Published in final edited form as:

J Pers Assess. 2014 ; 96(3): 283–292. doi:10.1080/00223891.2013.830263.

Toward a Model-Based Approach to the Clinical Assessment of Personality Psychopathology

Nicholas R. Eaton¹, Robert F. Krueger², Anna R. Docherty³, and Scott R. Sponheim⁴

¹Stony Brook University

²University of Minnesota

³University of Missouri

⁴Minneapolis VA Health Care Center

Abstract

Recent years have witnessed tremendous growth in the scope and sophistication of statistical methods available to explore the latent structure of psychopathology, involving continuous, discrete, and hybrid latent variables. The availability of such methods has fostered optimism that they can facilitate movement from classification primarily crafted through expert consensus to classification derived from empirically-based models of psychopathological variation. The explication of diagnostic constructs with empirically supported structures can then facilitate the development of assessment tools that appropriately characterize these constructs. Our goal in this paper is to illustrate how new statistical methods can inform conceptualization of personality psychopathology and therefore its assessment. We use *magical thinking* as example, because both theory and earlier empirical work suggested the possibility of discrete aspects to the latent structure of personality psychopathology, particularly forms of psychopathology involving distortions of reality testing, yet other data suggest that personality psychopathology is generally continuous in nature. We directly compared the fit of a variety of latent variable models to magical thinking data from a sample enriched with clinically significant variation in psychotic symptomatology for explanatory purposes. Findings generally suggested a continuous latent variable model best represented magical thinking, but results varied somewhat depending on different indices of model fit. We discuss the implications of the findings for classification and applied personality assessment. We also highlight some limitations of this type of approach that are illustrated by these data, including the importance of substantive interpretation, in addition to use of model fit indices, when evaluating competing structural models.

Keywords

magical thinking; schizotypy; factor mixture models; latent structure; classification

Corresponding author contact information: Nicholas R. Eaton, Ph.D., Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500, nicholas.eaton@stonybrook.edu.

Declarations of interest: None.

The goal of psychological assessment is to characterize an individual's standing on individual difference constructs of clinical relevance. The way that these constructs are defined has significant implications for their measurement. In assessment of normal-range personality, the constructs of interest are often conceptualized as traits—that is, dimensions of individual differences, for example, the domains of the Five-Factor Model, as instantiated in the NEO-PI-R (Costa & McCrae, 1992). Assessment of personality disorders, on the other hand, is typically reflective of official nosological systems such as the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)*; American Psychiatric Association, 1994), which defines personality psychopathology by means of a set of 10 putatively distinct, independent, and categorical entities.¹ While historically types (categories) and traits (dimensions) have served as the competing structures of personality-related variables, other possibilities exist. Indeed, advances in mixture modeling now allow personality and psychopathology researchers to test hybrid structures—structures that combine traits and types into a single model. The current report highlights the potential utility and limitations of this sort of hybrid modeling of personality data as well as its assessment implications.

The Relationship between Structure and Assessment

Structural assumptions about personality variables are inextricably linked to personality assessment. Reliable assessment of normal-range personality traits, and personality disorder categories, for instance, frequently takes different forms, given that the constructs of interest are presumed to have different structures (i.e., dimensions vs. categories, respectively). When assessing personality traits, the assessor needs to measure the full range of the trait dimension to determine where an individual falls on it. When assessing for the presence or absence of a *DSM-IV* personality disorder, the assessor needs to evaluate the presence or absence of the binary categorical diagnosis. Given the polythetic nature of the criterion sets, the purpose of the assessment is to determine which criteria are present, calculate the number of present criteria, and note whether this sum meets or exceeds a diagnostic threshold. If so, a personality disorder diagnosis can be rendered. As discussed below, categorical-dimensional hybrid structures similarly present unique assessment needs.

The nature of the personality assessment instrument—its content, its items' structure, its scoring and decision rules—reflect assumptions about the distributional characteristics of the constructs of interest. For example, a measure of a dimensional personality trait must employ a collection of items that measure the trait at various points along its continuum. In item response theory (IRT) models, such an inventory should have items of different *location* (or “difficulty”), thus covering the range of the trait dimension being assessed. Alternatively, measures of personality disorder under the *DSM-IV* framework are less concerned with sampling a broad universe of content, because the focus is on determining whether a specific threshold is crossed, counting what are presumed to be fungible criteria.

¹It is noteworthy that not all psychopathology measurement attempts are based solely on assessment devices detailing *DSM*-based constructs. For instance, the ongoing Research Domain Criteria (RDoC) effort from the National Institute of Mental Health attempts to understand psychopathology based on dimensions of neurobehavioral functioning. However, overall, personality disorder assessment reflecting a *DSM*-based tradition is pervasive.

Items on *DSM*-oriented inventories are usually intended to gather converging pieces of information about each criterion to determine whether or not it is present.

Distributional Assumptions of Personality Constructs

Historically, many assumptions about the distributions of data reflecting personality constructs resulted from expert opinion or theory. Both “type” theories (e.g., introverts versus extraverts) and dimensional theories (e.g., an introversion-extraversion dimension) of personality have been proposed, and assessment instruments have reflected this bifurcation in conceptualization. The resulting implications for assessment are far from trivial. The structure of a personality test designed to determine whether an individual is of one of two personality types, say introverted *or* extraverted, need only assess introverted characteristics and extraverted characteristics, as opposed to assessing characteristics that are more indicative of a mid-range level of introversion-extraversion (e.g., falling at the 50th percentile of an introversion-extraversion dimension). There is no middle-ground between introversion and extraversion in a type theory, so items covering a middle-ground are not relevant. The structure of a personality test designed to determine at what level an individual falls on an introversion-extraversion dimension would be quite different. Items tapping into the difficulty level found at the midpoint of the dimension would be critical, as would be items assessing very high and very low trait levels (i.e., the dimensional poles).

Categorical versus dimensional comparisons have often been conducted in previous personality and psychopathology studies. In contrast, there have been relatively few direct comparisons of these models with categorical-dimensional hybrid possibilities—models that share some characteristics of both typal and trait models. Testing these alternative hybrid structures and comparing them to categorical and dimensional structures are important steps toward precisely characterizing the latent constructs of interest.

