

The Maladaptive Schema Scale (MSS): Development and Validation of a Comprehensive Questionnaire for Beliefs Related to Psychopathology

Assessment
1–16
© The Author(s) 2025



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10731911251390083
journals.sagepub.com/home/asm



Ben Buchanan^{1,2}, Emerson Bartholomew^{2,3}, Carla Smyth², and David Hegarty^{2,4}

Abstract

The Maladaptive Schema Scale (MSS) was developed to assess dysfunctional cognitive frameworks linked to psychopathology, including personality disorders, trauma, and relational issues, using contemporary theoretical frameworks, addressing limitations in existing schema measures. This study aimed to validate the MSS, evaluate newly proposed schemas, and establish its psychometric properties using Rasch methodology. The scale was assessed in clinical and nonclinical respondents ($n = 2,182$) for overall and item fit, dimensionality, reliability, and measurement invariance. All 27 MSS schemas had an acceptable overall fit to the Rasch model, no item misfit, no local dependence, evidence of strict unidimensionality, measurement invariance by sex, age, time taken and clinical group, and convergent validity with the Young Schema Questionnaire (YSQ). The MSS is a valid, reliable, and comprehensive tool for assessing maladaptive schemas in clinical and research settings, offering advantages in both brevity and breadth over traditional schema measures.

Keywords

maladaptive schema, Rasch analysis, psychometric validation, schema therapy, psychopathology

A “schema” can be broadly defined as an organizing principle that an individual uses to make sense of their experiences (Young et al., 2003). Schemas are a central theoretical construct and focus of clinical interventions in numerous psychotherapeutic treatment models, including schema therapy (Young et al., 2003) and cognitive behavioral therapy (Moorey et al., 2020; Padesky, 1994). Literature reviews provide general support for the association between maladaptive schemas and adult psychopathology (e.g., Pilkington et al., 2023; Yalcin et al., 2020), supporting the importance of identifying and targeting these constructs in psychological therapy.

Schema therapy is an integrative treatment model developed by Jeffery Young and colleagues in response to the limitations of classical therapeutic approaches encountered in treating individuals with complex difficulties, such as personality disorders (Young et al., 2003). In this model, a schema can be defined as an organizing principle composed of cognitions, emotions, body sensations, and memories (Young et al., 2003). Maladaptive schemas are thought to form as a result of early life aversive experiences and underpin the chronic nature of some psychological disorders (Young et al., 2003). Identifying these maladaptive schemas is foundational to case conceptualization and treatment

(Green & Balfour, 2020). Self-report measures are a key data point in schema assessment, highlighting the importance of appropriate assessment tools that are both evidence-based and psychometrically robust. The Young Schema Questionnaire (YSQ; Young, 1990) is a widely used schema assessment instrument, particularly in schema therapy and more broadly in clinical practice and research.

Several groups have identified the need for a research agenda to address gaps in the theoretical and empirical evidence base for the schema therapy model (e.g., Arntz et al., 2021; Sempertegui et al., 2013), including the measurement and conceptualization of schemas (Pilkington et al., 2023). Existing measures such as the YSQ (Young, 1990) are lengthy, with more than 200 items, and have had mixed factor analysis results (Pilkington et al., 2023). A 2007 review

¹Monash University, Melbourne, Victoria, Australia

²NovoPsych Psychometrics, Melbourne, Victoria, Australia

³University of Auckland, New Zealand

⁴Southern Cross University, Coffs Harbour, New South Wales, Australia

Corresponding Author:

Ben Buchanan, NovoPsych Psychometrics, Level 1, 66 Whitehorse Road, Deepdene, Melbourne, Victoria 3103, Australia.
Email: ben@novopsych.com

of the psychometric properties of the YSQ concluded that the measurement of schemas remains challenging and recommended that, while the YSQ holds promise as an assessment tool, it should be used with caution in clinical settings, with further research needed (Oei & Baranoff, 2007). The YSQ and later derivatives use constructs that have not received a substantial revision and integrated more contemporary research, such as trauma research (see Substance Abuse and Mental Health Services Administration, 2014, for a review), developments in attachment theory (see Fraley, 2019, for a review), and empirical psychometric methods. Indeed, a recent consensus study of clinicians and researchers conducted by Pilkington et al. (2023) highlighted schema constructs and measures as priority areas for future research.

Recent research has sought to refine existing schema measures and address their limitations. For example, a 2020 factor analysis of the 232 items in the Young Schema Questionnaire Long-Form 3 (YSQ-L3) conducted by Yalcin et al. (2020) found 20 schemas rather than 18 as proposed by Young et al. (2003). The authors accordingly proposed revisions to the structure of the YSQ, including the separation of the punitiveness (into self and other) and emotional inhibition (into emotional constriction and fear of losing control) schemas. Recently, Rasch analysis was used in YSQ-L3 items to reduce the number of items, creating a revised version (YSQ-R). The YSQ-R showed appropriate individual item fit statistics for 116 items and measurement invariance by the clinical group. Yet, tests of the overall fit of schemas to the Rasch model were not reported, along with unidimensionality, response category threshold distributions, or measurement invariance by sex or age group (Yalcin et al., 2022).

In addition, there has been recognition of the need to integrate developments in research and theory to update the foundational principles of schema therapy and ensure that included constructs are empirically grounded. Arntz et al.'s (2021) international workgroup and position paper incorporated Dweck's (2017) unified theory of motivation, personality, and development into a reformulated schema therapy theory; this formed the basis of three proposed new schema constructs: Unfairness, Lack of Coherent Identity, and Lack of Meaningful World. These constructs have not yet been empirically tested but are suggested by the authors to provide a more comprehensive model for understanding areas of psychopathology not previously encompassed by schema theory and measures, such as dissociation and psychosis (Arntz et al., 2021).

Further theoretical and research developments are relevant to the development of a contemporary and comprehensive schema taxonomy. For example, adult separation anxiety disorder has only recently been recognized (Manicavasagar et al., 2000), and attachment theory, which once conceptualized attachment as a categorical construct,

has more recently been reconceptualized as a nuanced, dimensional construct (e.g., Fraley et al., 2015; Karantzas et al., 2010), pointing to a likely greater variance in attachment-related schema than is captured by current schema measures. A recent meta-analysis found higher levels of association between the YSQ's Abandonment schema and anxious attachment compared with avoidant attachment (Karantzas et al., 2023). As such, the inclusion of a construct reflecting attachment avoidance would provide a more comprehensive representation and assessment of the various maladaptive internal models that may result from the absence of early secure attachment and broaden the clinical utility of schema assessments. Furthermore, constructs associated with psychopathology and of clinical utility, such as self-efficacy and locus of control, do not currently appear to be represented in existing schema measures. Their integration would enable a more comprehensive approach to assessment and treatment.

Conceptual and empirical issues with current schema measures outlined above raise the need for an updated, unified, and comprehensive schema assessment tool. The first aim of the present study was to synthesize traditional early maladaptive schemas (Young, 1990; Young et al., 2023) with the aforementioned developments in theory and research to produce a contemporary, open-source schema measure, the Maladaptive Schema Scale (MSS; NovoPsych 2025).

The second aim of this research was to validate the MSS and its proposed constructs empirically using Rasch analysis. Rasch analysis is capable of investigating a range of psychometric properties not assessed by classical test theory analyses. This approach to psychometric assessment is based on an estimation of both item difficulty and person ability (Rasch, 1960). While more complex IRT models estimate both difficulty and discrimination parameters, we selected the Rasch model for several reasons. First, while the Rasch model has stringent requirements, if these can be met, this allows for greater confidence in the measurement properties of an instrument including sample-independent item calibration, invariant comparisons between populations, and the mathematical foundations for true interval-level (rather than ordinal-level) measurement properties. The Rasch assumption of equal item discrimination functions as a testable hypothesis that can support stronger inferences about scale properties. More complex models may achieve better fit by conforming to sample-specific response patterns through additional parameters, but this increases vulnerability to overfitting and can compromise the instrument's clinical utility and interpretability. Fit to the Rasch model indicates that a scale is approximately in alignment with the fundamental principles of measurement as defined by Thurstone (1928) of invariance, unidimensionality, and concatenability. The current study provides the initial validation of the MSS using Rasch methodology and confirmatory factor analysis (CFA).

Method

Participants

A total of 2,182 individuals in five samples contributed data between January and October 2024. All responses were de-identified and screened for authenticity by removing outliers on time taken to complete (<7.5 and ≥ 83 minutes were removed) and insufficient effort responding using the intra-individual response variability (IRV) index which detects low response variability (Dunn et al., 2018). Participants with IRV values in the lowest 5% were removed as this threshold effectively identifies cases with abnormally constrained response patterns while maintaining adequate sample representation. Across all samples, there were no missing data.

