

# A Taxometric Investigation of *DSM-IV* Major Depression in a Large Outpatient Sample

## Interpretable Structural Results Depend on the Mode of Assessment

John Ruscio

*The College of New Jersey*

Timothy A. Brown

*Boston University*

Ayelet Meron Ruscio

*University of Pennsylvania*

Most taxometric studies of depressive constructs have drawn indicators from self-report instruments that do not bear directly on the *Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV)* diagnostic construct of major depressive disorder (MDD). The present study examined the latent structure of MDD using indicator sets constructed from a semistructured clinical interview, self-report questionnaires, and a combination of the two. Taxometric analyses were performed in a large sample of outpatients with primary mood or anxiety disorders. For clinical rating data, results were more consistent with taxonic than dimensional structure, an interpretation supported by additional data obtained from 27 independent raters and objective indices of curve fit. Questionnaire indicators yielded inconclusive results, and combined rating-questionnaire indicators yielded results suggestive of taxonic structure. The findings highlight the importance of assessment in the study of a construct's latent structure and extend recent findings suggesting that MDD may be taxonic.

**Keywords:** *taxometrics; major depressive disorder; latent structure; clinical interview; self-report questionnaires*

Questions about the boundary separating mental disorders from normal psychological states have long been debated within the mental health community. Such questions have taken on renewed relevance and urgency as preparations begin for revisions of the major classification systems of mental disorders—the *Diagnostic and Statistical Manual of Mental Disorders (DSM)* and the *International Classification of Diseases*. For example, in response to criticisms of the purely categorical model of psychopathology that has existed in all editions of the *DSM* (e.g., Brown, 1996; Widiger & Frances, 2002), the question of whether mental disorders operate in a categorical or dimensional fashion has been identified as one of the most important areas of study in the research planning for *DSM-V*. Thus, the forthcoming revisions, perhaps more than any of their predecessors, are likely to be informed by research examining the latent structure of psychopathology,

research that is made possible by statistical techniques designed to test the relative fit of competing models of latent structure. One such technique, the taxometric method (e.g., Meehl, 1995; Waller & Meehl, 1998), addresses what may be the most fundamental of all boundary questions: whether a purely dimensional (continuous) model or a taxonic (categorical) model—which allows dimensional variation within categories—best represents the structure of a disorder (J. Ruscio & Ruscio, 2004a). Taxometrics thus has the potential to significantly influence psychiatric classification systems, helping to locate optimal diagnostic thresholds

**Authors' Note:** This research was supported by a grant from the National Institute of Mental Health (RO1 MH39096) awarded to the second author. Please address correspondence to John Ruscio, Psychology Department, The College of New Jersey, 2000 Pennington Road, Ewing, NJ 08628; e-mail: [ruscio@tnj.edu](mailto:ruscio@tnj.edu).

for taxonic disorders and challenging the categorical classification of disorders that lie on continua with normal functioning.

Although taxometrics has been applied to a variety of psychopathological constructs (Haslam & Kim, 2002; Ruscio, Haslam, & Ruscio, 2006), there has been particular interest in its application to the study of depression. This may be due not only to the high prevalence, cost, and health care policy significance of depressive disorders but also to disagreement in the field about the relation of depression to normal dysphoric states. Reviewing the evidence for competing structural models of depression, Flett, Vredenburg, and Krames (1997) noted the limitations of inferring latent structure from observed score distributions and called for taxometric research to help resolve the structure of depression. Answering this call, a small but growing number of taxometric studies have examined depressive phenomena. Some of these studies have focused on depressive subtypes (Ambrosini, Bennett, Cleland, & Haslam, 2002; Grove et al., 1987; Haslam & Beck, 1994) or vulnerability factors for depression (Gibb, Alloy, Abramson, Beevers, & Miller, 2004; Strong, Brown, Kahler, Lloyd-Richardson, & Niamura, 2004). However, seven studies investigated the latent structure of major depressive disorder (MDD) or closely related constructs.

J. Ruscio and Ruscio (2000) explored the structure of depression in two clinical samples: male combat veterans at an outpatient Veterans Administration clinic and the Hathaway Data Bank, an archive of Minnesota Multiphasic Personality Inventories (MMPIs; Hathaway & McKinley, 1943). Patients' responses to three self-report measures of depression—the Beck Depression Inventory (BDI), the Zung Self-rating Depression Scale, and the MMPI Depression scale (Scale 2)—were used to construct multiple sets of indicator variables, and all taxometric analyses yielded dimensional results. A. M. Ruscio and Ruscio (2002) extended this research to an analogue sample to determine whether commonly used BDI cut scores for identifying “depressed” or “dysphoric” individuals correspond to meaningful latent groups. Taxometric analyses using the 10 BDI items that best distinguished high scorers and low scorers on the questionnaire failed to support popular cut scores or a taxonic structural model.

Beach and Amir (2003) evaluated the latent structure of involuntary defeat syndrome, a depressive state characterized by chiefly somatic symptoms of

homeostatic disruption believed to serve the evolutionarily adaptive function of rendering the depressed individual nonthreatening to dominant others in competitive situations (Gilbert, 1992, 2000). Beach and Amir (2003) selected six BDI items with predominantly somatic content, analyzed them in a college student sample, and concluded that they had uncovered an involuntary defeat syndrome taxon. J. Ruscio, Ruscio, and Keane (2004) replicated Beach and Amir's analyses. Taxometric graphs were very similar to those obtained by Beach and Amir, but results for comparison data (see Method for an overview of this approach) showed they were more consistent with skewed indicators of a latent dimension than a taxon. The dimensional conclusion was supported by analyses in a clinical data set in which responses to the BDI were much less skewed.

These four studies favor a dimensional structural model for depression. At the same time, key limitations of this literature temper the confidence that can be held in this structural conclusion. First, taxometric studies of depression have relied heavily on one particular measure of the construct—the BDI. Although the BDI has been shown to have strong psychometric properties in a variety of populations (Beck & Steer, 1993; Beck, Steer, & Garbin, 1988; Lips & Ng, 1985), some authors have suggested that it may represent a nonspecific measure of general distress in nonclinical samples (Coyne, 1994; Deardorff & Funabiki, 1985). Moreover, BDI items have a highly constrained (4-point) response scale that, though able to be combined into composites for taxometric analysis, may be less desirable than items providing more precise, continuous measurement of each depressive symptom. In addition, the BDI item set does not correspond well to diagnostic definitions of MDD (e.g., the BDI contains items that assess symptoms beyond the diagnostic criteria of *DSM-IV*). Finally, the limited use of other measures has made it difficult to ascertain whether the dimensional results consistently obtained with the BDI reflect the latent structure of depression or the structure of this measure in particular.

Past taxometric studies of depression have not only used the BDI very often, but they have more generally been limited by a heavy reliance on self-report data. It is possible that self-report measures are not sufficiently sensitive to high trait levels of depressive symptoms to detect a severe depression taxon (P. E. Meehl, personal communication, May 10, 2001) and that ratings by clinicians trained to evaluate a wide range of trait levels may be more sensitive to such a

taxon. Several taxometric studies of depression have made use of clinical rating data. Two of these examined the latent structure of MDD among children and adolescents, with mixed results. Hankin, Fraley, Lahey, and Waldman (2005) used a general population-based sample of children and adolescents, with depressive symptoms assessed by self-responses and caregiver responses to a structured clinical interview of *DSM-IV* criteria. Their taxometric analyses yielded dimensional results. Solomon, Ruscio, Seeley, and Lewinsohn (2006) used data from the Oregon Adolescent Depression Project, a large longitudinal study that included clinical interviews with all participants at multiple time points. Their taxometric analyses yielded taxonic results.

In what may be the most similar study to the present investigation, J. Ruscio, Zimmerman, McGlinchey, Chelminsky, and Young (2007) performed analyses in a large outpatient sample of adults who had completed semistructured clinical interviews. These data afforded taxometric analyses of indicators that closely match *DSM-IV* criteria for MDD, and the results were taxonic. Because clinical interviews are generally regarded as the gold standard in the assessment of depression, further taxometric analysis of interview data seems warranted. Even among studies that included clinical ratings, the comparison of taxometric results obtained with self-report when compared with clinical interview data has not yet been examined.