Because the structure of personality assessment is reflective of the underlying distributional assumptions of the personality constructs of interest, reliance solely on expert opinion about these distributions is potentially problematic. The resulting problems of a mismatch between a model and the data can be seen clearly in the realm of personality disorders. Although the *DSM-IV* defines personality psychopathology via 10 categorical diagnostic entities (constituting a model), observations regarding the distribution of personality disorder features in individuals (the data) fail to fit into this classification system well. The myriad problems with the current personality disorder approach have been documented extensively elsewhere (e.g., Clark, 2007; Trull & Durrett, 2005; Widiger & Samuel, 2005). For example, the *DSM-IV* personality disorders are highly comorbid, elements of these disorders tend to vary continuously (a phenomenon poorly captured by dichotomous diagnoses), and multiple different constellations of diagnostic criteria can result in the same categorical diagnosis. Fortunately, it is no longer necessary to rely exclusively on expert opinion. Data-driven methodologies have emerged as an important tool in adjudicating these issues.

Model-Based Tests of Distributional Assumptions

It is critical for personality theory and assessment that underlying distributional assumptions of symptomatology be correct and justifiable, because different distributions impact the way

clinical and research constructs are conceptualized, measured, and applied to individuals. Indeed, characterizing these latent constructs properly is a prerequisite for efforts to assess them; it is of limited value to assess an improperly conceived construct with high reliability. How can we achieve this goal? Solely relying on expert opinion is clearly not enough, given that this approach has produced countless competing models. Instead, an empirically supported structure is needed, where the model is fit to the data, versus the historical precedent where data is fit to the model. Along these lines, statistical methods exist to compare competing distributional models directly so as to detail most accurately the nature of these latent constructs. We will discuss three such models below.

Latent trait models

Latent trait models posit the presence of one or more underlying continuous distributions. Individuals are dispersed along this trait dimension in a particular way (e.g., following a normal distribution) such that there are no *zones of rarity* (Kendell & Jablensky, 2003)—that is, there are no locations along the dimension that are unoccupied by some individuals. The latent trait is thus continuous and not segmented. (It is noteworthy, however, that some compelling research suggests that psychopathological constructs assumed to be dimensional might be better conceived as *quasi-continuous*. By quasi-continuous, we mean that the construct would be bounded at the low end by zero, a complete absence of the quality corresponding with the construct; see Reise & Waller, 2009.) Latent trait models form the basis of traditional exploratory factor analysis (EFA) and confirmatory factor analysis as well as IRT (e.g., Brown, 2006; Embretson & Reise, 2000). All items are associated with a particular location along the span of the latent trait. Further, in many such models, item *discrimination* can be modeled as well. Discrimination is a measure of how strongly the item taps into the latent trait. In terms of assessment, measures that reflect a latent trait model yield a score for each individual indicating his or her level of the latent trait. These dimensional scores can easily be transformed into percentiles for interpretation or compared to norms from a population of interest to facilitate extra-test inference.

Latent class models

Latent class models (McCutcheon, 1987) are based on the supposition of a latent group (class) structure for a personality construct's distribution, and they are typically evaluated via latent class analysis (LCA). A finite number of latent classes are posited. They are mutually exclusive and nominal, separated by zones of rarity of observations. Every individual is assigned to his or her most probable class of membership given his or her responses to items on the assessment measure. These classes are defined by patterns of item endorsement across individuals, assuming that inter-item correlations solely reflect class membership. This assumption is referred to as *conditional independence*.

Conditional independence has important implications for how the construct itself is conceptualized and assessed. Because inter-item correlations are accounted for solely by class membership, there are assumed to be no inter-item correlations within each class. In this way, assessment measures that reflect a latent class model yield a nominal (categorical, but not necessarily a dichotomy) result for each individual, such as Class A, Class B, or Class C. Given conditional independence, there is assumed to be no variation within classes,

so all individuals assigned to a particular class are expected to be of the same type and not to differ among one another on the construct of interest.

Hybrid models

Hybrid models are relative newcomers to the collection of possible models to test distributional assumptions (e.g., Clark et al., under review; Muthén, 2006; Muthén & Asparouhov, 2006). These models, often grouped together as *factor mixture models*, combine the continuous aspects of latent trait models with the discrete aspects of latent class models. Multiple forms of factor mixture models exist, such as latent class factor analysis, which can model factor distributions empirically by means of a series of mutually exclusive classes graded by trait level, thus creating a hybrid structure of a dimension composed of a set of classes. Some mixture models identify classes of individuals within a particular factor structure. Others identify classes of individuals, within which a dimensional factor structure differentiates individuals; these factor structures may be the same (invariant) or unique across classes. For present purposes, we will focus on exploratory factor mixture analysis (EFMA), which is a true hybrid of classes and dimensions. A set of latent classes are fit to the data and, within each class, a latent trait solution captures additional variance (overcoming assumptions of conditional independence within classes). Stated another way, EFMA conducts LCA on the data and then conducts EFA within each class, allowing for different latent trait structures within each class.

An example: Dimension, class, and hybrid possibilities for narcissism

To illustrate the three sorts of structures detailed above, we refer the reader to a simplified example given in Figure 1. In the figure, possible dimensional, class-based, and hybrid structures of narcissism data are depicted. In the dimensional structure, we see that the narcissism trait level dimension (theta) ranges from low to high. (We have illustrated only a unidimensional structure for narcissism for the purpose of simplicity, but it is possible, of course, that multiple dimensions could be necessary to capture narcissism's structure adequately; see Miller, Hoffman, Campbell, & Pilkonis, 2008.) Individuals are distributed continuously along the dimension in this example, with most individuals falling at an average trait level of narcissism.