Sample 1. MSS data were collected from practicing psychologists in Australia administering a pilot version of the MSS to clients in routine care from May to August 2024. A total of 210 responses were screened for authenticity using the time taken to complete or unrealistic response patterns. Seventeen responses were excluded, leaving 193 full sets. These data were used for the initial validation of the MSS. The sample had a mean age of 36.18(11.97), ranging from 18 to 78 years, and consisted of 54% female, 30% male, and 16% not reported.

Sample 2. A second set of MSS data was collected from practicing psychologists in Australia administering a pilot version of the MSS to clients as part of routine care, as well as those who anonymously self-administered the MSS after having located it on a public website from August to September 2024. A total of 274 responses were screened to 210. This sample was 16% male, 37% female, and 47% did not specify. Ages ranged from 18 to 67 for this sample, with a mean of 31.07 (10.80). For Samples 1, 2, and 3 used in Rasch analysis, sizes ranged from slightly lower to slightly higher than the recommended range of 250 to 500 (Hagell & Westergren, 2016). While falling outside the recommended ranges, these samples are still satisfactory for Rasch analysis. Simulations by Linacre (1994) show samples as low as 50 participants provide 99% confidence of stable item calibrations within ± 1 logit, and Samples 2 and 3 are within the 99% confidence range for stable item calibrations within $\pm .50$ logits.

Sample 3. An additional sample of MSS data was taken consisting of clients administered the MSS in three different contexts: (a) routine care by practicing psychologists in Australia, (b) psychologists self-administering the MSS, and (3) those who anonymously self-administered the MSS after having located it on a public website from September to October 2024. For the routine care context, practicing psychologists from private clinical settings in Australia were invited to

administer the MSS to their clients as part of their assessment protocols. The way the MSS was integrated into routine care varied by clinician on their discretion; however, all administrations occurred through the same secure online psychometrics platform, NovoPsych, with standardized digital instructions for participants. For the two self-administration contexts, practicing psychologists were invited to self-administer the MSS through the same platform to familiarize themselves with the instrument before administering it to their clients. Anonymous website respondents who discovered the assessment through online search accessed the scale via the same secure online psychometrics platform and completed it for personal interest, and were given a copy of their provisional results. No respondents were incentivized or compensated for participation in any context. A total of 1,257 responses were cut down to 1,018 based on the screening criteria described earlier, to be used for confirmatory factor modeling. This sample of 1,018 was 46.31% clinical, and 53.69% nonclinical respondents. A clinical-only subset ($n = 444$) was taken from the larger sample ($n = 1,018$) for the purposes of bias testing by age, sex, and time taken. This sample had a mean age of 37.46 (12.21), a range of 18 to 73 and was 32% male, 59% female, and 8% not specifying. Clinical respondents were identified as MSS administrations in routine care by psychologists in Australia. To complete further bias testing by clinical versus nonclinical group, 250 respondents were randomly selected from the 484 clinical and 574 nonclinical respondents ($n = 500$). Nonclinical respondents were identified as psychologists self-administering. Nonclinical respondents also included those who anonymously self-administered the MSS after having located it on a public website from September to October 2024.

Sample 4. A final sample of MSS data was collected, with respondents administered the final iteration of the MSS as well as the YSQ-R in two different contexts: (a) psychologists self-administering the MSS and YSQ-R and (b) those who anonymously self-administered the MSS and YSQ-R after having located them on a public website. A total of 53 respondents were screened with 51 remaining based on the criteria outlined above.

For clarity around the samples, for samples 1 ($n = 710$) and 2 ($n = 193$), all respondents were considered clinical as they were administered assessments in routine care by practicing psychologists. In samples 3 ($n = 210$), 4 ($n = 1,018$), and 5 ($n = 51$), respondents were classified as clinical if they were administered the scale in routine care by psychologists, whereas those who self-administered the MSS (or YSQ-R), either psychologists or anonymous individuals, were classified as nonclinical.

Preliminary Sample. Data for a preliminary analysis of the YSQ-R that informed the creation of the MSS were collected from practicing psychologists in Australia

administering only the YSQ-R to clients in routine care from January to May 2024. A total of 810 responses were reduced to 710 by the method outlined above. This sample consisted of 60.2% females, 35.8% males, and 4% not reporting; the mean age was 36.50 (12.34) years.

Procedure

Respondents in the clinical samples consisted of clients voluntarily seeking therapy and receiving the pilot versions of the MSS as part of routine care, where their clinicians had consented to the use of de-identified data for scale development. Therapists were not incentivized to administer the MSS, and data collection was conducted via a psychometrics platform, a secure online system used for psychological assessments. All data aggregation and analysis adhered to ethical requirements set out in the approval granted by the Human Research Ethics Committee (HREC) at Monash University. The HREC granted a waiver of consent for clients as the research involved de-identified secondary data analysis.

Measures

The MSS was created to assess the strength of belief in early maladaptive schemas. An extensive literature search was conducted to identify schemas, core beliefs, or world views associated with psychopathology. Research by Mussel (2023) using natural language processing techniques systematically identified and organized constructs related to schemas. To identify existing measures of maladaptive schemas, literature on the theoretical underpinnings of schema therapy was reviewed, and newly identified constructs, such as unfairness, lack of coherent identity, and lack of meaningful world (Arntz et al., 2021), were included in the shortlist of schemas. The results of recent analyses of the YSQ (Yalcin et al., 2020, 2022, 2023) were reviewed and proposed revisions to schema structures, such as the separation of the punitiveness schema (into self and other), were included in the short list.

In total, 30 existing psychometric instruments were reviewed, with the following scales being among those to be of key importance to item creation: YSQ (Young, 1990), Word Assumption Scale (Janoff-Bulman, 1989), Attachment Style Questionnaire-Short Form (Chui & Leung, 2016), and General Self-Efficacy Scale (Schwarzer & Jerusalem, 1995). From this process, an initial 138-item pool was developed by three PhD-trained psychologists and one psychometrician. Items reflect the synthesis of concepts found in the above scales, schemas identified by Mussel (2023), Arntz et al. (2021), and Dweck's (2017) unified theory of personality and development.

The MSS requires participants to indicate the strength of agreement to questions related to each schema, for

example, "I fear that my important relationships will end unexpectedly." A 5-point Likert-type scale was used, ranging from 0 = *strongly disagree* to 4 = *strongly agree*. Reverse scoring was used for 17 items. Total scores for each schema were calculated by adjusting for reverse scored items, and then summing the four items in a given schema for a total score. Higher scores indicate a stronger alignment with a maladaptive schema.

YSQ-R (Yalcin et al., 2022) is a 116-item measure of 20 schemas derived from a longer questionnaire, the YSQ-L3. The YSQ-R uses a 6-point Likert-type scale from 1 = *completely untrue* to 6 = *describes me perfectly*.

Data Analyses

IBM Statistical Package for the Social Sciences (SPSS) v27 was used for descriptive statistics, while Rasch analysis was facilitated by RUMM2030plus (Andrich et al., 2009). RStudio version 1.4.1717 was used for CFA modeling using the "lavaan" package (Rosseel, 2012).

To begin, a likelihood ratio test indicates which Rasch model was most suitable for the data (Lundgren-Nilsson & Tennant, 2011; Masters, 1982). Rasch analysis was completed in an iterative manner where the fit of the observed data to the Rasch model was examined by eight key indicators. (a) Overall model fit was assessed by the item-trait interaction chi-square. A nonsignificant value suggests that the scale functions appropriately across varying levels of schema belief strength (Gustafsson, 1980; Tennant & Conaghan, 2007). Ideally, item fit residual values (b) should fall within a range of -2.50 to $+2.50$ (Andrich, 2016). Unidimensionality, (c) a fundamental principle of measurement established by Thurstone (1928) and an assumption of the Rasch model (Rasch, 1960) was assessed by Smith's (2002) test. Item characteristic curves (ICCs) are used to investigate category threshold ordering (d). Sample targeting (e) assessed appropriate coverage of the abilities of the persons in the sample by the scale items. Ideal values fall between $+0.50$ and -0.50 for an item mean of 0, although it is argued that within 1 logit is also indicative of good targeting (Finger et al., 2012; Medvedev & Krägeloh, 2022). Local response dependence (f) is a potential issue that can result in increased measurement error and the distortion of both dimensionality and reliability (Fisher, 1992). Local dependence was assessed by examining residual correlations for values above .20 (Christensen et al., 2017). Items above this threshold are typically considered for removal or combined into subtests (Andrich et al., 2009). Scale reliability (g) was assessed by Cronbach's alpha (α), McDonald's omega (ω), the Person Separation Index (PSI), and the average inter-item correlation (IIC). PSI reflects the capacity of a scale to differentiate between varying levels of ability possessed by different individuals (Fisher, 1992; Medvedev & Krägeloh, 2022). Both omega and PSI use the same generally accepted .70 cutoff

Table 1. Higher-Order Category Organization for the MSS.