A final limitation of the existing database is that the samples in prior studies may not have afforded the most powerful tests between taxonic and dimensional structures. The majority of these studies were performed with college student or community samples, which may possess too few taxon members to permit detection of an MDD taxon. Even past clinical samples have had notable limitations. For example, the outpatient sample employed by J. Ruscio and Ruscio (2000) and J. Ruscio et al. (2004) consisted exclusively of male combat veterans presenting for assessment or treatment of posttraumatic stress disorder. The second sample studied by J. Ruscio and Ruscio (2000) was perhaps less selected but contained an unknown number of cases diagnosed (or diagnosable) with MDD; this number is important because taxometric procedures more easily detect taxa whose base rate in a sample is moderate rather than extreme (Meehl & Yonce, 1994, 1996). The fact that taxonic results emerged most clearly in J. Ruscio, Zimmerman, et al. (2007) may be due to the substantial representation of individuals qualifying for MDD

diagnoses in this sample (nearly 60%). There is thus a need for taxometric research with clinical samples containing a mixture of males and females as well as a sizable base rate of MDD.

The present study adds to this literature by addressing each of several limitations. Taxometric analyses were performed in a large, treatment-seeking sample presenting to an outpatient anxiety and mood disorders clinic. The sample contained both male and female patients, a substantial number of whom were diagnosed with MDD. In addition to completing a battery of self-report measures, all participants were assessed by a semistructured clinical interview—the Anxiety Disorders Interview Schedule for *DSM-IV*: Lifetime version (ADIS-IV-L; Di Nardo, Brown, & Barlow, 1994)—which represents a gold standard in the clinical assessment and differential diagnosis of mood and anxiety disorders. Each symptom of major depression was assessed along a 9-point Likert scale, providing a sufficient range of responses to serve as indicators in taxometric analyses.

The present study also made use of several procedural safeguards to protect against unwarranted or erroneous structural inferences. Unlike early studies, which relied on estimates of data parameters to evaluate the appropriateness of available data for taxometric analysis according to rules of thumb (Meehl, 1995), we generated taxonic and dimensional comparison data sets that reproduced the unique characteristics of our research data (J. Ruscio, Ruscio, & Meron, 2007). By submitting the comparison data to taxometric analysis, we were able to determine empirically whether differences across data known to be taxonic or dimensional could be detected while holding constant the characteristics of our data. This technique also enabled us to calculate an objective index that has been found to be valid, surpassing the performance of alternative tests in a number of rigorous studies (J. Ruscio, 2007; J. Ruscio & Marcus, 2007; J. Ruscio, Ruscio, et al., 2007). In addition to interpreting the taxometric results ourselves, both objectively and subjectively, we asked multiple judges who were blind to the purpose and hypotheses of the study to rate the results.

## Method

### Participants

The sample consisted of 1,500 outpatients who presented for assessment and treatment at the Center for

Anxiety and Related Disorders at Boston University. Women constituted the larger portion of the sample (61%); the average age was 33.15 years ( $SD = 11.21$ ; range = 18-75). The sample was predominantly Caucasian (89%; African American = 3.5%; Asian = 4%; Latino/Hispanic = 3%). Diagnoses were established with the ADIS-IV-L (Di Nardo et al., 1994), a semistructured interview designed to ascertain reliable diagnosis of the *DSM-IV* anxiety, mood, somatoform, and substance use disorders and to screen for the presence of other conditions (e.g., psychotic disorders). For each diagnosis, interviewers assign a 0 to 8 clinical severity rating (CSR) that indicates the degree of distress and impairment associated with the disorder (0 = none; 8 = very severely disturbing/disabling). In patients with two or more current diagnoses, the "principal" diagnosis is the one receiving the highest CSR. Other disorders that meet or surpass the threshold for a formal *DSM-IV* diagnosis are referred to as "additional" diagnoses. A reliability study of a subset of the current sample ( $N = 362$ ) who had two independent administrations of the ADIS-IV-L indicated good to excellent interrater agreement for principal disorders (range of  $\kappa$ s = .67-.86) except dysthymia ( $\kappa = .22$ );  $\kappa$  for MDD was .67 (Brown, Di Nardo, Lehman, & Campbell, 2001). For a detailed description of interviewers, training procedures, and related information, see Brown et al. (2001).

The rates of common principal disorders in the sample (i.e., >50 cases) were as follows: panic disorder with or without agoraphobia ( $n = 427$ ), social phobia ( $n = 327$ ), specific phobia ( $n = 144$ ), generalized anxiety disorder ( $n = 114$ ), major depression ( $n = 113$ ), and obsessive-compulsive disorder ( $n = 88$ ); 139 patients were assigned coprincipal diagnoses (i.e., two disorders were judged to be associated with equivalent levels of impairment/distress). Current MDD was assigned at a  $CSR \geq 4$  to 400 cases, or 26.7% of the sample (principal diagnosis  $n = 113$ ; coprincipal diagnosis  $n = 11$ ; additional diagnosis  $n = 276$ ). Slightly more than half (52%) of these 400 patients had recurrent MDD; nearly a third (31%) met criteria for the *DSM-IV* "chronic" specifier (i.e., duration greater than 2 years). More than half (55%) of the MDD cases met criteria for the *DSM-IV* "moderate" severity specifier ("mild" = 21%; "severe" = 17.5%; "partial remission" = 6.5%).

## Measures

*ADIS-IV-L.* In addition to rendering *DSM-IV* diagnoses, the ADIS-IV-L provides dimensional assessment of the key and associated features of disorders

(0-8 ratings). For the purposes of the present study and other studies at Boston University, the ADIS-IV-L was structured such that the features of MDD were dimensionally rated regardless of whether a formal *DSM-IV* diagnosis was under consideration. Thus, interviewers assigned dimensional ratings of the nine constituent symptoms of *DSM-IV* MDD to all patients. In the reliability study noted previously (Brown et al., 2001), the interrater reliability of these dimensional severity ratings of MDD features was found to be quite satisfactory ( $r = .74$ ).<sup>1</sup>

*Questionnaires.* In addition to the ADIS-IV-L, patients completed a battery of questionnaires as part of their initial intake evaluation at the Center for Anxiety and Related Disorders. With one exception (see below), all indicators used in the current study were selected from two intake questionnaires: (a) the BDI (Beck & Steer, 1993) and (b) the Depression Anxiety Stress Scales (DASS; Lovibond & Lovibond, 1995a). The BDI is a 21-item, widely used and psychometrically well-established measure of current (past week) depression. The DASS is a 42-item questionnaire that consists of three subscales (14 items each) designed to measure the current (past week) negative emotional states of depression, anxiety, and stress/negative affect. Items are self-rated on a scale from 0 (*did not apply to me at all*) to 3 (*applied to me very much, or most of the time*). Several studies using clinical and nonclinical participants have produced converging evidence attesting to the favorable psychometric properties of the DASS (e.g., Antony, Bieling, Cox, Enns, & Swinson, 1998; Brown, Chorpita, Korotitsch, & Barlow, 1997; Crawford & Henry, 2003; Lovibond & Lovibond, 1995b). In the present study, 11 items from the DASS-Depression scale, 3 items from the DASS-Stress scale, and 13 items from the BDI were used. Because only 1 item assessing concentration difficulties was available from the two questionnaires (BDI Item 13); a single item from the Padua Inventory (PI; Sanavio, 1988), a 60-item measure of obsessions and compulsions, was also used to construct the indicator of this MDD symptom ("I find it difficult to make decisions, even about unimportant matters," rated on a scale from 0 = *very slightly or not at all* to 4 = *extremely*).