In the class structure in Figure 1, a hypothetical two-class structure is depicted, based on research identifying two types of narcissism: grandiose and vulnerable (Miller & Campbell, 2008; Pincus & Lukowitsky, 2010). Grandiose narcissism reflects grandiosity, entitlement, aggression, and dominance, whereas vulnerable narcissism reflects a defensive and fragile grandiosity that serves as a mask for feelings of inadequacy. In this example, narcissistic individuals are parsed into grandiose and vulnerable classes, with a higher frequency of individuals falling into the grandiose class (as denoted by the bar heights). What is noteworthy about this class structure is that, although individuals within each class might differ in terms of their levels of grandiose or vulnerable narcissism, they are conceptualized in this class model as having the same amount of pathology as other members of their class; this is a basic illustration of conditional independence. Thus, while theoretically individuals in the grandiose class might vary in terms of their levels of grandiose narcissism, this model

treats them as having an identical level. The same is true for those in the vulnerable narcissism class.

Finally, in the hybrid structure in Figure 1, we see how classes and dimensions can be combined in a complementary fashion. In this structure, individuals are again assigned into a hypothetical two-class solution of grandiose and vulnerable narcissism. Within each class, however, a dimension allows us to characterize narcissism severity differences between class members. Individuals are distributed continuously along this dimension within both classes. In terms of interpretation, this result suggests that there are two general narcissism groups of individuals: grandiose and vulnerable. However, not everyone within each of those two general groups has the same degree of narcissistic pathology. The dimension of narcissism severity within each class captures these within-class narcissism differences, with most individuals falling at the average (mid-point) level of the dimension. Individuals within the grandiose class, for instance, might be “somewhat grandiose,” “moderately grandiose,” “extremely grandiose,” and all other points along the severity dimension. Different hybrid models place different assumptions on the similarity (invariance) of these dimensions within classes. So, some sorts of hybrid models would parameterize the dimensions within each class to be statistically invariant and representative of the same latent construct. Other sorts of models would allow unique dimensions to capture differences within each class. For excellent graphical depictions of a wide variety of dimensional, categorical, and hybrid models, see Masyn, Henderson, and Greenbaum (2010).

Model comparison

The availability of various latent structure models has allowed researchers to apply them to determine the latent traits underlying covariation among test items, to explore latent classes for assigning individuals to particular type-based groups, and, more rarely (owing to the corresponding models being newer), to search for mutually exclusive groups of individuals who can be sorted within-group in a dimensional fashion. Historically, researchers chose which model to apply *a priori* based on theory. This is no longer necessary, however. Now, information-theoretic fit indices, such as the Bayesian information criterion (BIC; Raftery, 1995) balance fit and parsimony of models and allow for direct comparison of non-nested models. In other words, the resulting BIC values for latent trait, class, and hybrid models can be compared directly to see which model best balances fit and parsimony. The data can now adjudicate between models with different distributional properties, limiting the necessity of subjective opinion and increasing the congruence between the theoretical model and the observed data. Of course, the utility and interpretability of the resulting models are of paramount importance, and thus rational judgment remains an essential component of this process, particularly when no model emerges as clearly optimal over the others (e.g., in the case of different fit indices supporting different models).

The implications for the availability of direct model comparison go beyond the accumulation of knowledge about latent constructs. They also clearly impact assessment, given that the constructs and their assessment are closely linked. Indeed, the logical outcome of such studies is improved measurement, reflective of the structures identified as optimal. This need to clarify latent structures is present not only for applied clinical purposes but for

research endeavors as well. For instance, it is unsurprising that molecular genetic studies, such as linkage analysis and genome wide association studies (GWAS), typically fail to identify likely subtle genetic effects underlying very heterogeneous diagnoses with poor construct validity. If we begin with a construct like *DSM-IV* borderline personality disorder that has a poor measurement model—it allows any of 256 different criterion combinations to yield the same diagnosis, and two patients with the diagnosis can share only a single diagnostic criterion (Johansen et al., 2004; Krueger & Eaton, 2010)—no amount of careful efforts to assess this construct reliably will counteract its poor construct validity. For these reasons, it is critical to characterize latent structures first, then develop appropriate measurement devices, and finally apply these assessment tools for clinical and research purposes (see e.g., Reise & Waller, 2009).

An Illustration by Application to Data on Magical Thinking

We now turn our attention to an application of model-based comparisons of latent trait, class, and hybrid distributions for illustrative purposes, by focusing on data relating to *magical thinking*. Magical thinking can involve beliefs in extra-sensory perception (ESP), telepathy, mind-reading, or fortune telling. Magical thinking can also involve a tendency for subjectively experiencing supernatural encounters, seeing the future, observing unidentified flying objects (UFOs), or communicating with others mentally. Sometimes, individuals with magical thinking believe that others can read their minds as well.

Distortions in reality testing such as magical thinking may characterize a variety of forms of personality psychopathology. As one example, magical thinking has a long history of being conceptualized as reflective of a latent class of schizotypy (in that literature, often referred to as a *taxon*). A good deal of research on the schizotypy construct—largely applying taxometric analytic methods (see Waller & Meehl, 1997), rather than the model-based approaches such as LCA discussed here—has been supportive of such a discrete distribution (Haslam, 2011). As Haslam (2011) points out, however, the trend in taxometric studies is that more recent and methodologically sophisticated analyses increasingly support latent dimensional distributions; the one taxometric study of schizotypy that has unambiguously supported dimensionality (Rawlings, Williams, Haslam, & Claridge, 2008) was the most recent. It is noteworthy, though, that most studies of latent psychopathology structures of which we are aware have supported latent dimensional, rather than latent class, distributions, when those sorts of models were compared directly in community based samples (e.g., Eaton et al., in press; Markon & Krueger, 2005; Walton, Ormel, & Krueger, 2011; Wright et al., in press).

In summary, discrete, continuous, and hybrid accounts of magical thinking are all plausible based on past theory and research. We therefore illustrate the application of a model-based approach to conceptualization and assessment of personality psychopathology using the example of magical thinking. In contrast to most previous model-based research, we apply these methods to a clinically enriched sample in the current paper. Notably, these data also illustrate potential limitations of these analyses and allow us to consider circumstances of potential misinterpretation. It is not our goal in this report to produce the final word on the structure of magical thinking, given notable limitations of our data and the use of a relatively

circumscribed set of possible models to avoid complexity. Instead, we provide this empirical example as a means to highlight the use, promise, and potential complications of a model-based approach to uncovering personality structure and its implications for applied assessment.