Schema	Needs	Focus
Abandonment/anxious attachment	Safety and attachment	Relationships
Excessive self-reliance/avoidant attachment	Safety and attachment	Relationships
Emotional deprivation	Safety and attachment	Relationships
Mistrust of others	Safety and attachment	Other people
Others are dangerous/malevolent	Safety and attachment	Other people
Social isolation/outsider	Safety and attachment	Relationships
Defectiveness/shame	Safety and attachment	Inadequate self
Vulnerability to dangerous world	Safety and attachment	World
Dependence	Autonomy and competence	Relationships
Failure/achievement inferiority	Autonomy and competence	Inadequate self
Low self-efficacy/weakness	Autonomy and competence	Inadequate self
Fatalistic/external locus of control	Autonomy and competence	Inadequate self
Enmeshment/diffuse boundaries	Autonomy and competence	Relationships
Subjugation/submission to others	Needs, opinions, and emotions	Relationships
Self-sacrifice	Needs, opinions, and emotions	Relationships
Approval-seeking/excessive need to be liked	Needs, opinions, and emotions	Relationships
Emotional inhibition	Needs, opinions, and emotions	Inadequate self
Pessimism/negativity	Spontaneity and play	World
Unrelenting standards	Spontaneity and play	Inadequate self
Punitiveness/unforgiving of self	Spontaneity and play	Inadequate self
Punitiveness/unforgiving of others	Spontaneity and play	Other people
Entitlement/specialness	Realistic and consistent limits	Inflated self
Full Control	Realistic and consistent limits	Inflated self
Over Reliance on Emotions	Realistic and consistent limits	Inflated self
Unfairness	Coherence and fairness	World
Meaningless World	Coherence and fairness	World
Lack of Coherent Identity	Coherence and fairness	Inadequate self

threshold as alpha (Tennant & Conaghan, 2007). The IIC was included as it assesses reliability while being robust to the number of scale items. Acceptable IIC values range between .15 and .50 (Clark & Watson, 1995). This range was suitable for schemas as they represent intermediate-level psychological constructs that are more specific than broad personality dimensions (which would warrant lower IICs of .15–.20) yet more inclusive than narrow behavioral indicators (which would require higher IICs of .40–.50). Measurement invariance (h) was established through differential item function (DIF) testing. If respondents from different person-level groups (e.g., age, sex, country) who possess the same latent trait level are observed to have significantly different response patterns to the same items, DIF would be indicated (Lundgren-Nilsson et al., 2013).

Analysis of Sample 1 ($n = 193$) was completed first followed by Sample 2 ($n = 240$). Sample 2 was used to assess added and reworked schemas. Next, confirmatory factor modeling was carried out on sample 3 ($n = 1,018$), to test the fit of items to schemas and schemas to two sets of higher-order organizing structures, one based on early unmet needs, and the other on relational focus (Table 1). Clinical and nonclinical norms were derived from subsets

of the larger MSS sample 3 ($n = 1,018$), and this sample was used to create subsets for further DIF testing by age (using three even groups; 18–30, 31–42, and ≥ 43), sex, and time taken to complete the MSS (using three even groups; <631 , 632–875, and ≥ 876 seconds). To test DIF by clinical versus nonclinical grouping, a random sample ($n = 250$) was taken from the clinical and nonclinical portions of the larger Sample 3 to ensure balanced numbers. Intercorrelation patterns were assessed between MSS schemas using the full sample 3 ($n = 1,018$), and then convergent validity was assessed between comparable MSS and YSQ-R schemas using sample 4 ($n = 51$), in which respondents had been assessed with both scales together. Regarding Sample 4, as neither the MSS nor YSQ-R are used to produce total scores, and with the aim of detecting moderate or stronger correlations ($r \geq .40$) between 4-item constructs, an a priori power analysis was conducted using G*Power 3.1. With parameters of $\alpha = .05$ (one-tailed), desired power of .80, and an expected correlation of .40, the analysis indicated a minimum sample size of $n = 37$. Our obtained sample of $n = 51$ provides approximately 91% statistical power to detect moderate or larger correlations between MSS and YSQ-R schemas. This level of power is appropriate given

that convergent validity between parallel measures of the same construct should demonstrate moderate to large effect sizes (Cohen, 1988).

Preliminary analysis of the YSQ-R data sample ($n = 710$) is reported in the supplementary materials (S1). This analysis served to inform the creation of the MSS and influenced it in several ways. The failure of 18 of 20 YSQ-R schemas to achieve acceptable Rasch model fit indicated the need for more tightly focused schema constructs. This informed our decision to aim for 4 items per schemas as the goal during the development of the MSS. Furthermore, the high number of observed disordered thresholds across the 6-point response scale of the YSQ-R suggested that respondents had difficulty reliably distinguishing between six levels of schema endorsement, guiding our decision to implement a 5-point Likert-type scale for the MSS. Furthermore, item issues such as the presence of local response dependencies (12 instances affecting 24 items) and observed item difficulty distributions informed MSS item creation.

Results

For Samples 1, 2, 3, and 4 where Rasch analysis was used, a likelihood ratio test indicated the unrestricted Partial Credit model was most suitable (Lundgren-Nilsson & Tennant, 2011; Masters, 1982). Results of the preliminary analysis of the YSQ-R are reported in the supplementary materials (S1).

Rasch Analysis of the MSS

Initial analysis of the 138 items in the first iteration of the MSS with sample 1 ($n = 193$) revealed an overall fit to the Rasch model for 22 of the 26 subscales tested, indicated by a nonsignificant item-trait interaction chi-square value (Table 2). Subscales that showed significant misfit included others are Dangerous, Entitlement, Unfairness, Meaningless World, and Lack of Coherent Identity. Reliability values were generally acceptable, with 22 schemas above the .70 PSI threshold. Schemas with below threshold reliabilities included Enmeshment, Entitlement, and Full Control (Table 2). Item fit values showed two misfitting items; Item 5 for Others are Dangerous; and Item 1 for Entitlement.

Local response dependency was observed in several instances, with residual correlations above .20 for Emotional Deprivation Items 3 and 4 (.18), Dependence Items 5 and 6 (.29), Entitlement Items 3 and 4 (.24), Unfairness Items 2 and 3 (.25), and Meaningless World Items 1 and 3 (.19). DIF by sex was assessed by values and inspection of plots, with no significant DIF observed across any items. The majority of schemas showed evidence for strict unidimensionality, with exceptions being Approval-Seeking, Punitiveness-Other, Entitlement, and Lack of Coherent Identity.

A process of item deletion was undertaken to make the MSS shorter, with the aim of four items per subscale. A total of 22 items were removed from various subscales, while no deletion was attempted for those subscales that began with four items such as Mistrust of Others, Failure, Fatalistic, Self-Sacrifice, Pessimism, and Punitiveness-Self. The item deletion process was guided by multiple empirical indicators and in consideration of content coverage. A systematic approach was followed using such criteria as item fit statistics (significant misfit outside of the ± 2.50 range), residual correlations exceeding .20, indicative of local response dependence issues, principal component loading strength, and item difficulty distributions (Wright maps), with a focus on balancing difficult and easy items, biased toward removing easier items given the intended use of the MSS for clinical populations. Impact of the PSI reliability was also considered when removing items, particularly when closer to the desired 4-item scale.

Before item deletion, the MSS had 138 items covering 26 maladaptive schemas with each schema having between 4 and 6 items each. After item deletion, overall model fit and reliability remained stable in those schemas that previously achieved fit, and was improved substantially in several cases of previous misfit, such as Others are Dangerous, $\chi^2(20) = 23.09, p = .246$; Unfairness, $\chi^2(24) = 22.54, p = .547$; Meaningless World, $\chi^2(28) = 35.40, p = .158$; and Lack of Coherent Identity, $\chi^2(28) = 22.54, p = .755$. Reliability was improved to an acceptable level for Punitiveness – Others (PSI = .71). The Entitlement and Enmeshment schemas had no item deletion solution that resulted in a substantial improvement in overall fit or reliability. Entitlement aside, no local response dependency was observed in subscales with previous residual correlations exceeding .20—Emotional Deprivation, Dependence, Unfairness, and Meaningless World. Item misfit was not observed in any of the modified schemas, or those that began with four items. After item deletion, subscales which achieved strict unidimensionality retained this, and all subscales that previously did not, subsequently showed evidence for strict unidimensionality, such as Approval Seeking, Punitiveness–Others, and Lack of Coherent Identity.

