., Barrantes-Vidal, N., & Kwapil, T. R. . (2015). The dimensional structure of short forms of the Wisconsin schizotypy scales. *Schizophrenia Research*, *166*(1-3), 80–85. doi:10.1016/j.schres.2015.05.016

Kemp, K. C., Gross, G. M., Barrantes-Vidal, N., & Kwapil, T. R. . (2018). Association of positive, negative, and disorganized schizotypy dimensions with affective symptoms and experiences. *Psychiatry Research*, *270*, 1143–1149. doi:10.1016/j.psychres.2018.10.031

Kwapil, T. R., & Barrantes-Vidal, N. . (2012). Schizotypal personality disorder: An integrative review. In T. A. Widiger. (Ed.), *The Oxford handbook of personality disorders* (pp. 437–477). New York, NY: 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., & Chun, C. A. . (2015). The psychometric assessment of schizotypy. In O. J. Mason, G. Claridge. (Eds.), *Schizotypy: New dimensions* (pp. 7–32). New York, NY: Routledge/Taylor & Francis Group.

Kwapil, T. R., Gross, G. M., Burgin, C. J., Raulin, M. L., Silvia, P. J., & Barrantes-Vidal, N. . (2018). Validity of the multidimensional schizotypy scale: Associations with schizotypal traits and normal personality. *Personality Disorders: Theory, Research, & Treatment*, *9*(5):458–466. doi:10.1037/per0000288

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*(3), 807–815. doi:10.1037/a0033759

Kwapil, T. R., Gross, G. M., Silvia, P. J., Raulin, M. L., & Barrantes-Vidal, N. . (2018). Development and psychometric properties of the multidimensional schizotypy scale: A new measure for assessing positive, negative, and disorganized schizotypy. *Schizophrenia Research*, *193*, 209–217. doi:10.1016/j.schres.2017.07.001

Lenzenweger, M. F. . (2006). The longitudinal study of personality disorders: History, design considerations, and initial findings. *Journal of Personality Disorders*, *20*(6), 645–670. doi:10.1521/pedi.2006.20.6.645

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

Mason, O. (2015). The assessment of schizotypy and its clinical relevance. *Schizophrenia Bulletin*, *41*(S2), S374–S385.

Mason, O., & Claridge, G. . (2006). The Oxford-Liverpool inventory of feelings and experiences O-LIFE: Further description and extended norms. *Schizophrenia Research*, *82*(2-3), 203–211. doi:10.1016/j.schres.2005.12.845

Mason, O., Claridge, G., & Jackson, M. . (1995). New scales for the assessment of schizotypy. *Personality and Individual Differences*, *18*(1), 7–13. doi:10.1016/0191-8869(94)00132-C

Mason, O., Claridge, G., & Williams, L. . (1997). Questionnaire measurement. In G. Claridge. (Ed.), *Schizotypy: Implications for illness and health* (pp. 19–37). New York, NY: Oxford University Press.

Mason, O., Linney, Y., & Claridge, G. . (2005). Short scales for measuring schizotypy. *Schizophrenia Research*, *78*(2-3), 293–296. doi:10.1016/j.schres.2005.06.020

Meehl, P. E. . (1990). Toward an integrated theory of schizotaxia, schizotypy, and schizophrenia. *Journal of Personality Disorders*, *4*(1), 1–99. doi:10.1521/pedi.1990.4.1.1

Raine, A. . (1991). The SPQ: A scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophrenia Bulletin*, *17*(4), 555–564. doi:10.1093/schbul/17.4.555

Raine, A., & Benishay, D. . (1995). The SPQ-B: A brief screening instrument for schizotypal personality disorder. *Journal of Personality Disorders*, *9*(4), 346–355. doi:10.1521/pedi.1995.9.4.346

Smith, G. T., McCarthy, D. M., & Anderson, K. G. . (2000). On the sins of short-form development. *Psychological Assessment*, *12*(1), 102–111. doi:10.1037/1040-3590.12.1.102