Method

Participants

The 301 participants were composed of 115 psychiatric outpatients and 186 of their first-degree relatives. Data were gathered within the context of a larger family study in which patients with schizophrenia, schizoaffective disorder, or bipolar disorder were recruited from the Minneapolis VA Health Care System, community outpatient programs, and a county mental health clinic. The sample was 52% female, primarily Caucasian (6% African-American), and the mean age was 48.9 ($SD = 14.0$). Outpatients had a range of diagnoses including schizophrenia ($n = 63$), schizoaffective disorder ($n = 18$), bipolar disorder ($n = 31$) and psychosis NOS ($n = 3$). First-degree relatives also had a broad range of lifetime diagnoses including schizophrenia-spectrum disorders ($n = 8$), depressive disorders ($n = 47$), anxiety disorders ($n = 20$) and alcohol dependence ($n = 9$). First-degree biological relatives were identified through interviews with patients and invited by letter and telephone to participate. Patients and relatives were excluded for histories of substance dependence but were not excluded for past alcohol dependence, as long as they had not abused alcohol in the past month.

All participants completed the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I; First et al., 1996) and the Diagnostic Interview for Genetic Studies (DIGS, Nurnberger et al., 1994). Lifetime Axis I diagnoses for participants were determined by doctoral-level psychologists and trained advanced graduate students through a consensus process consistent with published guidelines (Leckman et al., 1982), which involved review of the structured interview data, medical history, and family informant material. Sponheim and colleagues (2004) provided a full description of exclusion criteria, recruitment, and assessment procedures. All participants completed an informed consent process, and the Minneapolis VA Health Care System and University of Minnesota Institutional Review Boards approved the study protocol. For relatives who were unable to travel for in-person assessments, questionnaires were mailed to them after they had completed informed consent procedures by telephone and a mailed consent form had returned with their signature. These individuals completed SCID-I assessment via telephone and returned completed questionnaires by United States Postal Service.

Assessment

All participants completed the SPQ (Raine, 1991), a 74-item true-false questionnaire that measures interpersonal, cognitive-perceptual, and disorganized schizotypal characteristics with nine subscales: constricted affect, suspiciousness, no close friends, social anxiety, odd beliefs, odd behavior, odd speech, magical thinking/unusual perceptual experiences, and ideas of reference. The seven magical thinking items assessed the following content:

Item 1: supernatural experiences

Item 2: belief in telepathy

Item 3: thoughts accessible by other people

Item 4: belief in clairvoyance

Item 5: feelings accessible by other people

Item 6: paranormal experiences

Item 7: telepathic communication experiences

The SPQ has shown good criterion and discriminant validity for the detection of schizotypal personality features. Fifty-five percent of individuals in a normative sample who scored in the top 10% of SPQ total scores had a *DSM-III-R* diagnosis of schizotypal personality disorder as assessed by the SCID-I. In addition, the SPQ shows low correlations with scales that assess aspects of psychosis-proneness not included in schizotypal personality disorder criteria (e.g., $r = .18$ to $.19$ with anhedonia; Raine, 1991). Cronbach's alpha for the SPQ magical thinking subscale items was .80 in the current study.

Statistical Analyses

All analyses were conducted in Mplus version 6 (Muthén & Muthén, 2011), applying the default delta parameterization, which allows scale factors of categorical variables to be model parameters, and a robust maximum likelihood estimator (MLR). MLR allowed us to account for familial clustering within the participants and also to treat each of the SPQ magical thinking items as categorical. We used at least 100 sets of random starting values and at least 20 sets of final stage iterations for analyses requiring starting values. This use of multiple starting value sets is important in avoiding mixture model analyses identifying “local maxima.” In simple terms, by requiring multiple analyses begin at different values, it is less likely that the analysis will converge on an apparently optimal solution prematurely, without investigating other solutions with largely different values. For further information about this critical issue in mixture modeling, see Hipp and Bauer (2006). EFAs used the default geomin rotation, which has performed well in recovering latent oblique dimensional structures (Browne, 2001).

Model evaluation focused on comparison of BIC statistics for each model, with smaller BIC values indicating models that better balanced fit to the data and parsimony. A BIC difference of 10 between two models indicates the posterior odds of the model favored by BIC over the inferior model are 150:1. BIC frequently outperforms other statistical tests and fit indices in recovering known latent structures in simulations (Nylund, Asparouhov, & Muthén, 2007) and has become a standard means of comparing competing models (e.g., Eaton et al., in press; Kim & Muthén, 2009; Muthén & Asparouhov, 2006; Walton et al., 2011; Wright et al., in press). Although the BIC performs well, we also investigated a sample-sized adjusted BIC statistic, which can outperform the standard BIC when latent class sizes are unequal (Nylund, Asparouhov, & Muthén, 2007) but has been comparatively less used for model adjudication. The incorporation of familial clustering information into our analyses precluded use of the bootstrapped likelihood ratio test.

It is standard to fit model sets in a particular way, which we followed (see Clark et al., under review). First, latent trait models with an ascending number of traits are fit until BIC indicates inclusion of additional traits results in a poorer model. We thus conducted EFAs on SPQ magical thinking items beginning with one factor, then two factors, and so on. Second, latent class models with an ascending number of classes are fit until BIC indicates inclusion of additional classes results in a poorer model. We therefore conducted LCAs on SPQ magical thinking items beginning with one class, then two classes, and so on. Finally, hybrid dimension-category models are fit for up to the number of latent traits and classes identified by the previous EFAs and LCAs, respectively. If the EFAs and LCAs had suggested two factors and three classes, one would fit the following models: two classes with one factor, two classes with two factors, three classes with one factor, and three classes with two factors. (Fitting one-class EFMA models does not make conceptual sense because, in such models, there are no categorical class distinctions, making them equivalent to EFA models that assume a single population.) We thus conducted EFMAs on SPQ magical thinking items following this procedure.

Results

The correlations among the seven magical thinking items in the total sample are given in Table 1. It is notable that these items are all positively and significantly intercorrelated, ranging from $r = .14$ to $.60$. Table 1 also includes the means and variances for each item, given separately for patients and for relatives.