Second Rasch Analysis of the MSS

Results of the second MSS analysis, using sample 2 ($n = 210$), are marked with an asterisk in Table 2. This analysis added the schema Over Reliance on Emotions and revised versions of poorly performing subscales Entitlement and Enmeshment. The initial 6-item Over Reliance on Emotions subscale showed an acceptable overall model fit, $\chi^2(18) = 23.09, p = .187$, and reliability values of $\alpha = .68, \omega = .81$, and PSI = .72. Evidence of unidimensionality was observed,

Table 2. Summary of Fit Statistics for the Initial and Final Schemas.

Analyses by Schema	Person location		Goodness of fit					DIF	Unidimensionality		
	M	SD	$\chi^2(df)$	p	Alpha	Omega	PSI	YES/NO	%	^a LB95% CI	Achieved
Abandonment											
Initial	-0.14	1.21	36.46(30)	0.193	0.79	0.84	0.77	NO	2	1	YES
Final	-0.65	1.49	36.03(28)	0.142	0.79	0.85	0.76	NO	2	1	YES
Excessive self-reliance											
Initial	0.32	0.99	37.93(35)	0.337	0.72	0.76	0.71	NO	1	1	YES
Final	0.53	1.22	20.25(28)	0.855	0.73	0.77	0.70	NO	1	1	YES
Emotional deprivation											
Initial	-0.87	1.58	56.25(54)	0.390	0.84	0.89	0.84	NO	4	1	YES
Final	-1.05	1.85	28.89(24)	0.224	0.80	0.84	0.80	NO	1	1	YES
Mistrust of others											
Initial	-0.03	2.38	8.59(16)	0.929	0.87	0.88	0.86	NO	2	1	YES
Others are dangerous											
Initial	-0.77	1.28	85.62(25)	<.01	0.75	0.82	0.75	NO	5	1	YES
Final	-0.93	1.93	23.92(20)	0.246	0.81	0.86	0.80	NO	5	1	YES
Social isolation											
Initial	-0.14	1.19	36.69(30)	0.186	0.84	0.87	0.83	NO	4	1	YES
Final	-0.42	1.42	27.18(24)	0.296	0.75	0.80	0.75	NO	3	1	YES
Defectiveness											
Initial	-0.98	1.37	53.85(48)	0.260	0.82	0.86	0.81	NO	3	1	YES
Final	-0.48	1.67	37.19(32)	0.242	0.83	0.86	0.80	NO	2	1	YES
Vulnerability											
Initial	-0.69	1.76	23.93(15)	0.066	0.86	0.89	0.84	NO	4	1	YES
Final	-0.65	1.61	15.97(12)	0.192	0.80	0.85	0.78	NO	1	1	YES
Dependence											
Initial	-1.35	1.59	28.87(24)	0.225	0.87	0.92	0.81	NO	3	1	YES
Final	-1.43	1.68	37.44(28)	0.109	0.83	0.90	0.73	NO	1	1	YES
Failure											
Initial	0.01	1.63	29.20(20)	0.083	0.83	0.86	0.80	NO	5	1	YES
Low self-efficacy											
Initial	-0.96	1.69	28.85(25)	0.269	0.84	0.88	0.82	NO	2	1	YES
Final	-1.11	1.86	22.08(20)	0.336	0.81	0.85	0.80	NO	3	1	YES
Fatalistic											
Initial	-1.03	1.51	16.47(16)	0.420	0.75	0.78	0.73	NO	3	1	YES
Enmeshment											
Initial	-0.22	0.74	31.39(25)	0.176	0.57	0.67	0.57	NO	4	1	YES
Final ^a	-0.24	1.09	9.11(12)	0.694	0.68	0.70	0.68	NO	2	1	YES
Subjugation											
Initial	-1.62	1.41	18.47(25)	0.822	0.76	0.79	0.75	NO	4	1	YES
Final	-1.76	1.52	13.79(20)	0.840	0.74	0.84	0.71	NO	2	1	YES
Self-sacrifice											
Initial	0.32	1.39	11.29(16)	0.791	0.76	0.80	0.74	NO	4	1	YES
Approval seeking											
Initial	-0.44	1.57	32.75(30)	0.333	0.87	0.92	0.85	NO	7	3	YES
Final	-0.70	1.68	22.79(20)	0.299	0.82	0.84	0.80	NO	3	1	YES
Emotional inhibition											
Initial	-0.38	1.31	18.65(25)	0.813	0.79	0.84	0.78	NO	3	1	YES
Final	-0.58	1.39	24.90(20)	0.205	0.76	0.84	0.74	NO	3	1	YES
Pessimism											
Initial	0.12	1.24	14.24(16)	0.580	0.74	0.77	0.72	NO	2	1	YES

(continued)

Table 2. (continued)

Analyses by Schema	Person location		Goodness of fit		Alpha	Omega	PSI	DIF	Unidimensionality		
	M	SD	$\chi^2(df)$	p				YES/NO	%	^a LB95% CI	Achieved
Unrelenting standards											
Initial	0.37	1.37	25.45(25)	0.437	0.78	0.82	0.79	NO	4	1	YES
Final	0.34	1.38	17.88(20)	0.595	0.73	0.78	0.74	NO	4	1	YES
Punitiveness—self											
Initial	-0.31	1.67	11.03(16)	0.807	0.76	0.79	0.78	NO	2	1	YES
Punitiveness—other											
Initial	-1.30	1.25	20.96(25)	0.695	0.69	0.78	0.71	NO	8	2	YES
Final	-1.54	1.54	24.95(20)	0.203	0.71	0.74	0.71	NO	4	1	YES
Entitlement											
Initial	-1.15	0.85	115.53(48)	<.01	0.59	0.77	0.62	NO	7	2	YES
Final ^a	-1.29	1.51	6.63(4)	0.157	0.80	0.84	0.72	NO	2	1	YES
Full Control											
Initial	-0.28	0.97	18.55(20)	0.551	0.56	0.22	0.59	NO	2	1	YES
Unfairness											
Initial	-0.75	1.28	76.56(45)	<.01	0.74	0.82	0.76	NO	4	2	YES
Final	-0.33	1.39	22.54(24)	0.547	0.68	0.87	0.75	NO	1	1	YES
Meaningless World											
Initial	-1.33	1.63	57.22(40)	0.03	0.84	0.91	0.79	NO	5	1	YES
Final	-1.5	1.92	35.40(28)	0.158	0.86	0.85	0.79	NO	2	1	YES
Lack of Coherent Identity											
Initial	0.12	1.24	206.58(80)	<.01	0.82	0.85	0.79	NO	7	1	YES
Final	0.09	1.71	22.54 (28)	0.755	0.85	0.88	0.81	NO	2	1	YES
Over Reliance on Emotions											
Initial ^a	-0.41	0.97	23.09 (18)	0.187	0.68	0.81	0.72	NO	8	3	YES
Final ^a	-0.56	1.19	15.83(12)	0.199	0.65	0.87	0.68	NO	5	1	YES

^aSecond analysis with a revised version of the MSS, with new items and with sample 2 (n = 210).

but strict unidimensionality was not supported, with 8% of *t*-tests falling outside the -1.96 to 1.96 range. After the deletion of two items, similar overall model fit and reliability values observed ($\alpha = .65$, $\omega = .87$, $\text{PSI} = .68$), and strict unidimensionality was confirmed. For the revised subscales, Entitlement improved from a poor to acceptable overall model fit evidenced by a change from a nonsignificant to a significant item-trait interaction chi-square value—reliability also improved across all three indices (Table 2). Entitlement saw an improvement from its initial version with evidence of strict unidimensionality, and improved reliability ($\alpha = .57/\alpha = .68$, $\omega = .67/\omega = .70$, $\text{PSI} = .57/\text{PSI} = .68$), in addition to achieving acceptable overall model fit (Table 2).

Category threshold assessment via ICCs revealed that for the response options on the MSS's 5-point Likert-type scale (0, 1, 2, 3, and 4), the third category (Option 2) was never modal for 28 items out of 108. This suggests that respondents may not be able to reliably distinguish between more than four levels of schema endorsement. One solution is collapsing the third response category; however, eight schemas do not have this issue and of the remaining 15,

seven observe disorder on only one item. Collapsing response categories results in a loss of information, and can create inconsistencies between items, within subscales and between subscales.