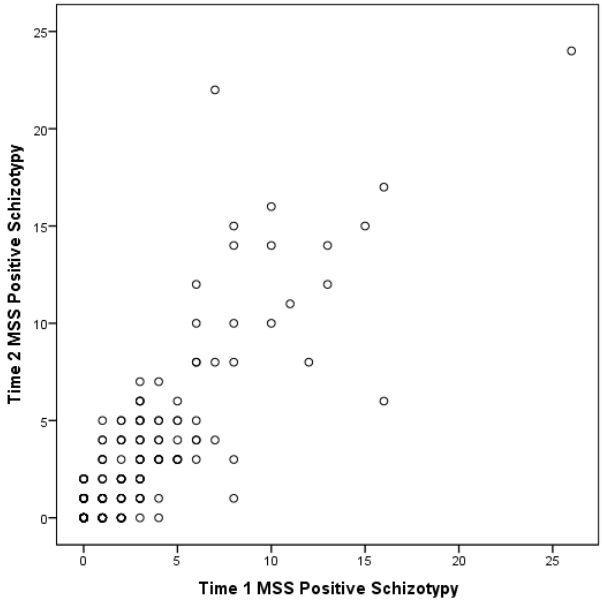
Stokel-Walker, C. . (2018, August 10). Bots on Amazon's Mechanical Turk are ruining psychology studies. *New Scientist*. Retrieved from https://www.newscientist.com/article/mg23931911-800-bots-on-amazons-mechanical-turk-are-ruining-psychology-studies/

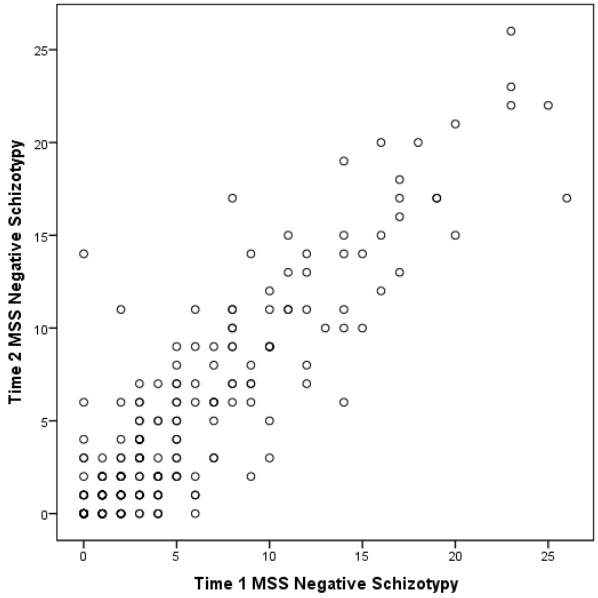
Vollema, M. G., & van den Bosch, R. J. . (1995). The multidimensionality of schizotypy. *Schizophrenia Bulletin*, *21*(1), 19–31. doi:10.1093/schbul/21.1.19

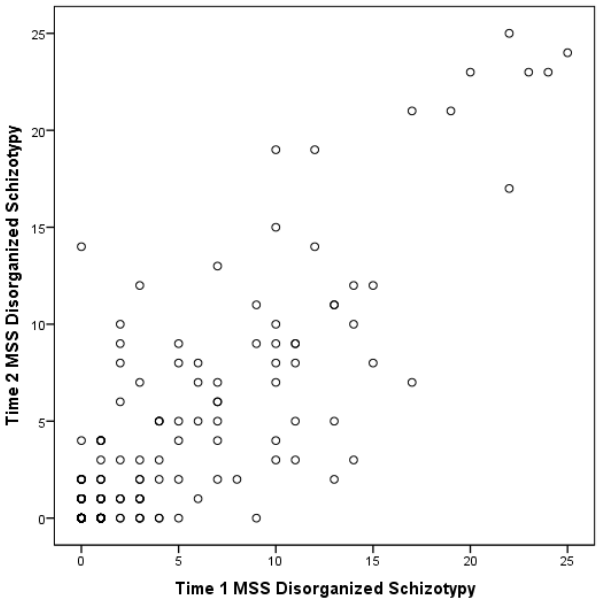
Watson, D. . (2004). Stability versus change, dependability versus error: Issues in the assessment of personality over time. *Journal of Research in Personality*, *38*(4), 319–350. doi:10.1016/j.jrp.2004.03.001

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*(8), 920–924. doi:10.1016/j.paid.2011.07.027

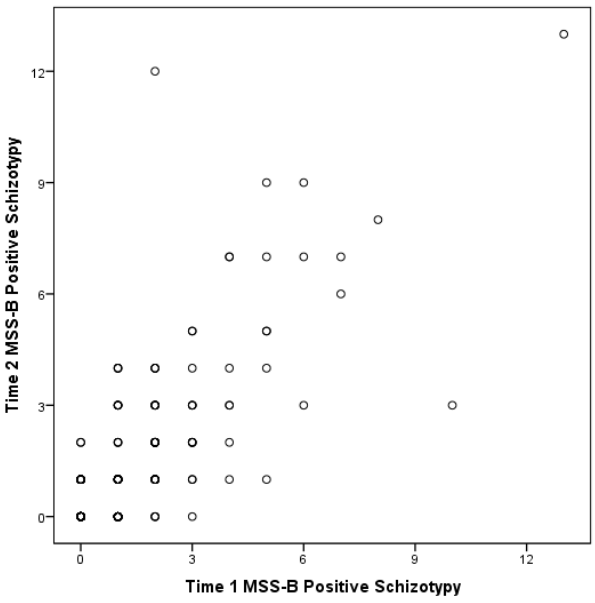
Supplemental Figure 1. Scatterplots of the Time 1 and Time 2 Assessments for the Multidimensional Schizotypy Scale (MSS) Subscales in Study 1