Results for latent trait, class, and hybrid models are given in Table 2. The appendix provides modified example Mplus code, with annotations, for each of the three types of analyses. EFAs indicated that one latent dimensional factor (trait) provided a superior balance of fit and parsimony to a two-factor model. Given the relatively close fit of the one- and two-factor models, we also fit a three-factor model, which was markedly poorer. The one-factor model provided the best fit by BIC, although the adjusted BIC favored the two-factor model. However, the two-factor model also tended to produce Heywood cases (implausible factor loadings) in some analyses, even with alternative factor rotations, suggesting over-extraction. This further supported the superiority of the one-factor model. Factor loadings for the one-factor EFA are given in Table 3. It is noteworthy, though, that the availability of only seven magical thinking items made extraction of a multi-factor solution somewhat unlikely.

LCAs supported the presence of two classes of magical thinking (Table 2), defined by item endorsement patterns. The two-class solution provided a superior fit by BIC to the one-class solution, and improvements in BIC ended with the three-class solution, which provided inferior fit. The three-class solution also produced a notably small class composed of only 23 individuals, and examination of the item endorsement probabilities within each of its three classes did not yield a clear substantive interpretation, in contrast to the two-class solution. For these reasons, the two-class solution was selected as optimal. The proportions of individuals assigned to each of the two classes are given in Table 2. The probabilities of item endorsement in each class are given in Table 3. These endorsement patterns indicate that one class of individuals (class 1) tended to have much higher probabilities of endorsing

each item than did individuals in the other class. For instance, the probability of endorsing item 1 was .71 in class 1 but only .07 in class 2, indicating how much more frequently magical thinking items were endorsed by class 1 members. Finally, it is noteworthy, that the adjusted BIC supported a three-class solution over a two-class solution. Given that neither of these models was supported as the optimal overall model by BIC or the adjusted BIC, we focus our discussion below on the two-class model for simplicity.

A two-class, one-factor EFA was fit to the data (Table 2). Only one EFMA was fit, because LCAs and EFAs supported the presence of two classes and one factor, respectively. The proportions of individuals assigned to each class are given in Table 2. Unique factor loadings, and item endorsement probabilities, for each class are given in Table 3. While not favored by BIC, it is noteworthy that the adjusted BIC favored this model.

Overall, the superior model by BIC was the one-factor (trait) EFA model. We therefore expressed this model as a two-parameter IRT model of the seven items on the scale, to illustrate our results. Item characteristic curves, and test information for all seven items, are given in Figure 2. As can be seen in the figure, the seven items assessed for magical thinking trait (theta) levels generally between .5 to 1.5 standard deviations above the mean (i.e., zero). What this means is that the items tended to “tap into” the construct of magical thinking near its mean level for the sample. In other words, the items largely did not assess quite low or quite high ends of the magical thinking dimension. Given that the items tended to measure mean-level magical thinking best, the information provided by the items taken together as a single test centered also around the mean value and tapered off toward the ends of the magical thinking dimension. The scale as a whole provided the most information about magical thinking trait levels approximately one standard deviation above the mean. This explanation of the peaked information curve assumes a bipolar dimensional construct, but there are alternative explanations. In particular, another interpretation is that the peaked information curve could suggest that the latent magical thinking dimension is a quasi-trait, being bounded at the lower end by zero (absence of magical thinking; see Reise & Waller, 2007).

Discussion

Distributional assumptions about personality psychopathology are critical to how it is conceptualized and assessed. Measures of personality psychopathology that reflect trait, class, and hybrid disorder distributions will necessarily differ markedly, reflecting the fundamentally different ways that these competing models characterize individual differences. To illustrate this point, we applied a model-based approach to test the latent distribution of magical thinking. Magical thinking is a fundamental aspect of various forms of personality psychopathology, and previous theory and research suggests various plausible distributions. We now turn our attention to discussion of the potential assessment implications of each possible magical thinking model.

Latent trait implications—In comparison of one-, two-, and three-factor models—and all other types of models—a one-factor model provided superior fit by BIC. This suggested that the SPQ magical thinking items are structurally unidimensional. It should be noted,

however, that our use of only seven items may have made it less likely that solutions with more factors would be identified and/or favored. This dimension ranges from very low levels of magical thinking to very high levels, and there are no zones of rarity—that is, individuals occupy every trait location along the continuum. Investigation of the item characteristic curves and test information function revealed that the seven SPQ items tend to tap into the construct primarily at the mid-range ($\theta = 0$) and above, as would be expected of a measure of schizotypy.

In terms of applied personality assessment of a magical thinking trait dimension, the SPQ seems to take the appropriate approach: Multiple items of varying difficulty tap into the latent continuum. For testing magical thinking levels in the general population, it would likely be beneficial to include additional items with lower difficulty, so as to characterize lower levels of magical thinking (e.g., $\theta = -2$) with more precision. Further, some items had relatively poorer discrimination parameters (as seen by the less steep slope of their item characteristic curves), indicating that they were less strongly related to the magical thinking dimension than others, and thus did a poorer job determining participant's levels of magical thinking (θ). Finally, the SPQ magical thinking items can be summed to yield a full magical thinking index score. Each item in this summation is accorded equal weight. Our results suggest that such a summation approach is a coarse one. A finer means of creating a composite score would reflect that not all items tap into the magical thinking trait dimension at the same level or to the same degree. Using the IRT model formally to assign scores to test takers would be a better incorporation of test information for applied assessment of magical thinking.

Latent class implications—In comparison of one-, two-, and three-class models, a two-class model provided the best fit by BIC, and the adjusted BIC supported a three-class model. For brevity, we focus on interpretation of the two-class model, given that neither was selected as optimal by BIC or adjusted BIC overall. A two-class finding would have suggested that magical thinking, as assessed by the SPQ, had a latent dichotomous categorical structure. Individuals would fall into one category or the other in terms of magical thinking. Investigation of the item response probabilities in each class suggested an interpretation for the two classes. Class 1 appeared to represent individuals who were more likely to endorse magical thinking items, while class 2 appeared to represent individuals who were quite unlikely to endorse magical thinking items. In other words, this class solution provided a “magical thinkers” class (class 1) and a “non-magical thinkers” class (class 2). Also worthy of note is that, in class 1 (“magical thinkers”), items 3 and 5 had the lowest probability of endorsement; this is congruent with EFA and IRT findings that these items tap into the magical thinking trait comparatively weakly.