## Constructing Indicator Sets

We constructed multiple sets of indicators not only to evaluate the consistency of results, as is common in taxometric investigations, but also to determine

**Table 1**  
**Correlations and Validity Estimates for Clinical Rating Indicators**

	Correlations: Full Sample (Taxon/Complement)						
	#1/2	#3	#4	#5	#6	#7	#8
Symptom #3	.50 (.23/.21)						
Symptom #4	.55 (.26/.25)	.44 (.28/.22)					
Symptom #5	.51 (.28/.24)	.38 (.17/.21)	.46 (.26/.30)				
Symptom #6	.63 (.34/.40)	.43 (.22/.21)	.52 (.29/.33)	.49 (.29/.33)			
Symptom #7	.55 (.20/.28)	.35 (.02/.17)	.41 (.15/.17)	.37 (.06/.22)	.45 (.15/.24)		
Symptom #8	.61 (.38/.33)	.39 (.11/.20)	.46 (.20/.25)	.48 (.25/.34)	.53 (.21/.37)	.50 (.16/.34)	
Symptom #9	.47 (.26/.19)	.26 (.05/.03)	.30 (.04/.13)	.31 (.13/.09)	.32 (.04/.15)	.38 (.21/.16)	.36 (.17/.14)

	Validity Estimates							
	#1/2	#3	#4	#5	#6	#7	#8	#9
Taxon <i>M</i>	1.28	0.83	0.88	0.79	0.93	0.89	0.91	0.77
Taxon <i>SD</i>	0.63	1.24	0.93	1.16	0.78	0.94	0.92	1.46
Complement <i>M</i>	-0.46	-0.30	-0.32	-0.29	-0.34	-0.32	-0.33	-0.28
Complement <i>SD</i>	0.64	0.68	0.82	0.75	0.85	0.80	0.81	0.54
Validity ( <i>d</i> )	2.73	1.31	1.41	1.23	1.54	1.45	1.47	1.19
Skew	0.67	1.44	0.61	1.18	0.32	0.59	0.54	2.40

Note: MDD = major depressive disorder.  $N = 1,500$ ; taxon  $n = 400$  MDD+; complement  $n = 1,100$  MDD-. All indicators have been standardized. Symptom numbers refer to *DSM-IV* criteria for MDD: #1 = depressed mood; #2 = loss of interest/pleasure in activities; #3 = weight change/appetite disturbance; #4 = sleep disturbance; #5 = psychomotor agitation/retardation; #6 = fatigue/energy loss; #7 = worthlessness/guilt; #8 = concentration difficulties/indecisiveness; #9 = suicidality/thoughts of death. Indicator validity is expressed as Cohen's *d*, the mean difference between the taxon and complement standardized using pooled within-groups variances (weighted by *df*).

whether some instruments would yield more informative taxometric results than others. Indicator sets were drawn from the rating data, the questionnaire data, and combinations of the rating and questionnaire data. In each case, we constructed indicators first by broadly representing as many *DSM-IV* symptoms of MDD as possible and second by narrowing the set of indicators to those specific to MDD. Thus, we began by examining nine potential indicators that corresponded to all *DSM-IV* MDD symptoms. In the rating data, there was one rating per symptom; each was standardized prior to the analysis. In the questionnaire data, there were differing numbers of BDI and DASS items relevant to each symptom. Each of the authors independently judged which items assessed each symptom, and disagreements were subsequently discussed to arrive at a consensual aggregation of items used to represent each symptom.<sup>2</sup> These composites were then standardized; for composites including both rating and questionnaire data, the corresponding variables for each symptom were averaged and then restandardized.

Three sources of information were used to evaluate the appropriateness of each candidate set of indicators for the planned taxometric analyses. First, we

compared the indicator correlation matrix in the full sample to the corresponding matrices computed within each putative group (using current MDD diagnosis to denote putative taxon and complement classes). Although some researchers have suggested threshold values for the largest acceptable within-group correlations, the *difference* between correlations in the full sample (which result from the mixture of groups plus within-group associations) and correlations within groups may be more informative in assessing the likely appropriateness of data for taxometric analyses (Ruscio et al., 2006). For example, within-group correlations of  $r = .20$  may be tolerably low with full-sample correlations of  $r = .50$  but highly problematic with full-sample correlations of  $r = .25$ . Thus, rather than applying a threshold in judging whether within-group correlations were tolerably low, we considered the difference between full-sample and within-group indicator correlations (see the top portions of Tables 1, 2, and 3).

Second, we estimated the validity with which each indicator distinguished the putative taxon and complement classes. Using current MDD diagnosis as a fallible criterion, we calculated the *M* and *SD* of each

**Table 2**  
**Correlations and Validity Estimates for Questionnaire Indicators**

	Correlations: Full Sample (Taxon/Complement)			
	#1/2/7/9	#3	#5	#6
Symptom #3	.34 (.27/.22)			
Symptom #5	.58 (.41/.53)	.28 (.21/.21)		
Symptom #6	.64 (.48/.53)	.32 (.21/.24)	.46 (.30/.37)	
Symptom #8	.55 (.42/.46)	.23 (.15/.15)	.44 (.26/.39)	.49 (.40/.38)

	Validity Estimates				
	#1/2/7/9	#3	#5	#6	#8
Taxon <i>M</i>	0.97	0.45	0.59	0.79	0.59
Taxon <i>SD</i>	0.94	1.26	0.86	0.96	0.95
Complement <i>M</i>	-0.35	-0.16	-0.22	-0.29	-0.22
Complement <i>SD</i>	0.76	0.83	0.96	0.85	0.93
Validity ( <i>d</i> )	1.63	0.63	0.87	1.22	0.87
Skew	0.88	1.85	0.05	0.51	0.30

Note: MDD = major depressive disorder.  $N = 1,500$ ; taxon  $n = 400$  MDD+; complement  $n = 1,100$  MDD-. All indicators have been standardized. Symptom numbers refer to *DSM-IV* criteria for MDD. Indicator validity is expressed as Cohen's *d*, the mean difference between the taxon and complement standardized using pooled within-groups variances (weighted by *df*).

indicator within groups and then Cohen's *d* as the measure of group separation. Monte Carlo studies provide limited guidance regarding how valid indicators must be to yield informative taxometric results. Meehl (1995) suggested as a rule of thumb that  $d = 1.25$  constitutes an acceptable minimum validity, and a similar threshold value of  $d = 1.20$  was recommended by Beauchaine and Beauchaine (2002) for one particular taxometric procedure. Rather than applying a single threshold in determining when to retain or revise particular indicators, we considered indicator validity estimates (presented in the bottom portions of Tables 1, 2, and 3) along with the two other sources of information relevant to indicator construction.

Third, for each candidate indicator set, we generated and analyzed taxonic and dimensional comparison data. This empirical approach enables a judgment of the adequacy of a particular indicator set, submitted to a particular analysis, by testing whether different results emerge for data sets known to be taxonic or dimensional. To the extent that results differ for taxonic and dimensional comparison data, this suggests that the characteristics of the indicator set are acceptable for this specific taxometric analysis. Not only does this help to evaluate the adequacy of the data and the analysis plan but also provides an invaluable interpretive aid if one does proceed with the analysis (more on this below).

In all three data sets (rating data, questionnaire data, and combined rating and questionnaire data), data on the first two MDD symptoms (depressed mood and anhedonia) had to be combined because of problematically high within-group correlations (e.g., the clinical ratings of these symptoms were correlated .59 and .64 in the MDD+ and MDD- groups, respectively). In the rating data and the combined rating and questionnaire data, this left a total of eight indicators that appeared satisfactory for the planned analyses (see Tables 1 and 3, respectively). Within each of these two broad sets of indicators, a subset of five depression-specific indicators was selected for additional analysis. These depression-specific indicators corresponded to Symptoms 1 and 2 (depressed mood, anhedonia), 3 (weight change/appetite disturbance), 5 (psychomotor agitation/retardation), 7 (worthlessness/guilt), and 9 (suicidality).

In the questionnaire data, additional considerations (e.g., high within-group correlations, poor group separation, poor discrimination of taxometric results for taxonic and dimensional comparison data) led us to combine Symptoms 7 (worthlessness/guilt) and 9 (suicidality) with Symptoms 1 and 2 and to drop Symptom 4 (sleep disturbance). This resulted in a set of five indicators drawn from the questionnaire data (see Table 2; for supplementary analyses of an additional set of eight indicators constructed from the questionnaire data, see Footnote 3).