Figure 1 shows the person-item threshold distribution for the final MSS abandonment schema, demonstrating good alignment between person abilities and item threshold coverage. The person distribution shows a near-normal distribution centered at $.073$ logits ($SD = 1.108$), with the majority of respondents falling within the -1 to $+1$ logit range. The item threshold distribution covers respondents well, indicating adequate measurement precision across the continuum of Abandonment schema endorsement. Good sample targeting was observed in general, with 13 schemas within $\pm .50$, and 9 schemas within 1 logit. Five schemas were greater than 1 logit; Dependence, Subjugation, Punitiveness—Others, Entitlement, and Meaningless World (Supplementary Material S2). Of these <1 logit schemas, the percentage of persons grouped outside of the item range did not exceed 20%, indicating the absence of floor effects (Murugappan et al., 2022). However, the Meaningless World subscale was close to this threshold, with

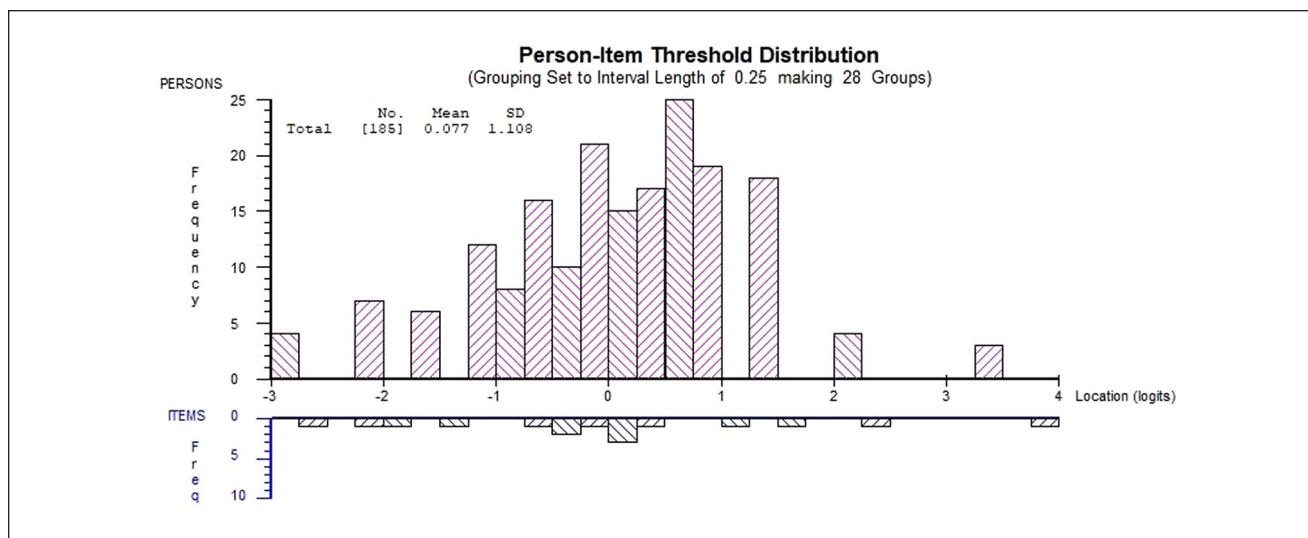


Figure 1. Person-Item Threshold Distribution for the Final MSS Abandonment Schema.

19% outside the item range on the lower end of the ability distribution—bordering on a floor effect. Further details can be seen in the supplementary material (S2), where it can be generally observed that the sample’s abilities are well covered by schema items.

Convergent Validity Between the MSS and YSQ-R

Convergent validity was assessed using sample 4 ($n = 51$) between comparable MSS and YSQ-R schemas. Each MSS schema displayed a significant, positive correlation to its YSQ-R counterpart with an average of .680, supporting convergent validity between the two measures (Table 3). High correlations were observed for Abandonment/Anxious Attachment (.889), Defectiveness/Shame (.858), and Entitlement/Specialness (.770) in particular, suggesting close conceptual alignment between these constructs across instruments. The schema showing the lowest correlation was between Emotional Inhibition and its counterpart Emotional Constriction (.367).

Intercorrelations Between MSS Schemas

Within the MSS, between-schema correlation patterns were also examined (using sample 3, $n = 1018$) and are detailed in Table 4. Notable convergent correlations include Low-Self Efficacy and the subscales Dependence (.74), Subjugation (.62), and Pessimism (.62). Further notable findings are observed with Punitiveness—Self and Defectiveness (.53), Low Self-Efficacy and Fatalism (.45), and Pessimism and Meaningless World (.63). Examples of divergent correlations can be seen between Emotional

Inhibition and Over Reliance on Emotions (−.31), and Full Control and Fatalistic (−.28). Furthermore, Full Control showed negative correlations to several other schemas such as Failure, Defectiveness, Dependence, Subjugation, Pessimism, and Lack of Coherent Identity. Exhaustive correlation matrices are provided in the supplementary materials (S3) for convergent/divergent MSS and YSQ-R schemas, and intercorrelations between schemas within the MSS (S4).

Lack of coherent identity had the highest standard deviation and was associated with greater variance in responding across other subscales. For example, people in the lowest quartile for lack of coherent identity were more consistent in their responses on other subscales, whereas those in the top quartile had more variation in their responses. Specifically, the mean variance of responses within other subscales between participants scoring in high and low quartiles on lack of coherent identity was compared using an independent sample t -test. Mean variances were obtained for high scorers (.81) and low scorers (.70), with the t -test results, $t(50.18) = 1.77, P = .04$, indicating that high scorers had significantly greater response variability across other subscales.

Confirmatory Factor Modeling, DIF Testing, and Normative Data

Results of CFA modeling for the MSS subscales into organizing structures using Sample 3 data ($n = 1018$) are presented in Table 5. MSS items mapped to their respective schemas, with excellent RMSEA and SRMR values. The fit of MSS subscales to broader organizing structures showed a similarly strong fit, with the “relational focus” structure fitting slightly better than subscales to “early childhood

Table 3. Convergent Correlations of Comparable MSS and YSQ-R Schemas (n = 51).

MSS	YSQ-R	Correlation
Abandonment/anxious attachment	Abandonment	.889**
Emotional deprivation	Emotional deprivation	.608**
Mistrust	Mistrust	.708**
Social isolation/outsider	Social isolation	.775**
Defectiveness/shame	Defectiveness	.858**
Vulnerability to dangerous world	Vulnerability to harm	.642**
Dependence	Dependency/incompetence	.731**
Low self-efficacy/weakness	Insufficient self-control	.586**
Failure	Failure	.754**
Enmeshment/diffuse boundaries	Enmeshment	.566**
Subjugation/submission to others	Subjugation	.633**
Self-sacrifice	Self-sacrifice	.598**
Approval seeking/excessive need to be liked	Approval-seeking	.751**
Emotional inhibition	Emotional constriction	.367**
Pessimism/negativity	Negativity	.523**
Unrelenting standards	Unrelenting standards	.773**
Punitiveness/unforgiving of self	Punitiveness (self)	.606**
Punitiveness/unforgiving of others	Punitiveness (other)	.731**
Entitlement/specialness	Entitlement	.770**

** $p < .01$ (one-tailed).

Table 4. Intercorrelations Between Final MSS Schemas.

Schema	DEF	DEP	LSE	FTL	EMI	PES	PUNS	FC	ORE	MW
Defectiveness										
Dependence	.50*									
Low self-efficacy	.52*	.74*								
Fatalistic	.45*	.43*	.45*							
Emotional Inhibition	.57*	.35*	.41*	.37*						
Pessimism	.64*	.54*	.62*	.50*	.47*					
Punitiveness-self	.53*	.34*	.35*	.33*	.47*	.47*				
Full Control	-.22*	-.28*	-.41*	-.28*	-.08	-.23*	.03			
Over Reliance on Emotions	-.14*	-.03	-.05	-.01	-.31*	-.16*	-.17*	.08		
Meaningless World	.62*	.47*	.55*	.52*	.54*	.63*	.38*	-.24*	-.06	

Pearson's r ; * $p < .01$. DEF = defectiveness, DEP = dependence, LSE = low self-efficacy, FTL = fatalistic, EMI = emotional inhibition, PES = pessimism, PUNS = punitiveness self, FC = Full Control, ORE = Over Reliance on Emotions, MW = Meaningless World.