In terms of applied personality assessment of this two-class magical thinking structure, the goal would be the ascertainment of what every individual's most likely group membership would be. Certain items (e.g., items 1, 2, 4, and 6) had high probabilities of endorsement by “magical thinkers” and low probabilities of endorsement by “non-magical thinkers.” Items such as these are ideal for assigning individuals to a particular class, given that they are discriminate between groups so well. Items such as item 5 do a poorer job, because both classes have relatively low endorsement probabilities. Another assessment option would be

to use a dimensional measure like the SPQ, sum the seven items, and demarcate a cut-point or threshold along the dimension that best discriminated between classes.

Hybrid implications—Due to the results of the LCAs and EFAs, we fit only one hybrid EFMA model: a two-class, one-factor solution. Had this been the optimal model by BIC, it would have suggested that SPQ magical thinking comprised two latent classes of individuals; within each class, a severity dimension would be present to characterize class members' experiences. However, these dimensions would be defined differently within each class. Class 1 endorsement probabilities suggested that this class was similar to the LCA “non-magical thinkers” class, and class 2 probabilities suggested this class was similar to the LCA “magical thinkers” class. The class 1 severity dimension bore a notable similarity to the one-factor EFA model, with high loadings for most items. The exception here was for item 5, which had a strong, negative loading, suggesting that, in this class, this item tended to reflect *lower* levels of magical thinking. The class 2 severity dimension was quite different. It was defined largely by item 6 and also had notable loadings for items 2 and 4. Even still, this complex model was selected as the optimal model by the adjusted BIC, somewhat complicating interpretation of results overall.

In terms of applied personality assessment of this two-class, one-factor magical thinking structure, the goal would be the ascertainment of what every individual's most likely group membership would be and then to assess their level of magical thinking within-class via the severity dimension for that class. The approach to doing this would likely be complicated. First, class membership would have to be determined, as outlined above. Then, an appropriate assessment measure—that taps into that class's severity dimension—would be administered to characterize severity. This highlights the complexity of personality assessment if hybrid structures are present.

Concerning Model Selection

In the current study, we illustrated a wholly model-based approach to adjudicating between competing possible latent structures of magical thinking. This approach allows one or more numerical indices to indicate the optimal model; in our case, this was the BIC, which balanced each model's fit to the data with an adjustment favoring more parsimonious models. This approach has several strengths, and key among them is that it largely removes subjective decision-making from model selection. In other words, this approach can take a “dustbowl empiricist” approach to determining the optimal model.

While we believe that model-based approaches have been underutilized, it is noteworthy that it has relative weaknesses, while other approaches have relative strengths of their own. In terms of weakness, the current report notes a critical complication: The decision of which fit indices to use to adjudicate between models can have a large impact on interpretation. If multiple fit indices converge on the same model, this can strengthen inference about this model's optimality. However, if they diverge—as the BIC and the adjusted BIC did somewhat in the present analyses—one is left in something of a statistical quandary. If simulation research has strongly supported one index over another, this is less of an issue; if

there is no clear superior index, however, one is left again with making a somewhat subjective choice.

It is also noteworthy that competing, non-model-based approaches have relative strengths. Given that the goal of personality assessment research is constructing reliable and valid instruments that are also *useful*, some degree of clinical judgment and expert decision-making is inescapable. If the BIC had suggested a three-class solution as optimal, for instance, but the classes did not make intuitive sense or had limited utility in applied settings, a solely model-based approach would not be ideal. Indeed, this point can be illustrated by the two-class solution identified as best among the latent class analytic solutions. How worthwhile would it be to create an assessment device with a sole purpose of distinguishing those individuals who experience some degree of magical thinking (“magical thinkers”) from those who largely do not (“non-magical thinkers”)? To summarize, while the current report focused on the illustration of a wholly model-based approach to personality assessment, it is not the case that practical concerns such as applied utility can be forgotten. Rather, the most effective approaches will likely combine model-based assessment with real world clinical concerns to select the optimal model for a particular use and setting.

Summary

A set of statistical models and methods allows for direct tests of various distributional assumptions about core elements of personality psychopathology, but the application of these models to personality assessment is a relatively recent development. This represents a notable advance for the field, and increased application of these approaches will likely help resolve many outstanding questions regarding the nature of clinically relevant personality variation. Once accurate latent distributions are ascertained for personality construct data, appropriate assessments must be developed, because each possible distribution has different aims, and requires assessments with different psychometric properties. In the future, such investigations will likely be critical to reliable and valid personality psychopathology assessment.

One interpretive consideration here is that the current study had limitations insofar as producing the final word on the structure of magical thinking would be concerned. Although our sample was relatively small and artificially mixed (i.e., patients and relatives), it was also purposefully oversaturated with psychopathology to increase the likelihood of proper coverage of the construct. This has the advantage of ensuring that we were studying clinically relevant variation, but the potential disadvantage of creating a sample that does not represent a specific defined population (e.g., community dwelling residents in a specific catchment area). Nevertheless, we took this approach because the majority of previous modeling research has been in community as opposed to clinical samples, such that extension to more clinically saturated samples is important in ensuring direct clinical relevance. Ultimately, however, it will be important to consider findings from both diverse models *and* diverse sampling frames to arrive at a fuller understanding of the basic nature of variation in clinically relevant individual differences.

While magical thinking was used as an illustration in the current report, other core elements of personality psychopathology should similarly be subjected to such analysis. Subsequently, how those core constructs combine into syndromal presentations can then be explored empirically (e.g., Eaton, Krueger, South, Simms, & Clark, 2011). Indeed, schizotypy is likely a constellation of multiple constructs, of which magical thinking is only one. It could be additionally beneficial to examine mixtures of items from more than one measure of a given construct. Investigation and elaboration of such core elements is a first step toward understanding the broader, higher-order syndromes of personality psychopathology.