**Table 3**  
**Correlations and Validity Estimates for Combined Clinical Rating**  
**and Questionnaire Indicators**

	Correlations: Full Sample (Taxon/Complement)						
	#1/2	#3	#4	#5	#6	#7	#8
Symptom #3	.51 (.31/.29)						
Symptom #4	.51 (.21/.28)	.44 (.33/.25)					
Symptom #5	.64 (.41/.47)	.42 (.24/.26)	.45 (.24/.28)				
Symptom #6	.72 (.47/.55)	.45 (.25/.28)	.49 (.19/.34)	.58 (.34/.42)			
Symptom #7	.71 (.41/.57)	.37 (.11/.18)	.38 (.07/.17)	.53 (.19/.41)	.56 (.28/.37)		
Symptom #8	.63 (.44/.43)	.37 (.15/.21)	.41 (.23/.21)	.56 (.31/.43)	.59 (.35/.44)	.60 (.32/.48)	
Symptom #9	.72 (.52/.63)	.37 (.12/.17)	.36 (.02/.20)	.47 (.20/.32)	.50 (.18/.34)	.63 (.43/.51)	.49 (.25/.32)

	Validity Estimates							
	#1/2	#3	#4	#5	#6	#7	#8	#9
Taxon <i>M</i>	1.22	0.76	0.80	0.85	0.97	0.95	0.86	0.95
Taxon <i>SD</i>	0.68	1.23	0.95	0.94	0.81	0.92	0.88	1.24
Complement <i>M</i>	-0.44	-0.28	-0.29	-0.31	-0.35	-0.35	-0.31	-0.35
Complement <i>SD</i>	0.68	0.73	0.85	0.83	0.82	0.78	0.85	0.60
Validity ( <i>d</i> )	2.45	1.17	1.25	1.34	1.62	1.58	1.38	1.58
Skew	0.65	1.60	0.71	0.53	0.40	0.72	0.42	1.65

Note: MDD = major depressive disorder.  $N = 1,500$ ; taxon  $n = 400$  MDD+; complement  $n = 1,100$  MDD-. All indicators have been standardized. Symptom numbers refer to *DSM-IV* criteria for MDD. Indicator validity is expressed as Cohen's  $d$ , the mean difference between the taxon and complement standardized using pooled within-groups variances (weighted by  $df$ ).

In all, then, we created five indicator sets: eight indicators from clinical rating data (R8), five depression-specific indicators from rating data (R5), five indicators from questionnaire data (Q5), eight indicators from combined questionnaire and rating data (QR8), and five depression-specific indicators from combined questionnaire and rating data (QR5). Before proceeding, it should be noted that these five sets of indicators were by no means equally well-suited for the planned taxometric analyses. By all criteria, the clinical rating data seemed most likely to provide informative results and the questionnaire data the least, with the combined rating and questionnaire data of intermediate utility. Because one goal of this study was to compare results across different assessment modalities, we performed analyses for all five indicator sets. As described earlier and detailed below, we put several methodological safeguards in place to minimize the risk of misinterpreting ambiguous results.

### Taxometric Procedures

We planned to submit each indicator set to three taxometric procedures: maximum eigenvalue (MAXEIG;

Waller & Meehl, 1998), mean above minus below a cut (MAMBAC; Meehl & Yonce, 1994), and L-Mode (Waller & Meehl, 1998); all these procedures have been described in detail in their original sources as well as in taxometric tutorials (e.g., J. Ruscio, 2007; J. Ruscio & Ruscio, 2004b) and reports of taxometric investigations (e.g., A. M. Ruscio & Ruscio, 2002; A. M. Ruscio, Ruscio, & Keane, 2002). Because of their multivariate nature, we anticipated that MAXEIG and L-Mode would provide stronger discrimination between taxonic and dimensional data than MAMBAC. However, analyses of comparison data suggested that this expectation was only partly met. Specifically, whereas MAXEIG analyses of taxonic and dimensional comparison data often did yield a clearer distinction between these structures than MAMBAC analyses, L-Mode analyses yielded highly ambiguous and poorly discriminating results for each indicator set. Because of the potential for uninformative or misleading results, we chose not to use the L-Mode procedure. We therefore describe our implementation decisions only for the MAXEIG and MAMBAC procedures.

MAXEIG was performed using 50 windows that overlapped 90% with adjacent windows. Because there were 400

MDD+ diagnoses, the putative taxon was substantial both in terms of its absolute size and its relative frequency (26.7%). This means that there was no need to perform the inchworm consistency test (Waller & Meehl, 1998) and that allowing  $n = 254$  cases within each MAXEIG window should be sufficient to detect a taxon of this size (Ruscio et al., 2006). For each curve, one variable served as the “input” indicator and all others served as “output” indicators; this yielded a panel of  $k$  curves for  $k$  indicators. MAXEIG analyses were performed with 10 internal replications to reduce the obfuscating influence of arbitrarily assigning tied-score cases to adjacent windows (Ruscio et al., 2006). Final results for each curve were used to generate an estimate of the taxon base rate using a method that was originally developed by Meehl (1973) for use with the MAXCOV procedure but can be adapted for use with the MAXEIG procedure (Ruscio et al., 2006, Appendix C).

MAMBAC was performed by locating 50 equally spaced cutting scores along the input indicator, beginning and ending 25 cases from either end of its range. For each curve, one variable served as the input indicator and another served as the output indicator. By forming all possible indicator pairs as well as swapping their roles as input and output indicators, this yielded a panel of  $k(k - 1)$  curves for  $k$  indicators. MAMBAC analyses were performed with 10 internal replications to reduce the obfuscating effects of arbitrarily locating cutting scores between tied-score cases. Final results for each curve were used to generate an estimate of the taxon base rate using the formula in Meehl and Yonce (1994).

## Interpretive Aids

Because taxometric curves for actual research data—as opposed to somewhat idealized Monte Carlo data—can be somewhat ambiguous, we took three additional steps to facilitate the accurate interpretation of results for each analysis of each indicator set.

*Taxonic and dimensional comparison data.* Because there are a number of sample-specific factors that can influence the shape of taxometric curves in ways that have yet to be thoroughly explored in Monte Carlo studies, we analyzed comparison data sets with known latent structures. Populations of taxonic and dimensional data were generated that reproduced the unique distributional and correlational characteristics of each indicator set using the programs provided by J. Ruscio and Kaczetow (2008). Analyses of multiple random samples drawn from these populations provide a customized benchmark for interpretation (J. Ruscio et al.,

2007). Because the generation of taxonic comparison data requires a (fallible) criterion variable (Ruscio, Ruscio, et al., 2007), we supplied current MDD diagnoses for this purpose. Thus, taxonic comparison data were generated such that the indicator distributions and correlations within the 400 MDD+ cases and the 1,100 MDD– cases were reproduced. Generating dimensional comparison data requires no criterion variable. At every stage of the interpretive process, results for the research data were accompanied by the results for taxonic and dimensional comparison data. This facilitated not only the subjective interpretation of taxometric curves but also the interpretation of taxon base rate estimates—which can be biased and therefore misleading in a number of ways when interpreted outside the context of parallel estimates for taxonic and dimensional comparison data (J. Ruscio, 2007, Study 1).

*Independent raters.* In addition to subjectively interpreting the taxometric curves ourselves, we enlisted the assistance of 27 minimally trained, independent raters who were blind to the hypotheses of the investigators. These raters consisted of 23 undergraduate students in a research methods and statistics course who completed the curve-judging task in exchange for extra credit after a lab session had ended, plus 4 psychology faculty members who completed the task as time allowed and returned the materials to the experimenter. None of these individuals possessed any special expertise in taxometrics; few were familiar with the method at all. Thus, their minimal knowledge of taxometric curves came only from our rating task instructions, which are reproduced in the Appendix. These instructions described the fundamental research question that taxometric analysis addresses, the way that comparison data can help interpret the obtained results, and the two ratings to be made for each panel of three taxometric curves. Raters were asked first to indicate, using a 5-point scale, how different the results for the taxonic and dimensional comparison data appeared. This provided a measure of the perceived informativeness of the results. Raters were then asked to make a forced-choice interpretive judgment of whether the results for the research data looked more like those for the taxonic or dimensional comparison data.

Before beginning the judgment task, raters were encouraged to peruse all the graphs to get a sense for the full range of results to be assessed and then to ask clarification questions; copies of these graphs are available on request. Once the instructions were

understood and all questions had been answered, it took no more than 10 min for each rater to judge the MAXEIG and MAMBAC curves for the five indicator sets. All raters worked independently and at their own pace.