Table 5. Confirmatory Factor Modeling for the MSS Schema and Groupings.

Model	$\chi^2(df)$	CFI	TLI	RMSEA [CI]	SRMR
Items to schemas (e.g., q1–q4 to abandonment)	10,344(5,319)	.98	.98	.03 [.024-.036]	.04
Schemas to needs (e.g., abandonment to safety and attachment)	24,530(5,655)	.95	.95	.06 [.053-.068]	.07
Schemas to focus (e.g., defectiveness to inadequate Self)	10,344(5,319)	.97	.97	.04 [.031-.047]	.05

χ^2 = chi-square, df = degrees of freedom, CI = confidence interval.

needs.” In all cases, incremental fit indices (CFI and TLI) were in the ideal range of $\geq .95$.

DIF testing (using a subset of Sample 3 data, $n = 500$) by age, time taken, and clinical versus nonclinical groups revealed no instances of item bias for any of the 108 MSS items. Clinical and nonclinical sample distributions are presented in Table 6. All schemas in the clinical group had higher mean scores than in the nonclinical groups, with the exceptions of Entitlement and Over Reliance on Emotions. An independent samples t -test (using Sample 3, $n = 1018$) of the total MSS score between clinical ($M = 201.08$, $SD = 37.18$) and nonclinical groups ($M = 181.28$, $SD = 40.02$), $t(1015) = 8.07$, was significant ($p < .001$), with a medium effect size ($d = .51$). Further independent samples t -tests between clinical and nonclinical samples at the schema level, adjusted for multiple comparisons using the Benjamini–Hochberg correction, showed mean differences were significant in 25 of the 27 schemas, with the two non-significant schemas being punitiveness-others, and full control. The Benjamini–Hochberg procedure was selected for multiple comparison adjustment as it offers superior statistical power while maintaining adequate control of the false discovery rate (Benjamini & Hochberg, 1995). This approach is appropriate for our analysis as the Bonferroni correction assumes independence between tests, and this assumption does not hold when examining intercorrelated psychological constructs like schemas (Cramer et al., 2016). Furthermore, recent methodological and simulation studies have demonstrated that the Bonferroni correction becomes increasingly conservative as the number of tests grows, potentially obscuring clinically meaningful differences (Gelman et al., 2012; Nakagawa, 2004; Stevens et al., 2017).

Discussion

This study detailed the development and validation of the MSS using Rasch methodology. The MSS was developed in a series of analyses with the final 27 subscales, 108-item scale possessing overall fit to the Rasch model for all subscales, no item misfit, adequate reliability, no local dependency, strict unidimensionality and measurement invariance by age, sex, time taken, and clinical group.

After initial analysis of a pilot version of the MSS, item reduction and second MSS analysis, all 27 MSS subscales showed overall fit to the Rasch model. Convergent validity was supported between comparable MSS and YSQ-R schemas, with each MSS schema positively correlating to its YSQ-R equivalent. For example, MSS abandonment/anxious attachment and YSQ-R abandonment correlated at $.889$ ($p \leq .01$), with the MSS schema having only four items compared with eight in the YSQ-R. The lowest observed convergent correlation was between emotional inhibition to its counterpart emotional constriction ($.367$). This moderate correlation may reflect differences in content

coverage as the MSS emotional inhibition items focus on fundamental beliefs about whether emotions themselves are helpful or harmful (e.g., “My emotions do more harm than good,” “Emotions are not useful, so I need to ignore them”), capturing an evaluation of emotional utility. In contrast, the YSQ-R emotional constriction items are all centered around relational dynamics. As schemas are conceptualized as “broad, pervasive themes” (Young et al., 2003) that organize understanding across multiple domains, not solely interpersonal relationships, the MSS approach attempted to capture the cognitive structure (the belief that emotions are fundamentally harmful) rather than its relational manifestations, as individuals may suppress emotional expression due to various factors (social anxiety, cultural norms, autism spectrum traits) without necessarily holding the underlying maladaptive belief about emotional danger. Furthermore, measuring the core belief directly identifies the primary therapeutic target in schema therapy, the fundamental evaluation that emotions are harmful, rather than just the downstream social expressions that may vary across individuals. While interpersonal expression difficulties are indeed related to and often result from the emotional inhibition schema, the MSS aims to isolate and measure the core cognitive pattern that emotions are dangerous or harmful and remove related but distinct elements.

While many of the subscales included in the MSS are aligned with previously established schemas, the MSS also included additional schemas. Patterns of convergent and divergent validity were observed between MSS subscales, for instance, the highest positive correlation was between low-self efficacy and dependence ($.74$), which share a common core of inability, where individuals who feel weak and inept are also likely to believe they cannot manage on their own. In addition, low self-efficacy showed positive correlations to subjugation ($.62$) and pessimism ($.62$). These correlations are consistent with the divergent correlation observed between full control and low self-efficacy ($-.41$). A sense of helplessness would expectedly be inversely related to a sense of total control of one’s environment and suggests that as the belief in control increases, feelings of helplessness and ineptitude may decrease. Several subscales such as failure, defectiveness, dependence, subjugation, pessimism, and lack of coherent identity also showed negative correlations to full control. Like low self-efficacy, these can be considered “negative” schemas, sharing themes of unrealistic underestimation related to inadequacy or limited agency. This naturally opposes the unrealistic overestimation characterizing the full control schema, which is an excessively positive self-view. This is consistent with research findings that both negative and positive schemas are potentially problematic when inflexibly held or extreme (e.g., see Steffen et al., 2017). Full Control provides a sense of agency, predictability, and hope, which relieves the vulnerability and painful emotions associated with these other,

Table 6. Means, Standard Deviations, and Distributions Statistics for Nonclinical and Clinical MSS Samples.

Schema	Nonclinical					Clinical					
	M	SD	% ≥ 3	SE	Kurtosis	M	SD	% ≥ 3	SE	Skew	Kurtosis
Abandonment/anxious attachment	8.16 ^a	3.90	14.3	.12	.03	9.23	2.68	21.2	.12	-.40	-.05
Excessive self-reliance/avoidant attachment	8.77 ^a	3.56	24.6	.15	-.14	10.20	3.29	39.2	.15	-.40	-.46
Emotional deprivation	6.91 ^a	3.32	2.1	.09	.33	7.23	2.15	3.4	.10	.41	-.25
Mistrust of others	7.77 ^a	3.32	4.9	.10	.13	8.59	2.32	11.3	.11	-.12	-.50
Others are dangerous/malevolent	6.09 ^a	2.92	4.2	.12	.30	6.63	3.31	8.6	.15	.41	-.09
Social isolation/outsider	7.96 ^a	3.30	8.2	.10	.26	8.53	2.29	9.9	.10	.22	-.10
Defectiveness/shame	5.98 ^a	4.09	13.4	.17	.49	7.52	3.93	19.1	.18	.07	-.81
Vulnerability to a dangerous world	6.42 ^a	3.44	1.6	.10	.28	7.10	2.27	2.9	.10	.15	-.44
Dependence	5.47 ^a	3.31	.9	.09	1.12	5.83	2.10	.9	.10	.57	-.28
Failure/achievement inferiority	8.40 ^a	4.01	15.7	.12	.22	9.49	2.78	26.8	.13	-.26	-.77
Low self-efficacy/weakness	5.86 ^a	2.98	1.7	.09	1.11	6.36	2.08	1.4	.10	.46	.12
Fatalistic/external locus of control	6.21 ^a	2.83	.9	.08	.48	6.82	2.11	2.7	.10	.65	1.06
Enmeshment/diffuse boundaries	5.61 ^a	3.53	7.3	.15	.47	6.88	3.27	8.8	.15	.14	-.39
Subjugation/submission to others	4.31 ^a	2.74	1.2	.12	.60	5.58	2.96	4.3	.14	.39	-.02
Self-sacrifice	9.06 ^a	3.07	15.4	.10	-.17	9.45	2.39	17.8	.11	-.32	.35
Approval-seeking/excessive need to be liked	6.39 ^a	3.62	9.6	.15	.26	7.70	3.62	17.1	.17	-.05	-.72
Emotional inhibition	6.97 ^a	3.47	4.0	.10	.45	7.91	2.35	5.4	.11	.12	-.30
Pessimism/negativity	6.42 ^a	3.42	3.5	.10	.50	7.28	2.61	5.4	.12	.18	-.34
Unrelenting standards	7.50 ^a	3.38	1.7	.07	.36	7.55	1.68	2.3	.08	.07	.78
Punitiveness/unforgiveness of self	7.17 ^a	3.17	5.0	.10	.44	7.68	2.31	5.4	.11	-.13	-.32
Punitiveness/unforgiveness of others	8.93	2.76	14.3	.10	-.01	9.18	2.25	15.8	.10	-.04	.11
Entitlement/specialness	4.32 ^a	2.77	1.4	.12	.56	3.89	2.55	.2	.12	.57	.15
Full Control	7.77	2.47	4.4	.09	.12	7.79	2.16	4.1	.10	.15	.35
Over Reliance on Emotions	7.01 ^a	2.61	4.9	.11	.08	6.59	2.56	3.8	.12	.15	-.25
Unfairness	5.80 ^a	3.23	5.4	.14	.42	6.45	3.17	7.7	.15	.43	-.25
Meaningless World	3.77 ^a	3.37	4.5	.14	1.06	5.18	3.81	7.0	.18	.57	-.23
Lack of Coherent Identity	6.10 ^a	4.13	14.1	.17	.42	8.43	4.10	26.1	.19	-.12	-.80

% ≥ 3 refers to the proportion of the sample who had an average score of 3 (Likert-type = agree) or above.