Appendix

Example Mplus code for EFA analysis with annotations

```
TITLE: Magical thinking;

DATA: FILE IS SPQ_Data.dat;

!Identifying ID and family numbers and the seven magical thinking items;

VARIABLE: NAMES ARE idnum famnum mag1 mag2 mag3 mag4 mag5 mag6 mag7;

!Defining missing values and familial clustering;

MISSING = ALL(-999);

CLUSTER IS famnum;

!Denoting the seven magical thinking items for analysis and as categorical;

USEVARIABLES ARE mag1 mag2 mag3 mag4 mag5 mag6 mag7;

CATEGORICAL ARE mag1 mag2 mag3 mag4 mag5 mag6 mag7;

!Indicating the number of factors to extract. In this case, one-, two-, and

!three-factor solutions will be modeled. The COMPLEX command allows for

!incorporation of familial clustering;

ANALYSIS: TYPE = EFA 1 3 COMPLEX;

!Selecting the use of a robust maximum likelihood estimator;

ESTIMATOR = MLR;
```

Example Mplus code for LCA analysis with annotations

```

TITLE: Magical thinking;

DATA: FILE IS SPQ_Data.dat;

!Identifying ID and family numbers and the seven magical thinking items;

VARIABLE: NAMES ARE idnum famnum mag1 mag2 mag3 mag4 mag5 mag6 mag7;

!Defining missing values and familial clustering;

MISSING = ALL(-999);

CLUSTER IS famnum;

!Denoting the seven magical thinking items for analysis and as categorical;

USEVARIABLES ARE mag1 mag2 mag3 mag4 mag5 mag6 mag7;

CATEGORICAL ARE mag1 mag2 mag3 mag4 mag5 mag6 mag7;

!Indicating the number of classes to model, in this case two classes;

CLASSES = c(2);

!Indicating a class-based analysis. The COMPLEX command allows for

!incorporation of familial clustering;

ANALYSIS: TYPE = MIXTURE COMPLEX;

!Defining the number of starting values and final stage iterations;

STARTS = 100 50;

```

Example Mplus code for EFMA analysis with annotations

```

TITLE: Magical thinking;

DATA: FILE IS SPQ_Data.dat;

!Identifying ID and family numbers and the seven magical thinking items;

VARIABLE: NAMES ARE idnum famnum mag1 mag2 mag3 mag4 mag5 mag6 mag7;

!Defining missing values and familial clustering;

```

```

MISSING = ALL(-999);

CLUSTER IS famnum;

!Denoting the seven magical thinking items for analysis and as categorical;

USEVARIABLES ARE mag1 mag2 mag3 mag4 mag5 mag6 mag7;

CATEGORICAL ARE mag1 mag2 mag3 mag4 mag5 mag6 mag7;

!Indicating the number of classes to model, in this case two classes;

CLASSES = c(2);

!Indicating the number of factors to extract. In this case, one-, two-, and

!three-factor solutions will be modeled. The COMPLEX command allows for

!incorporation of familial clustering;

ANALYSIS: TYPE = MIXTURE COMPLEX EFA 1 3;

!Defining the number of starting values and final stage iterations;

STARTS = 100 50;

```

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd. Washington, DC: Author; 1994.
- Brown, TA. Confirmatory factor analysis for applied research. New York: Guilford Press; 2006.
- Browne MW. An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*. 2001; 36(1):111–150.
- Clark LA. Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual Review of Psychology*. 2007; 58:227–257.
- Clark, SL.; Muthén, B.; Kaprio, Jaako; D'Onofrio, BM.; Viken, R.; Rose, RJ.; Smalley, SL. Models and strategies for factor mixture analysis: Two examples concerning the structure underlying psychological disorders. 2012. Manuscript submitted for publication available at <http://statmodel.com/papers.shtml>
- Costa, PT., Jr; McCrae, RR. Professional manual for the revised NEO Personality Inventory. Odessa, FL: Psychological Assessment Resources; 1992.
- Eaton NR, Krueger RF, Markon KE, Keyes KM, Skodol AE, Wall M, Hasin DS, Grant BF. The structure and predictive validity of the internalizing disorders. *Journal of Abnormal Psychology*. in press.
- Eaton NR, Krueger RF, South SC, Simms LJ, Clark LA. Contrasting prototypes and dimensions in the classification of personality pathology: Evidence that dimensions, but not prototypes, are robust. *Psychological Medicine*. 2011; 41:1151–1163. [PubMed: 20860863]
- Embretson, SE.; Reise, SP. Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.