*Comparison curve fit index (CCFI).* As a final interpretive aid, we calculated a quantitative curve-fit index developed for use in taxometric analyses (Ruscio, Ruscio, et al., 2007). This requires results for comparison data, and it quantifies the relative fit of taxonic and dimensional structural models. The first step is to calculate the root mean square residual of the  $y$  values on the averaged curves of the research data and the comparison data, once to evaluate the fit of the taxonic comparison data and once to evaluate the fit of the dimensional comparison data:

$$Fit_{RMSR} = \sqrt{\frac{\sum (y_{res.data} - y_{comp.data})^2}{N}}$$

where  $y_{res.data}$  refers to a data point on the curve for the research data,  $y_{comp.data}$  refers to the corresponding data point on the curve for simulated taxonic or dimensional comparison data, and  $N$  refers to the number of points on each curve. Lower values of  $Fit_{RMSR}$  reflect better fit, with perfect fit represented by a value of 0. The second step is to combine the two  $Fit_{RMSR}$  values ( $Fit_{RMSR-tax}$  and  $Fit_{RMSR-dim}$ ) into a single fit index:

$$CCFI = \frac{Fit_{RMSR-dim}}{Fit_{RMSR-dim} + Fit_{RMSR-tax}}$$

CCFI values can range from 0 to 1, with lower values indicative of dimensional structure and higher values indicative of taxonic structure. When  $Fit_{RMSR-dim} = Fit_{RMSR-tax}$ , which corresponds to equivalent fit for both structures,  $CCFI = .50$ .

A number of studies suggest that the CCFI distinguishes taxonic and dimensional data with an impressive degree of validity (J. Ruscio, 2007; J. Ruscio & Marcus, 2007; Ruscio, Ruscio, et al., 2007).

## Results

### Analyses of Rating Indicators

Both R8 and R5 yielded peaked MAXEIG curves that appeared considerably more similar to those of the taxonic data than the dimensional comparison

data (see Figure 1). Here, we present only the averaged curves that are accompanied by results for comparison data and that were provided to our independent raters; panels of individual curves are available on request. Our 27 raters judged these to be the most informative of all analyses, and they overwhelmingly selected taxonic structure (96% agreement for R8, 100% agreement for R5; see Table 4 for all curve-judging results). CCFI values of .568 for R8 and .655 for R5 also favored a taxonic interpretation of these MAXEIG results (see Table 4). Estimates of the taxon base rate were consistent within and between these analyses (for R8,  $M = 0.28$ ,  $SD = 0.04$ ; for R5,  $M = 0.24$ ,  $SD = 0.05$ ), and estimates for the averaged curves were more similar to those for the taxonic than the dimensional comparison data (see Table 5).

The MAMBAC analyses of the rating data provided results that also were more consistent with the taxonic structure than the dimensional structure. Although the curves were not neatly peaked (see Figure 1), they nonetheless appeared somewhat more similar to those of the taxonic data than the dimensional comparison data. The difference appears slight, evidenced only by a small alteration in shape and height toward the far right end of the curves, but our impressions in this regard were supported by the independent ratings and even more so by the CCFI. Raters found the MAMBAC results more difficult to interpret than the MAXEIG results (as reflected in the mean ratings of the difference between results across structures), yet the results for the rating data were judged to be taxonic by a strong majority (for R8, 100% agreement; for R5, 85% agreement for all ratings, 92% agreement for only the ratings of those who judged the distinction to be at least moderately clear). CCFI values of .799 for R8 and .729 for R5 were strongly supportive of a taxonic interpretation. Taxon base rate estimates were consistent with those for MAXEIG, and also more similar to those for the taxonic than the dimensional comparison data.

### Analyses of Questionnaire Indicators

In many ways, the results for the questionnaire data were ambiguous. Neither MAXEIG nor MAMBAC curves appeared to be interpretable (see Figure 1), a conclusion reinforced by the independent raters. Ratings of the difference between results across structures were among the lowest of all analyses, and whereas MAXEIG results were interpreted as dimensional by

**Table 4**  
**Comparison Curve Fit Index (CCFI) and 27 Raters' Judgments of All Taxometric Curves**

Indicator Set <sup>a</sup>	Procedure	CCFI	All Judgments		Judgments $\geq 3$	
			<i>M</i> ( <i>SD</i> ) <sup>b</sup>	Percentage Taxonic <sup>c</sup>	<i>N</i> <sup>d</sup>	Percentage Taxonic <sup>c</sup>
R8	MAXEIG	.568	4.04 (0.90)	96	25	96
R8	MAMBAC	.799	3.04 (0.94)	100	17	100
R5	MAXEIG	.655	4.22 (0.80)	100	26	100
R5	MAMBAC	.729	2.44 (0.93)	85	12	92
Q5	MAXEIG	.393	2.07 (0.92)	33	6	33
Q5	MAMBAC	.406	2.56 (1.09)	81	14	86
QR8	MAXEIG	.435	3.26 (0.86)	59	22	41
QR8	MAMBAC	.816	2.89 (1.05)	93	17	100
QR5	MAXEIG	.590	3.22 (0.93)	56	21	57
QR5	MAMBAC	.831	2.70 (1.17)	100	14	100

Note: MAXEIG = maximum eigenvalue; MAMBAC = mean above minus below a cut. CCFI values range from .00 (strongest support for dimensional structure) to 1.00 (strongest support for taxonic structure), with .50 representing an ambiguous result.

a. R = clinical rating data; Q = questionnaire data; QR = combined questionnaire and rating data; the abbreviation is followed by the number of indicators.

b. Judgments of how different the results for the taxonic and dimensional comparison data appeared, rated on a scale from 1 (*not at all different*) to 5 (*very different*).

c. Percentage of raters who judged the results for the research data to be more typical of those for taxonic than dimensional comparison data; *M* across all 10 analyses = 80.3% for all judgments (Column 5), 82.8% for judgments  $\geq 3$  (Column 7).

d. Number of raters (out of 27) who judged the difference between the results for taxonic and dimensional comparison data to be at least a 3 on the 1 to 5 scale described above.

**Table 5**  
**Taxon Base Rate Estimates for the Averaged MAXEIG and MAMBAC Curves for Each Indicator Set**

Indicator Set <sup>a</sup>	Procedure	Research Data	Comparison Data	
			Taxonic	Dimensional
R8	MAXEIG	.30	<b>.27</b>	.22
R8	MAMBAC	.21	<b>.22</b>	.20
R5	MAXEIG	.26	<b>.24</b>	.17
R5	MAMBAC	.23	<b>.19</b>	.17
Q5	MAXEIG	.24	<b>.24</b>	.26
Q5	MAMBAC	.41	.34	<b>.36</b>
QR8	MAXEIG	.25	<b>.28</b>	.19
QR8	MAMBAC	.32	<b>.31</b>	.29
QR5	MAXEIG	.24	<b>.24</b>	.18
QR5	MAMBAC	.30	<b>.29</b>	.29
Mean		.28	<b>.26</b>	.23