^aSignificance under correction for multiple comparisons.

more vulnerable schemas that may be unconsciously held or denied. An interesting implication is that high scores on Full Control may reflect biased responding and underreporting of other schemas on the MSS.

Of note, the nonclinical group scored higher than the clinical group on two of the 27 subscales: Entitlement (.43 difference) and Over Reliance on Emotions (.42 difference). These results can be explained by the composition of the samples, where the nonclinical sample was made up of practicing psychologists who, because of their relatively privileged positions, may feel more entitled than the typical mental health patient. Likewise, mental health patients often have difficulties with identifying and accepting emotions (Gratz & Roemer, 2004), which may explain why Over Reliance on Emotions was lower among the clinical group. Given the MSS's broad nature, minimal average differences are expected between groups. The major difference expected would be the proportion with extreme scores, defined as an average score of 3 (*agree*) or more on the Likert-type scale. Indeed, the clinical group had a higher frequency of high scores across 22 subscales, with one tied and four lower (Table 6). The nonclinical group had a higher proportion of an average score of 3 for Low Self-Efficacy (.03% difference), Full Control (.03% difference), Entitlement (1.2% difference), and Over Reliance on Emotions (1.1% difference).

The MSS represents a significant advancement in schema assessment through its integration of contemporary theoretical constructs and modern psychometric construction. The MSS includes nine new subscales that integrate developments in attachment theory, trauma research, and contemporary psychopathology models (Supplementary Material S5 contains the full MSS scale). Specifically, the MSS introduces to clinicians a tool to measure schema related to avoidant attachment, dissociative and identity-related phenomena, perceived control, and self-efficacy not otherwise represented in traditional schema measures. This scale also integrates Dweck's (2017) theory of motivation, personality, and development, which underpins the three schemas (Meaningless World, Unfairness, and Lack of Coherent Identity) proposed by Arntz et al. (2021). This is the first study to empirically investigate and successfully validate all three proposed schemas. Each showed good model fit, reliability, unidimensionality, and expected correlations with other subscales. In particular, Lack of Coherent Identity was further validated by its significant association with greater response variability across schemas. As hypothesized, higher scores on Lack of Coherent Identity had a higher degree of response variance within each subscale, validating that high scorers have a less consistent sense of self compared with low scorers. By validating and integrating these schemas and further including Full Control, the MSS can more comprehensively represent schemas arising from frustrated core emotional needs consistent with Dweck's (2017) theory. Method improvements

are also apparent with reversed items to help reduce response biases associated with one-directional items, and the brief 4-item structure per subscale enables between-schema comparisons while keeping the length practical.

The MSS is currently the only schema measure to show fit to the Rasch model expectations of a useful measurement instrument, and suggests the MSS accurately represents its latent schema constructs across their continuums of severity. Regarding measurement invariance, the MSS demonstrated an absence of DIF (measurement invariance is the absence of DIF) across all items at the clinical group level, by age group, and by sex. In contrast, the vast majority of schema measures have not been tested for invariance at all. The findings of the current study support observed score differences in the MSS as representative of true differences in underlying constructs rather than the result of measurement artifacts, in the case of comparisons between males and females, young, middle, and older age groups, those who complete the scale slowly or quickly, and those in and not in clinical treatment.

The current findings support the validity of the MSS schemas, and the use of a clinical sample in this study strengthens the validity and applicability of the MSS. In clinical practice, the MSS can provide a more comprehensive assessment of schemas linked to contemporary understandings of trauma, attachment, and psychopathology while maintaining measurement precision, representing a significant advancement over the traditional maladaptive schema measures in accuracy and interpretability.

Limitations

Two MSS subscales, Enmeshment and Full Control, fell short of a .70 threshold for PSI reliability. The Enmeshment scale, despite being only 4 items, was comparable to the 7-item YSQ-R Enmeshment scale. Both MSS and YSQ-R enmeshment scales having low reliability, despite being conceptualized and measured slightly differently, call into question the construct of enmeshment itself. We chose to retain enmeshment as it represents a foundational construct in schema therapy for identifying and understanding interpersonal boundary issues, identity development, patterns in personality disorders, and attachment-related psychopathology (Jacobvitz et al., 2004). Given its clinical importance, further refinement of the conceptual framework of enmeshment is warranted. The subscale Full Control also fell below the .70 PSI cutoff. Despite Full Control exhibiting poor reliability, we chose to retain it because of its clinical utility and its negative correlation with other subscales, providing a potential avenue to identify individuals with low reflective capacity.

There are several important avenues for future work on the MSS, primarily to identify cut-offs for each schema, report illustrative clinical case studies, and assess longitudinal psychometric properties such as

temporal stability. Overall, the MSS uses contemporary theoretical frameworks and modern psychometric analysis to produce a schema measure that is reliable, more comprehensive, and yet shorter than traditional measures. Proposed schemas were empirically supported, and through a series of iterations, a new, open-source scale useful for clinical and research applications was produced.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: B.B., E.B., C.S., D.H. work in a professional capacity at NovoPsych Psychometrics.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Ethical Considerations

This research was approved by a Human Research Ethics Committee at Monash University, Australia.

Consent to Participate

Data were collected from practicing psychologists who consented for de-identified data to be used for psychometric research. The Human Research Ethics Committee granted a waiver of consent for clients because de-identified data were collected during routine care, with no imposition from research created.

Consent for Publication

Not applicable.

ORCID iDs

Ben Buchanan  <https://orcid.org/0009-0001-8700-034X>
Emerson Bartholomew  <https://orcid.org/0000-0003-0735-0436>
David Hegarty  <https://orcid.org/0000-0002-4079-1017>

Supplemental Material

Supplemental material for this article is available online.

References

- Andrich, D. (2016). Rasch rating-scale model. In W. J. van der Linden (Ed.), *Handbook of Item Response Theory. Vol. 1: Models* (pp. 76–94). CRC Press.
- Andrich, D., Sheridan, B., & Luo, G. (2009). *RUMM 2030*. RUMM Laboratory.
- Arntz, A., Rijkeboer, M., Chan, E., Fassbinder, E., Karaosmanoglu, A., Lee, C. W., & Panzeri, M. (2021). Towards a reformulated theory underlying schema therapy: Position paper of an international workgroup. *Cognitive Therapy and Research, 45*, 1007–1020. <https://doi.org/10.1007/s10608-021-10209-5>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*(1), 289–300.
- Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q3: Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement, 41*(3), 178–194. <https://doi.org/10.1177/0146621616677520>
- Chui, W.-Y., & Leung, M.-T. (2016). Adult attachment internal working model of self and other in Chinese culture: Measured by the Attachment Style Questionnaire—Short Form (ASQ-SF) by confirmatory factor analysis (CFA) and item response theory (IRT). *Personality and Individual Differences, 96*, 55–64. <https://doi.org/10.1016/j.paid.2016.02.068>
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. <https://doi.org/10.1037/1040-3590.7.3.309>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cramer, A. O., Van Borkulo, C. D., Giltay, E. J., Van Der Maas, H. L., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLOS ONE, 11*(12), Article e0167490.
- Dunn, A. M., Heggestad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences. *Journal of Business and Psychology, 33*(1), 105–121. <https://doi.org/10.1007/s10869-016-9479-0>
- Dweck, C. S. (2017). From needs to goals and representations: Foundations for a unified theory of motivation, personality and development. *Psychological Review, 124*, 689–719. <https://doi.org/10.1037/rev0000082>
- Finger, R. P., Fenwick, E., Pesudovs, K., Marella, M., Lamoureux, E. L., & Holz, F. G. (2012). Rasch analysis reveals problems with multiplicative scoring in the macular disease quality of life questionnaire. *Ophthalmology, 119*(11), 2351–2357. <https://doi.org/10.1016/j.ophtha.2012.05.031>
- Fisher, W. (1992). Reliability statistics. *Rasch Measurement Transactions, 6*(3), 238.
- Fraley, R. C. (2019). Attachment in adulthood: Recent developments, emerging debates, and future directions. *Annual Review of Psychology, 70*(1), 401–422. <https://doi.org/10.1146/annurev-psych-010418-102813>
- Fraley, R. C., Hudson, N. W., Heffernan, M. E., & Segal, N. (2015). Are adult attachment styles categorical or dimensional? A taxometric analysis of general and relationship-specific attachment orientations. *Journal of Personality and Social Psychology, 109*(2), 354–368. <https://doi.org/10.1037/pspp0000027>
- Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness, 5*(2), 189–211.
- Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment, 26*(1), 41–54.