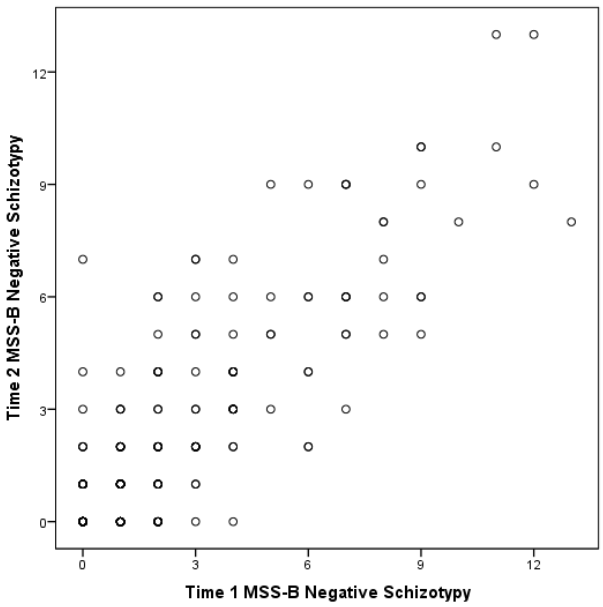


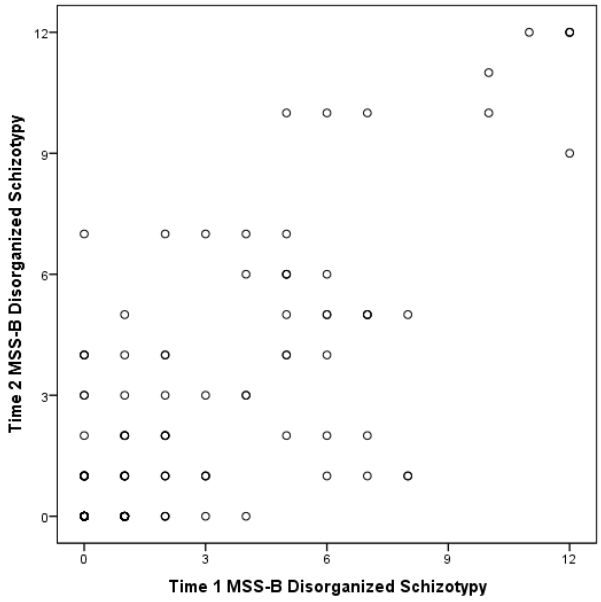




Supplemental Figure 2. Scatterplots of the Time 1 and Time 2 Assessments for the Multidimensional Schizotypy Scale-Brief (MSS-B) Subscales in Study 1







Supplementary Table 1. Correlations of the Subscales of the Multidimensional Schizotypy Scale and Multidimensional Schizotypy Scale-Brief in Study 1

MSS	Positive	Negative	Disorganized
Positive		.15*	**.39***
Negative	.12		.29*
Disorganized	**.35***	**.33***	

MSS-B	Positive	Negative	Disorganized
Positive		.12	**.36***
Negative	.09		.27*
Disorganized	**.33***	**.35***	

* $p<.05$

Note: Values above diagonal are from Time 2, and values below diagonal are from Time 1; medium effect sizes in bold

Supplementary Table 2. Pearson Correlations of the Multidimensional Schizotypy Scale (MSS) and Multidimensional Schizotypy (MSS-B) Items at Test and Retest Assessments in Study 1