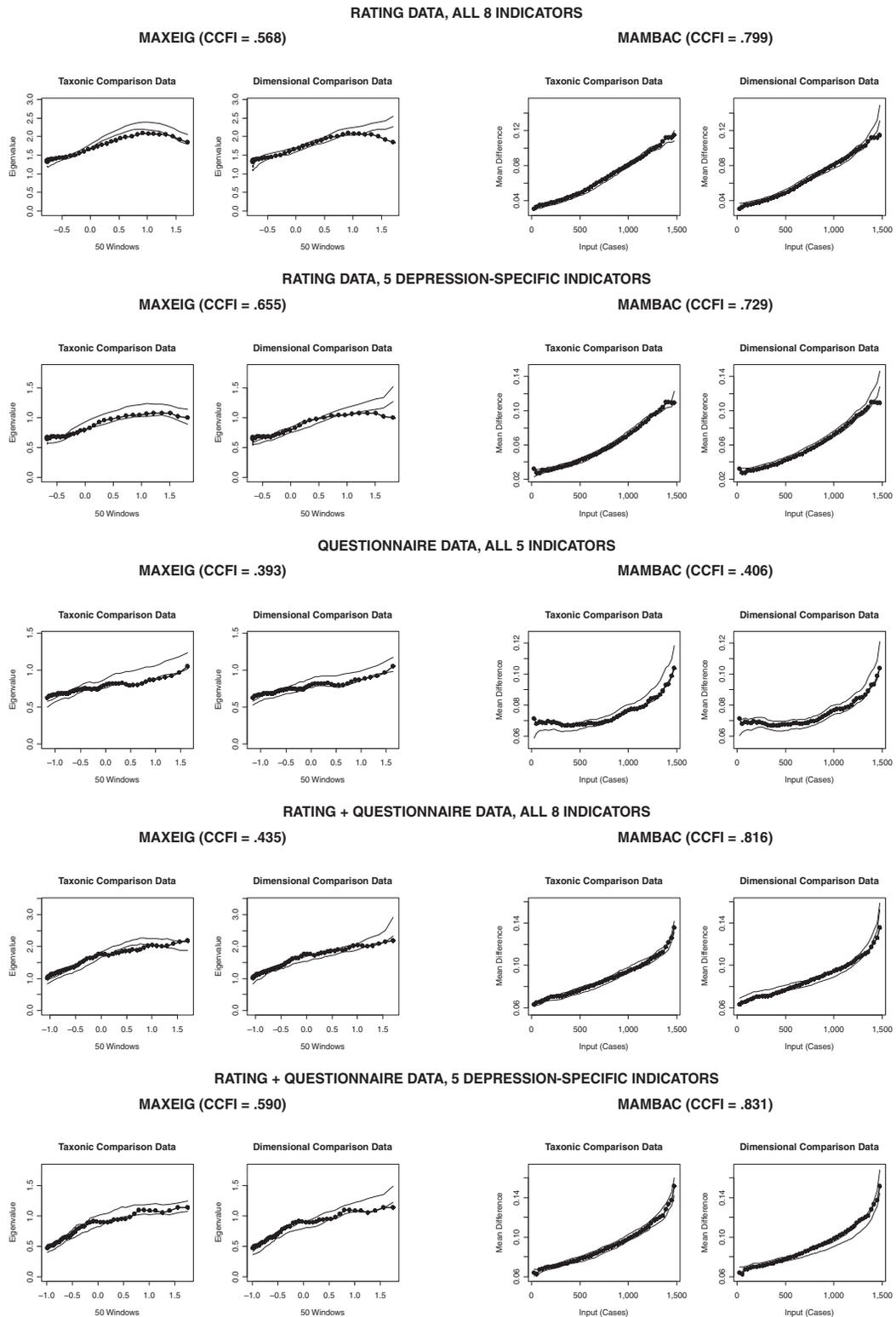
Note: MAXEIG = maximum eigenvalue; MAMBAC = mean above minus below a cut. Taxon base rate estimates calculated using the averaged curve for each taxometric procedure and indicator set, along with the mean of these estimates calculated across all samples of taxonic and dimensional comparison data. The bottom row is calculated as the mean across all 10 indicator set and taxometric procedure combinations. Within each row, the mean for the type of comparison data closer to the corresponding mean for the research data appears in boldface (this was assessed prior to rounding values for the table).

a. R = clinical rating data; Q = questionnaire data; QR = combined questionnaire and rating data; the abbreviation is followed by the number of indicators.

a majority of raters, MAMBAC results were interpreted as taxonic by a larger majority. The CCFI values of .393 (for MAXEIG) and .406 (for MAMBAC) weakly

suggest dimensional structure. Not surprisingly, although MAXEIG estimates of the taxon base rate were consistent with previous analyses, MAMBAC

**Figure 1**  
**MAXEIG (Maximum Eigenvalue) and MAMBAC (Mean Above Minus Below a Cut) Analyses of Each Indicator Set**



Note: Each graph shows the averaged curve for the research data (data points connected by line segments) superimposed on a band representing the averaged curves for comparison data sets. At each data point, this band is centered on the  $M$  for the comparison data and extends to  $\pm 1 SD$ .

estimates were considerably higher than those obtained in analyses of any indicator set. In sum, this pattern of results suggested to us that no structural inference should be drawn from analyses of the questionnaire data.<sup>3</sup>

### Analyses of Combined Rating and Questionnaire Indicators

As might be expected from the previous results, blending the rating and questionnaire data yielded taxometric results whose informational value was intermediate to either assessment modality alone. The MAXEIG analysis yielded ambiguous curves (see Figure 1); particularly toward the right end of each curve—the area critical for interpretation when the base rate of the putative taxon is less than .50—the research data neither evidenced a decline like that for the taxonic comparison data nor a continuous rise like that for the dimensional comparison data. Although the raters judged the MAXEIG results to be reasonably informative, their interpretations of these results were highly ambiguous: 59% and 56% of all ratings were taxonic for QR8 and QR5, respectively, and when only the ratings of those who judged the results to be at least moderately informative were examined, these values further declined to 41% and 57%, respectively. CCFI values were equivocal for QR8 (CCFI = .435) and taxonic for QR5 (CCFI = .590). In contrast, taxon base rate estimates were similar to those from previous analyses and were again more consistent with those for the taxonic than the dimensional comparison data. Thus, the only conclusion that could be drawn from these MAXEIG results was that no consensus emerged.

MAMBAC analyses, in contrast, yielded more interpretable results. As was the case for the rating indicators, MAMBAC analyses of the combined rating and questionnaire data yielded rising curves that appeared somewhat more similar to those for the taxonic than the dimensional comparison data. Raters overwhelmingly selected taxonic structure, with 93% agreement for QR8 and 100% agreement for QR5 (both values were 100% among raters who judged the distinction to be at least moderately clear). Likewise, CCFI values of .816 for QR8 and .831 for QR5 were the strongest in this study and supportive of a taxonic interpretation. Although taxon base rate estimates were barely distinguishable across comparison data sets, nonetheless they were consistent with those from other analyses and slightly closer to the values for taxonic than dimensional comparison data.

### Symptom and BDI Differences by Group Assignment and Gender

In a final series of analyses, we examined differences in symptom ratings and BDI scores across cases assigned to the taxon and complement groups. To classify cases using Bayes's theorem with MAXEIG results, we chose the indicator set that provided the clearest evidence of taxonic structure in MAXEIG analyses: the five depression-specific ratings indicators. This assigned 361 cases to the taxon, and the classification agreed with *DSM-IV* diagnosis for 1,325 cases (293 MDD+ and 1,032 MDD–),  $\kappa = .69$ . Using this Bayesian classification of cases, we tested for group differences, along with gender differences and Group  $\times$  Gender interactions, in between-subjects factorial ANOVAs. Taxon versus complement differences were statistically significant (all  $p < .001$ ) and large (all  $d \geq 1.43$ ) for each symptom and for BDI scores. Effect size was larger for most symptoms than for BDI scores ( $d = 1.50$ ), with the largest differences observed for depressed mood ( $d = 1.96$ ) and loss of interest ( $d = 1.98$ ). Gender differences were small (all  $|d| \leq .11$ ), with only two statistically significant differences (women had higher BDI scores,  $d = -.09$ , and ratings of weight change,  $d = -.10$ ) and one marginally significant difference (men had higher ratings for loss of interest,  $d = .10$ ). Table 6 shows the results for all tests of group and gender main effects; there were no statistically significant interaction effects.

### Discussion

Taken as a whole, the present results suggest that in outpatients with mood and anxiety disorders, a taxonic structural model (which allows for dimensional variation within categories) better represents the *DSM-IV* construct of MDD than a purely dimensional model. At least as noteworthy is that results were clearer when depressive symptoms were assessed using clinical ratings rather than self-report questionnaires. In 10 series of analyses using multiple indicator sets yielded by these assessment methods (clinical ratings, questionnaires, or a combination of both) and submitted to multiple taxometric procedures (MAXEIG, MAMBAC), independent raters judged the results to be taxonic 80% of the time, on average. This rate increased to 83% for analyses in which the raters believed the results for taxonic and dimensional comparison data could be distinguished at least moderately well (Table 4). In

**Table 6**  
**Differences in Depressive Symptom Ratings and BDI Scores**  
**by Group Assignment and Gender**

Symptom	Group Differences			Gender Differences		
	<i>d</i>	Taxon, <i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	<i>d</i>	Taxon, <i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )
Depressed mood	1.96*	4.50 (1.64)	1.20 (1.70)	0.09	2.12 (2.20)	1.92 (2.19)
Loss of interest	1.98*	4.14 (1.94)	0.84 (1.56)	0.11**	1.78 (2.24)	1.55 (2.14)
Weight change	1.76*	2.90 (2.07)	0.40 (1.14)	-0.10*	0.90 (1.73)	1.07 (1.81)
Sleep disturbance	1.44*	3.97 (2.01)	1.24 (1.85)	0.01	1.91 (2.22)	1.89 (2.23)
Psychomotor disturbance	1.86*	3.19 (1.98)	0.55 (1.19)	0.04	1.24 (1.82)	1.16 (1.81)
Fatigue	1.43*	4.34 (1.86)	1.56 (1.97)	-0.04	2.17 (2.27)	2.26 (2.29)
Worthlessness/guilt	1.60*	4.12 (2.05)	1.21 (1.75)	-0.01	1.90 (2.23)	1.92 (2.19)
Impaired concentration	1.49*	4.03 (1.99)	1.27 (1.81)	.05	2.01 (2.20)	1.89 (2.20)
Suicidality	1.51*	1.74 (1.84)	0.15 (0.62)	0.06	0.58 (1.30)	0.50 (1.22)
BDI total score	1.50*	23.92 (9.23)	11.71 (7.74)	-0.09*	14.13 (9.44)	14.98 (9.79)

Note: BDI = Beck Depression Inventory. Cases were assigned to groups using Bayes's theorem with parameter estimates from MAXEIG analyses; taxon  $n = 361$ ; complement  $n = 1,139$ . There were 583 men and 917 women. Cohen's  $d$  was calculated standardized using pooled within-groups variances (weighted by  $df$ ), and statistical significance is based on  $F$  values for main effects tested in a factorial ANOVA. There were no statistically significant Group  $\times$  Gender interaction effects.