- Green, T. C., & Balfour, A. (2020). Assessment and formulation in schema therapy. In G. Heath, & H. Startup (Eds.), *Creative methods in schema therapy: Advances and innovation in clinical practice* (pp. 19–47). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9781351171847-1>
- Gustafsson, J.-E. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 33(2), 205–233. <https://doi.org/10.1111/j.2044-8317.1980.tb00609.x>
- Hagell, P., & Westergren, A. (2016). Sample Size and Statistical Conclusions from Tests of Fit to the Rasch Model According to the Rasch Unidimensional Measurement Model (Rumm) Program in Health Outcome Measurement. *Journal of Applied Measurement*, 17(4), 416–431.
- Jacobvitz, D., Hazen, N., Curran, M., & Hitchens, K. (2004). Observations of early triadic family interactions: Boundary disturbances in the family predict symptoms of depression, anxiety, and attention-deficit/hyperactivity disorder in middle childhood. *Development and Psychopathology*, 16(3), 577–592. <https://doi.org/10.1017/s0954579404004675>
- Janoff-Bulman, R. (1989). Assumptive worlds and the stress of traumatic events: Applications of the schema construct. *Social Cognition*, 7, 113–136. <https://doi.org/10.1521/soco.1989.7.2.113>
- Karantzas, G. C., Feeney, J. A., & Wilkinson, R. (2010). Is less more? Confirmatory factor analysis of the Attachment Style Questionnaires. *Journal of Social and Personal Relationships*, 27(6), 749–780. <https://doi.org/10.1177/0265407510373756>
- Karantzas, G. C., Younan, R., & Pilkington, P. D. (2023). The associations between early maladaptive schemas and adult attachment styles: A meta-analysis. *Clinical Psychology*, 30(1), 1–20. <https://doi.org/10.1037/cps0000108>
- Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, 7(4), 328.
- Lundgren-Nilsson, Å., Jonsdottir, I. H., Ahlborg, G., Jr., & Tennant, A. (2013). Construct validity of the Psychological General Well Being Index (PGWBI) in a sample of patients undergoing treatment for stress-related exhaustion: A Rasch analysis. *Health and Quality of Life Outcomes*, 11, 2. <https://doi.org/10.1186/1477-7525-11-2>
- Lundgren-Nilsson, Å., & Tennant, A. (2011). Past and present issues in Rasch analysis: The functional independence measure (FIM™) revisited. *Journal of Rehabilitation Medicine*, 43(10), 884–891. <https://doi.org/10.2340/16501977-0871>
- Manicavasagar, V., Silove, D., Curtis, J., & Wagner, R. (2000). Continuities of separation anxiety from early life into adulthood. *Journal of Anxiety Disorders*, 14(1), 1–18. [https://doi.org/10.1016/s0887-6185\(99\)00029-8](https://doi.org/10.1016/s0887-6185(99)00029-8)
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174. <https://doi.org/10.1007/BF02296272>
- Medvedev, O. N., & Krägeloh, C. U. (2022). Rasch measurement model. In O. N. Medvedev, C. U. Krägeloh, R. J. Siegert, & N. N. Singh (Eds.), *Handbook of assessment in mindfulness research* (pp. 1–18). Springer.
- Moorey, S., Byrne, S., & Ruths, F. (2020). Brief work: Schema informed CBT. In G. Heath, & H. Startup (Eds.), *Creative methods in schema therapy: Advances and innovation in clinical practice* (pp. 301–315). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9781351171847-19>
- Murugappan, M. N., King-Kallimanis, B. L., Mangir, C., Howie, L., Bhatnagar, V., Beaver, J. A., Basch, E. M., Henson, S. R., & Kluetz, P. G. (2022). Floor and ceiling effects in the EORTC QLQ-C30 Physical Functioning Subscale among patients with advanced or metastatic breast cancer. *Cancer*, 128(4), 808–818. <https://doi.org/10.1002/cncr.33959>
- Mussel, P. (2023, May 15). *Core beliefs about the self: A structural model based on a systematic review and natural language processing*. <https://doi.org/10.31234/osf.io/ky4vu>
- Nakagawa, S. (2004). A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behavioral Ecology*, 15(6), 1044–1045.
- NovoPsych. (2025, August 7). *Maladaptive Schema Scale*. <https://novopsych.com/assessments/formulation/maladaptive-schema-scale-mss/>
- Oei, T. P. S., & Baranoff, J. (2007). Young schema questionnaire: Review of psychometric and measurement issues. *Australian Journal of Psychology*, 59(2), 78–86. <https://doi.org/10.1080/00049530601148397>
- Padesky, C. A. (1994). Schema change processes in cognitive therapy. *Clinical Psychology and Psychotherapy*, 1, 267–278. <https://doi.org/10.1002/cpp.5640010502>
- Pilkington, P. D., Younan, R., & Karantzas, G. C. (2023). Identifying the research priorities for schema therapy: A Delphi consensus study. *Clinical Psychology & Psychotherapy*, 30(2), 344–356. <https://doi.org/10.1002/cpp.2800>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment test*. Danish Institute for Educational Research.
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Schwarzer, R., & Jerusalem, M. (1995). Self-efficacy measurement: Generalized self-efficacy scale (GSES). In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio* (pp. 35–37). NFER-Nelson.
- Sempértegui, G. A., Karreman, A., Arntz, A., & Bekker, M. H. (2013). Schema therapy for borderline personality disorder: A comprehensive review of its empirical foundations, effectiveness and implementation possibilities. *Clinical Psychology Review*, 33(3), 426–447. <https://doi.org/10.1016/j.cpr.2012.11.006>
- Smith, E. V. (2002). Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3, 205–231.
- Steffen, P. R., Elliott, C. H., Lassen, M. K., Olsen, J., & Smith, L. L. (2017). Expanding schema conceptualisation and assessment: Towards a richer understanding of adaptive and maladaptive functioning. *Australian Journal of Psychology*, 69(3), 200–209. <https://doi.org/10.1111/ajpy.12141>
- Stevens, J. R., Al Masud, A., & Suyundikov, A. (2017). A comparison of multiple testing adjustment methods with block-correlation positively-dependent tests. *PLOS ONE*, 12(4), Article e0176124.
- Substance Abuse and Mental Health Services Administration. (2014). *Trauma-informed care in behavioral health services*. U.S. Substance Abuse and Mental Health Services Administration.

- Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism*, 57(8), 1358–1362.
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529–554. <https://doi.org/10.1086/214483>
- Yalcin, O., Lee, C., & Correia, H. (2020). Factor structure of the Young Schema Questionnaire (Long Form-3). *Australian Psychologist*, 55(5), 546–558. <https://doi.org/10.1111/ap.12458>
- Yalcin, O., Marais, I., Lee, C. W., & Correia, H. (2022). Revisions to the Young Schema Questionnaire using Rasch analysis: The YSQ-R. *Australian Psychologist*, 57(1), 8–20. <https://doi.org/10.1080/00050067.2021.1979885>
- Yalcin, O., Marais, I., Lee, C. W., & Correia, H. (2023). The YSQ-R: Predictive validity and comparison to the short and long form Young Schema Questionnaire. *International Journal of Environmental Research and Public Health*, 20(3), 1778. <https://doi.org/10.3390/ijerph20031778>
- Young, J. E. (Ed.). (1990). *Cognitive therapy for personality disorders: A schema-focused approach*. Professional Resource Press.
- Young, J. E., Klosko, J. S., & Weishaar, M. E. (2003). *Schema therapy: A practitioner's guide*. Guilford Press.