Positive Schizotypy Items	Pearson *r*
I believe that dreams have magical properties.	0.63
I believe that ghosts or spirits can influence my life.	0.62
I believe that I could read other people's minds if I really tried.	0.66
I have had the momentary feeling that I might not be human.	0.74
Some people can make me aware of them just by thinking about me.	0.62
I have had the momentary feeling that someone's place has been taken by a look-alike.	0.34
I often wonder if everyone in the world is part of a secret experiment.	0.42
I have worried that people on other planets may be influencing what happens on Earth.	0.56
I occasionally have the feeling that my thoughts are not my own.	0.28
I have sometimes felt that strangers were reading my mind.	0.52
I have felt that there were messages for me in the way things were arranged, like furniture in a room.	0.21
Sometimes I feel that a television show or movie has a special message just for me.	0.35
I believe that there are secret signs in the world if you just know how to look for them.	0.53
I sometimes wonder if there is a small group of people who can control everyone else's behavior.	0.43
I occasionally worry that people I see on the street are spying on me.	0.76
I often worry that other people are out to get me.	0.55
I often think that I hear people talking only to discover that there was no one there.	0.54
Occasionally I have felt as though my body did not exist.	0.62
At times I have wondered if my body was really my own.	0.67
I have felt that something outside my body was a part of my body.	0.38
There are times when it feels like someone is touching me when no one is actually there.	0.41
Sometimes when I look at ordinary objects they seem strange or unreal.	0.45
There are times when I think I see another person, but there is actually no one there.	0.43
I have had experiences with seeing the future, ESP or a sixth sense.	0.74
I often worry that someone or something is controlling my behavior.	0.21
I often find hidden meanings or threats in things that people say or do.	0.33

Negative Schizotypy Items	Pearson *r*
Throughout my life I have noticed that I rarely feel strong positive or negative emotions.	0.41
I rarely feel strong emotions even in situations in which other people usually do.	0.53
Throughout my life there have been very few things that interest me.	0.68
My emotions have almost always seemed flat regardless of what is going on around me.	0.46
Generally I do not have many thoughts or emotions.	0.43
I often look forward to upcoming events.	0.56
Throughout my life, very few things have been exciting or interesting to me.	0.49
I tend to have few interests.	0.63

I have always preferred to be disconnected from the world.	0.43
Having close friends is not as important as people say.	0.61
I have never really been interested in having close relationships.	0.61
In general, it is important for me to have close relationships with other people.	0.64
When I move to a new place, I feel a strong desire to make friends.	0.65
If given the choice, I would much rather be with another person than alone.	0.58
Although there are things I enjoy doing by myself, I usually have more fun when I do things with other people.	0.59
I enjoy meeting new people and making new friends.	0.74
It has never been important to me to be involved with other people.	0.45
Most of the time I feel a desire to be connected with other people.	0.61
Throughout my life, I have had little interest in dating or being in a romantic relationship.	0.60
I generally am not interested in being emotionally close with others.	0.65
There are just not many things that I have ever enjoyed doing.	0.57
I have little or no interest in sex or romantic relationships.	0.70
I greatly enjoy traveling to new places.	0.58
Just being with other people can make me feel good.	0.67
Spending time with close friends and family is important to me.	0.58
Having a meal with other people is almost always better than eating alone.	0.66

Disorganized Schizotypy Items	**Pearson *r***
Most of the time I find it is very difficult to get my thoughts in order.	0.57
No matter how hard I try, I can't organize my thoughts.	0.37
Even when I have time, it is almost impossible to organize my thoughts.	0.51
Most of the time my thoughts seem clear and organized.	0.51
My thoughts are so hazy and unclear that I wish that I could just reach up and put them into place.	0.58
My thoughts almost always seem fuzzy and hazy.	0.58
Things slip my mind so often that it's hard to get things done.	0.45
I have a hard time staying on topic while speaking.	0.53
My thoughts often feel so jumbled that I have difficulty doing anything.	0.66
My thoughts are almost always hard to follow.	0.49
I find that I am very confused about what is going on around me.	0.40
I often find that when I talk to people I don't make any sense to them.	0.52
People find my conversations to be confusing or hard to follow.	0.58
I have trouble following conversations with others.	0.58
When people ask me a question, I often don't understand what they are saying.	0.64
It is usually easy for me to follow conversations.	0.42
My lack of organization often makes it hard to do the things I am supposed to do.	0.50
My thoughts and behaviors are almost always disorganized.	0.43
I often feel so disconnected from the world that I am not able to do things.	0.57

My thoughts and behaviors feel random and unfocused.	0.55
I often have difficulty organizing what I am supposed to be doing.	0.72
When I try to do one thing, I often become confused and start doing something else.	0.54
I often feel so mixed up that I have difficulty functioning.	0.64
I often struggle to stay organized enough to complete simple tasks throughout the day.	0.56
I often have difficulty following what someone is saying to me.	0.45

Note: Italicized items are included in the Multidimensional Schizotypy Scale-Brief