\* $p < .05$ . \*\* $p < .10$ .

most instances, the objective CCFI values supported a taxonic interpretation.

At the same time, results differed across indicator sets. Results were clearest for the clinical rating indicators drawn from the semistructured ADIS-IV-L (R8, R5), where MAXEIG and MAMBAC analyses provided considerably stronger support for taxonic than dimensional structure. This judgment was upheld by the independent raters, who evidenced good agreement that the distinction could be made at least moderately well and overwhelmingly (97%) chose in favor of taxonic structure.

The questionnaire indicators drawn primarily from the BDI and DASS (Q5) appeared marginally adequate, at best, for taxometric analysis (e.g., high within-group correlations, modest separation between groups; poor discriminability of results for taxonic and dimensional comparison data). Accordingly, 63% of the time the independent raters indicated that these results could not be distinguished well. In instances where the raters felt this distinction was reasonably clear, the percentage of judgments in favor of taxonic structure varied considerably across taxometric procedures (33% for MAXEIG, 86% for MAMBAC). In addition to demonstrating that the interpretability of taxometric results can depend on how one chooses to assess the construct, these results provide some assurance that our procedural safeguards against the misinterpretation of ambiguous results were helpful.

Analyses of combined rating and questionnaire data (QR5, QR8) yielded results of intermediate clarity. For instance, 68% of the independent ratings suggested that the results could be distinguished at least moderately well across structures. On average, 75% of the ratings in these cases were in support of taxonic structure, though discrepancies were observed between the MAXEIG and MAMBAC results.

The clinical rating data required relatively little work to construct useful indicators that better represented the broad array of cognitive, somatic, and affective features of *DSM-IV* MDD. This suggests that clinical rating data may be more appropriate for studying some psychopathological phenomena than self-report measures. For instance, questionnaire data may be insufficiently sensitive to the high trait levels on MDD criteria required to adequately represent the construct. Moreover, it may be more challenging to construct indicator sets that are appropriate for taxometric analysis in clinical than in nonclinical samples, perhaps in part because of higher levels of general distress and negative affect found in patient samples relative to college student or community samples. The fact that questionnaires were found to be less useful than clinical ratings could be viewed as consistent with prior evidence that the former assessment modality is more prone to bias arising from nonspecific distress (e.g., Widiger, Verheul, & van den Brink, 1999). Whereas J. Ruscio and Ruscio (2002) discussed the importance of understanding

latent structure for assessing a construct most efficiently, the present study suggests that obtaining interpretable taxometric results can depend on how one chooses to assess the construct.

Because the final allocation of symptoms to indicators differed across rating and questionnaire data, one might wonder whether we conflated the issue of method of assessment with representation of the depressive construct. To examine this possibility as directly as our data allowed, we constructed an indicator set using the rating data by assigning symptom ratings to indicators as in the questionnaire data. Analyses of this set of five indicators (Symptoms 1/2/7/9, 3, 5, 6, and 8) yielded CCFI values of .779 for MAMBAC and .593 for MAXEIG, which are at least as supportive of taxonic structure as other analyses.

Despite the apparently superior utility of clinical ratings in the present study, criticisms have been levied against the use of rating data in taxometric analyses. For instance, Beauchaine and Waters (2003) contended that rating data (obtained by means of either self-report or expert opinion) are more prone to yield pseudotaxonic results because of measurement artifacts (e.g., systematic response tendencies) and a preexisting tendency of raters to hold categorical beliefs about the nature of various psychological constructs (i.e., natural inclination to classify others or oneself in terms of the presence/absence of a given characteristic, e.g., shy vs. outgoing). On the other hand, McGrath, Neubauer, Meyer, and Tung (in press) performed taxometric analyses of rating data obtained under taxonic or dimensional instruction sets and concluded that under normal conditions for the use of rating scales, instructional set does not influence structural results. At present, it remains uncertain when and how much rater expectations may influence taxometric results. With regard to the present data, precautions had been taken to minimize the potential influence of *DSM-IV* categorical diagnostic guidelines on the clinical ratings.

In an effort to guard against artifacts in the rating procedure that could bias the results in favor of taxonicity, interviewers were trained to assign the dimensional (0-8) severity ratings of MDD symptoms from the perspective of a patient, without considering *DSM-IV* diagnostic guidelines (e.g., clinical threshold, differential diagnosis, diagnostic hierarchy) other than the MDD temporal criterion (i.e., severity rating in reference to the past 2 weeks). Of course, it cannot be determined how well interviewers adhered to this directive when making their ratings, and it is possible that some ratings

were artificially inflated based on the expectation that a patient would be assigned a current MDD diagnosis. A recent study from the Center for Anxiety and Related Disorders at Boston University (Kollman, Brown, Liverant, & Hoffman, 2006) used a similar set of clinical ratings (0-8) in a taxometric investigation of *DSM-IV* social anxiety disorder. Encouragingly, the results provided unequivocal evidence for a dimensional latent structure of social anxiety disorder, an outcome that allays concerns that these rating procedures have an inherent method bias toward taxonicity.

A related concern is that the individuals who provided ratings of the taxometric curves may have been biased in favor of perceiving evidence of taxonic structure. Though our instructions were not leading, raters' prior beliefs about the structure of clinical depression may have influenced judgments. However, if any bias existed, we believe that the raters may have been predisposed toward perceiving evidence of dimensional structure. Each of the student raters had taken an introductory psychology course, usually the semester prior to taking the research methods and statistics course during which they made ratings for the present study. The assigned text for that course (Weiten, 2001) explicitly stated that mental disorders are not categorical phenomena and that individuals differ along a continuum ranging from normality to abnormality; a diagram in the margin illustrated such a severity continuum. The faculty raters had chosen this text and assigned it in their sections of the introductory psychology course, which suggests that they are at least familiar with this perspective and aware that many psychologists share it. It is our impression that the prevailing belief among psychologists is that mental disorders are dimensional, not taxonic, and we have no reason to suspect that our raters tended to hold the opposite view.

The present findings provide suggestive evidence that *DSM-IV* MDD may be taxonic, and the methods and results are most consistent with those of J. Ruscio, Zimmerman, et al. (2007). These findings are at odds with those from prior taxometric investigations of unipolar depression that have supported a dimensional structural model. The methodology of the current study (e.g., sampling, indicator sets) differed from previous studies in ways that may account for this, however. Most of these studies relied on indicators drawn from self-report questionnaires (e.g., BDI) that were not specific to the *DSM-IV* construct of MDD. Many studies also assessed depressive symptoms in student or community samples with very low base

rates of diagnosable MDD. The present study included heteromethod (semistructured clinical interview ratings, questionnaire) indicators that were constructed to correspond closely to the *DSM-IV* definition of MDD, a large outpatient sample carefully assessed for *DSM-IV* mood and anxiety disorders, and a sizeable base rate of MDD (.27). At the same time, although the current study used a sample that is appropriate to examine the latent structure of MDD (e.g., large clinical sample with a sizeable base rate of MDD, reliably determined by structured interview), it was characterized by a predominance of anxiety disorders, with MDD more likely to be an additional diagnosis than the principal diagnosis. In addition, the sample was dominated by mild-to-moderate MDD presentations; given the outpatient setting, severe or suicidal cases of MDD were uncommon. Thus, it would be important to extend this work to other (e.g., inpatient) clinical settings where severe MDD is more prevalent.