- First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV). Washington, DC: American Psychiatric Press, Inc; 1996.
- Haslam N. The latent structure of personality and psychopathology: A review of trends in taxometric research. *Scientific Review of Mental Health Practice*. 2011; 8:17–29.
- Hipp JR, Bauer DJ. Local solutions in the estimation of growth mixture models. *Psychological Methods*. 2006; 11(1):36–53. [PubMed: 16594766]
- Johansen M, Karterud S, Pedersen G, Gude T, Falkum E. An investigation of the prototype validity of the borderline *DSM-IV* construct. *Acta Psychiatrica Scandinavica*. 2004; 109:289–298. [PubMed: 15008803]
- Kendell R, Jablensky A. Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry*. 2003; 160:4–12. [PubMed: 12505793]
- Kim Y, Muthén BO. Two-part factor mixture modeling: Application to an aggressive behavior measurement instrument. *Structural Equation Modeling*. 2009; 16:602–624. [PubMed: 20717486]
- Krueger RF, Eaton NR. Personality traits and the classification of mental disorders: Toward a more complete integration in *DSM-5* and an empirical model of psychopathology. *Personality Disorders: Theory, Research, and Treatment*. 2010; 1(2):97–118.
- Leckman JF, Sholomskas D, Thompson WD, Belanger A, Weissman MM. Best estimate of lifetime psychiatric diagnosis: a methodological study. *Archives of General Psychiatry*. 1982; 39:879–83. [PubMed: 7103676]
- Markon KE, Krueger RF. Categorical and continuous models of liability to externalizing disorders: A direct comparison in NESARC. *Archives of General Psychiatry*. 2005; 62:1352–1359. [PubMed: 16330723]
- Masyn KE, Henderson CE, Greenbaum PE. Exploring the latent structures of psychological constructs in social development using the dimensional-categorical spectrum. *Social Development*. 2010; 19(3):470–493. [PubMed: 24489441]
- McCutcheon, AL. Latent class analysis. Newbury Park, CA: Sage Publications, Inc; 1987.
- Miller JD, Campbell WK. Comparing clinical and social-personality conceptualizations of narcissism. *Journal of Personality*. 2008; 76(3):449–476. [PubMed: 18399956]
- Miller JD, Hoffman BJ, Campbell WK, Pilkonis PA. An examination of the factor structure of *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*, narcissistic personality disorder criteria: one or two factors? *Comprehensive Psychiatry*. 2008; 49:141–145. [PubMed: 18243885]
- Muthén B. Should substance use disorders be considered as categorical or dimensional? *Addiction*. 2006; 101(Suppl. 1):6–16. [PubMed: 16930156]
- Muthén B, Asparouhov T. Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors*. 2006; 31:1050–1066. [PubMed: 16675147]
- Muthén, LK.; Muthén, BO. *Mplus user's guide*. 6th. Los Angeles: Author; 2011.
- Nurnberger J Jr, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T. Diagnostic Interview for Genetic Studies: Rationale, unique features, and training. NIMH Genetics Initiative. *Archives of General Psychiatry*. 1994; 51:849–859. [PubMed: 7944874]
- Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*. 2007; 14(4):535–569.
- Pincus AL, Lukowitsky MR. Pathological narcissism and narcissistic personality disorder. *Annual Review of Clinical Psychology*. 2010; 6:421–446.
- Raine A. The SPQ: a scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophrenia Bulletin*. 1991; 17:555–64. [PubMed: 1805349]
- Raftery AE. Bayesian model selection in social research. *Sociological Methodology*. 1995; 25:111–163.
- Rawlings D, Williams B, Haslam N, Claridge G. Taxometric analysis supports a dimensional latent structure for schizotypy. *Personality and Individual Differences*. 2008; 44:1640–1651.

- Reise SP, Waller NG. Item response theory and clinical measurement. *Annual Review of Clinical Psychology*. 2009; 5:27–48.
- Sponheim SR, Steele VR, McGuire KA. Verbal memory processes in schizophrenia patients and biological relatives of schizophrenia patients: intact implicit memory, impaired explicit recollection. *Schizophrenia Research*. 2004; 71:339–348. [PubMed: 15474904]
- Trull TJ, Durrett CA. Categorical and dimensional models of personality disorder. *Annual Review of Clinical Psychology*. 2005; 1:355–380.
- Waller, NG.; Meehl, PE. *Multivariate taxometric procedures: Distinguishing types from continua*. Newbury Park, CA: Sage; 1997.
- Walton KE, Ormel J, Krueger RF. The dimensional nature of externalizing behaviors in adolescence: Evidence from a direct comparison of categorical, dimensional, and hybrid models. *Journal of Abnormal Child Psychology*. in press.
- Widiger TA, Samuel DB. Diagnostic categories or dimensions? A question for the *Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition*. *Journal of Abnormal Psychology*. 2005; 114:494–504. [PubMed: 16351373]
- Wright AGC, Krueger RF, Hobbs MJ, Markon KE, Eaton NR, Slade T. The structure of psychopathology: Toward an expanded quantitative empirical model. *Journal of Abnormal Psychology*. in press.

A dimensional structure *A class structure* *A hybrid class-dimension structure*

Figure 1. Hypothetical distributions of individuals in dimensional, class, and hybrid structures of narcissism.

Figure 2.
Item information curves (top) and test information (bottom) for SPQ magical thinking items.

Table 1
Means, variances, and correlations among magical thinking items

	Item						
	1	2	3	4	5	6	7
							M_{Patient}
							s^2_{Patient}
Item 1							.31
Item 2	.45						.37
Item 3	.39	.28					.22
Item 4	.36	.56	.17				.43
Item 5	.35	.18	.33	.14*			.28
Item 6	.49	.49	.23	.60	.23		.38
Item 7	.47	.53	.21	.42	.25	.41	.42
M_{Relative}	.09	.17	.06	.16	.13	.16	.15
s^2_{Relative}	.08	.14	.06	.14	.11	.14	.13

Note: M (mean) and s^2 (variance) are given for both patient and relative sub-groups. All values are significant at $p < .001$ except * which is significant at $p = .004$. See text for item content.

Table 2

Model fit results

Model	LL	BIC	BIC _A	Class Proportions		
				1	2	3
<i>EFA</i>						
1-factor	-866.456	1812.811	1768.411			
2-factor	-849.936	1814.015	1750.586			
3-factor	-845.850	1834.377	1755.091			
<i>LCA</i>						
1-class	-1074.273	2188.496	2166.296	1.00		
2-class	-870.069	1825.744	1778.173	.27	.73	
3-class	-852.749	1836.762	1763.819	.08	.73	.19
<i>EFMA</i>						
2-class, 1-factor	-838.139	1841.785	1749.814	.81	.19	

Note: *LL*: log-likelihood. BIC: Bayesian information criterion. BIC_A: sample-size adjusted Bayesian information criterion. EFA: exploratory factor analysis. LCA: latent class analysis. EMFA: exploratory factor mixture analysis. Class proportions reflect the proportion of individuals assigned to each class in class-based models. The optimal model by BIC and BIC_A are bolded.

Table 3

EFA and LCA results

	EFA Loadings	LCA Probabilities		EFMA Loadings		EFMA Probabilities	
		Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
Item 1	.81	.71	.07	.82	.39	.17	.61
Item 2	.83	.70	.07	.87	.62	.17	.58
Item 3	.56	.44	.08	.30	-.33	.07	.67
Item 4	.78	.69	.10	.92	.69	.22	.48
Item 5	.58	.32	.04	-.95	.30	.03	.52
Item 6	.85	.71	.07	.78	.99	.18	.52
Item 7	.87	.58	.01	.99	.34	.05	.64

Note: EFA: exploratory factor analysis. LCA: latent class analysis. EFMA: exploratory factor mixture analysis. LCA and EFMA probabilities are the probabilities of individuals in each class endorsing each item in the positive (present) direction in the two-class model. See text for item content.