All correlations significant, $p < .05$

Correlations > .30 indicted medium effect sizes and > .50 indicate large effect sizes

Supplementary Table 3. Descriptive Statistics for the Multidimensional Schizotypy Scale (MSS) and Brief Version (MSS-B) for Samples A (*n* = 153), B (*n* = 97), and C (*n* = 89) at Test (Time 1) and Retest (Time 2) Assessments in Study 2

Criterion	Mean	S.D.	Range	Coefficient Alpha	Skew	Standard Error	Kurtosis	Standard Error
SAMPLE A – MSS-B Time 1								
Positive Schizotypy	1.95	2.37	0 – 10	.78	1.38	0.20	1.36	0.39
Negative Schizotypy	2.20	2.86	0 – 11	.85	1.45	0.20	1.32	0.39
Disorganized Schizotypy	1.63	2.76	0 – 12	.89	2.07	0.20	3.84	0.39
SAMPLE A – MSS-B Time 2								
Positive Schizotypy	1.79	2.36	0 – 10	.80	1.46	0.20	1.37	0.39
Negative Schizotypy	2.05	2.81	0 – 12	.86	1.74	0.20	2.48	0.39
Disorganized Schizotypy	1.20	2.57	0 – 12	.91	2.69	0.20	6.83	0.39
SAMPLE B – MSS-B Time 1								
Positive Schizotypy	1.59	2.22	0 – 11	.79	1.73	0.25	3.29	0.49
Negative Schizotypy	1.79	2.70	0 – 13	.86	2.10	0.25	4.53	0.49
Disorganized Schizotypy	1.26	2.42	0 – 9	.88	2.17	0.25	3.68	0.49
SAMPLE B - MSS Time 2								
Positive Schizotypy	2.72	4.03	0 – 17	.89	2.10	0.25	4.15	0.49
Negative Schizotypy	4.35	5.70	0 – 26	.93	1.85	0.25	3.39	0.49
Disorganized Schizotypy	2.41	4.64	0 – 22	.94	2.53	0.25	6.18	0.49
SAMPLE C – MSS Time 1								
Positive Schizotypy	3.04	5.48	0 – 25	.95	2.78	0.26	7.68	0.51
Negative Schizotypy	3.83	5.00	0 – 25	.91	1.81	0.26	3.32	0.51
Disorganized Schizotypy	2.73	5.20	0 – 23	.95	2.72	0.26	7.22	0.51

SAMPLE C – MSS-B Time 2								
Positive Schizotypy	1.47	2.27	0 – 10	.82	1.99	0.26	3.96	0.51
Negative Schizotypy	1.71	2.50	0 – 12	.84	1.79	0.26	2.98	0.51
Disorganized Schizotypy	0.92	2.05	0 – 11	.87	2.88	0.26	8.90	0.51

Supplementary Table 4. Comparison of Participants who Completed Both Assessments and who were Lost to Attrition on T1 MSS and MSS-B scores in Study 2

Sample A – Comparison on T1 MSS-B Scores

Positive Schizotypy: $t(198) = 1.48$, $p=.14$

Negative Schizotypy: $t(198) = 0.74$, $p=.46$

Disorganized Schizotypy: $t(198) = 1.14$, $p=.26$

Sample B – Comparison on T1 MSS-B Scores

Positive Schizotypy: $t(125) = 2.77$, $p<.01$

Negative Schizotypy: $t(125) = 2.64$, $p<.01$

Disorganized Schizotypy: $t(125) = 2.42$, $p<.05$

Sample C – Comparison on T1 MSS Scores

Positive Schizotypy: $t(114) = 0.96$, $p=.34$

Negative Schizotypy: $t(114) = 0.66$, $p=.51$

Disorganized Schizotypy: $t(114) = 1.51$, $p=.14$

Supplementary Table 5. Correlations of the MSS and MSS-B Subscales at T1 and T2 for Samples A (*n* = 153), B (*n* = 97), and C (*n* = 89) in Study 2.

Correlation	Sample A Time 1 MSS-B	Sample A Time 2 MSS-B	Sample B Time 1 MSS-B	Sample B Time 2 MSS	Sample C Time 1 MSS	Sample C Time 2 MSS-B
Positive Schizotypy- Negative Schizotypy	.13	.00	.13	.21*	.28*	.11
Positive Schizotypy – Disorganized Schizotypy	**.35***	**.28***	**.35***	***.52****	***.65****	***.62****
Negative Schizotypy – Disorganized Schizotypy	**.48***	**.30***	**.38***	**.38***	**.45***	.14

* $p<.05$

medium effect sizes in bold, large effects in bold and italics