Future studies should also employ suitable measures (e.g., indicators that directly assess the *DSM-IV* disorder construct under study) and procedural safeguards to ensure that structural inferences are drawn in a cautious manner using data that are appropriate for taxometric analysis. A careful and deliberate approach to the selection and construction of indicators can help refine the indicator sets submitted to taxometric analysis, leading to the combination or exclusion of indicators that possess too little validity to distinguish putative groups or that covary too highly within these groups. Even with concerted effort, some data may not yield any indicator sets that prove to be adequate for taxometric analysis. Moreover, software is available to allow the researcher to generate and analyze taxonic and dimensional comparison data that reproduce the distributional and correlational properties of the research data (Ruscio, Ruscio, et al., 2007). This approach provides an interpretative aid that, in the present study, revealed the ambiguity of the results for questionnaire indicators.

The use of independent raters can supplement the objectively calculated CCFI. In this study, raters' judgments also suggested that a taxonic model provided better fit than the dimensional model. In a number of studies, Walters and his colleagues have tested their conclusions against those of independent raters (e.g., Walters, Diamond, Magaletta, Geyer, & Duncan, 2007; Walters, Duncan, & Mitchell-Perez, 2007; Walters et al., 2007). Because it requires the use of raters without a bias toward categorical or continuous conclusions, this safeguard may be best

implemented with raters who are not only blind to the investigators' hypotheses but also naive to the taxometric method. Individuals with no preconceived notions may be most objective when interpreting empirical results against those for comparison data.

Sound knowledge of the structure of MDD may profoundly impact the manner in which this disorder is classified in future nosological systems such as *DSM-V*. This is one of the initial studies to entail a large clinical sample with a sizeable base rate of MDD and clinical ratings of *DSM-IV* MDD criteria. Initial evidence suggests that MDD may be characterized better by taxonic than dimensional structure, representing a qualitatively distinct phenomenon from normal sadness and dysphoria. Nonetheless, this study also highlighted some of the methodological challenges associated with applying the taxometric method to clinical samples and *DSM-IV* criteria sets. We hesitate to draw strong conclusions from this single study, especially when other studies have supported different conclusions about the structure of MDD, but it is reassuring to note that two recent studies (J. Ruscio, Zimmerman, et al., 2007; Solomon et al., 2006) that share some of the features of the present investigation (e.g., inclusion of clinical interview data and, in the former, analyses of comparison data) also produced taxonic results. We recommend that future patient-based studies adopt some of the methodological recommendations outlined in this article to further our understanding of the latent structure of MDD and other *DSM-IV* disorders.

Another useful direction for future research involving an MDD taxon is the development of a user-friendly assessment tool. Using a classification of cases provided by taxometric analyses, one can examine the sensitivity and specificity of candidate items to determine which most efficiently distinguish between MDD+ and MDD- individuals. One or more scales can be constructed from item sets, thresholds then studied using receiver operating characteristic (ROC) curve analyses, and the scale validity can be tested using external criteria. Such work extends beyond the scope of the present study but would provide a helpful foundation for subsequent exploration of an MDD taxon's relationship with other constructs in basic and applied psychopathology research.

There are other depressive constructs that warrant study and that are relevant to understanding the structure of the broader depression construct. These include known vulnerability factors for depression, characteristics of depression that cut across episodes over an individual's life (e.g., chronic/recurrent depression),

proposed alternative manifestations of depression symptoms beyond MDD (e.g., dysthymic disorder, brief recurrent depression, depressive personality disorder), and putative depressive subtypes (e.g., atypical depression, psychotic depression). What is called *depression* is heterogeneous, complex, and multiply defined. Because there is particular interest in understanding and treating current major depressive episodes, it makes sense that taxometric studies, including ours, have focused mainly on this level of analysis. Nonetheless, a comprehensive picture of the structure of depression—and efforts to tie this back to etiology and mechanisms—is likely to require parallel, complementary investigations focusing on different levels of analysis.

We examined differences in symptom ratings across taxon and complement members. J. Ruscio, Zimmerman, et al. (2007) and Zimmerman, Chelminski, McGlinchey, and Young (2006) report the sensitivity and specificity of *DSM-IV* symptoms, and Andrews, Slade, Sunderland, and Anderson (2007) also suggest ways that diagnostic criteria might be simplified to increase utility in future editions of the *DSM*. The present findings add to this growing literature on ways to improve the diagnostic criteria for MDD or develop new measures of the construct, and future research should use additional criteria of convergent and discriminant validity to build on this foundation.

## Appendix

### Rating Task Instructions

---

Individuals can differ from one another in categorical or continuous ways. For example, men and women belong to distinct categories based on their biological sexes, whereas tall and short people vary along a continuum of heights. In a study of clinical depression, I performed statistical procedures to determine whether depression is best characterized as a categorical (i.e., depressed vs. nondepressed) or a continuous (i.e., severity of depression) phenomenon.

These statistical procedures produce graphs, not *p* values, so the proper interpretation of the curves requires some subjective judgment. To help interpret the results for each analysis, I created artificial, or “simulated,” data sets whose structure (categorical or continuous) is known. These simulated data sets were analyzed using the same statistical procedures, so their results provide a benchmark for comparison that can be helpful when interpreting the results for the actual data, whose structure is unknown.

On each of the following pages (one page for each data set), you will be asked to judge graphs (one series for each

statistical procedure). Each series of graphs is arranged horizontally, and it contains the results of one particular kind of statistical procedure applied to both the actual data and the simulated data sets.

- The graph on the left contains a single curve that was generated by analyzing one set of actual research data.
- The middle graph contains 10 curves (plus, in bold, their average) that were generated by analyzing 10 data sets simulated to be “taxonic.” This means that individuals in these simulated data sets belong to one or the other of two categories (as in the possibility of depressed vs. nondepressed individuals).
- The graph on the right contains 10 curves (plus, in bold, their average) that were generated by analyzing 10 data sets simulated to be “dimensional.” This means that individuals in these simulated data sets lie somewhere along a continuum of possible scores (as in the possibility that individuals differ in terms of depressive severity).

For each series of graphs, you will be asked to make two judgments:

1. How different are the results for the simulated taxonic and dimensional data? To answer this question, compare the set of curves in the middle graph to the set of curves in the graph on the right. If these two sets of curves are not different, you would circle the lowest value on the scale (1). If these two sets of curves are very different, you would circle the highest value on the scale (5). Alternatively, you could circle an intermediate value (2 through 4) to represent a degree of difference somewhere between these extremes. *Please circle one (and only one) number for this question.*
2. Do the results for the research data look more like the results for the simulated *taxonic* or *dimensional* data? To answer this question, you need to compare the curve in the graph on the left to the sets of curves in the other graphs to see where the better match is achieved. If you think the curve in the graph on the left looks more typical of the curves in the middle graph, circle the italicized word “taxonic” in the question. If you think the curve in the graph on the left looks more typical of the curves in the graph on the right, circle the italicized word “dimensional” in the question. *Please circle one (and only one) word for this question.*

Please take a moment to flip through all the pages to get a sense for what the curves tend to look like, and ask me any questions that you have before you begin to make your judgments.

---

## Notes

1. This published reliability estimate pertained to the unidimensional composite of the MDD ratings. Interrater  $r$ s for individual symptoms ranged from .33 to .72 ( $M = 0.59$ ). MDD symptom #5 (psychomotor agitation/retardation) was the only rating with an interrater  $r$  less than .55.

2. Items were combined to represent each of the nine *DSM-IV* diagnostic features of MDD as follows: #1 (BDI Item 1; DASS Items 13 and 16); #2 (BDI Item 4; DASS Items 3, 16, 24, and 31); #3 (BDI Items 18 and 19); #4 (BDI Item 16); #5 (DASS Items 8, 33, and 39); #6 (BDI Item 17; DASS Item 5); #7 (BDI Items 3, 5, 7, and 8; DASS Items 17 and 34); #8 (BDI Item 13; PI item); #9 (BDI Items 2 and 9; DASS Items 10, 21, 37, and 38).

3. A follow-up series of taxometric analyses was performed using a Q8 indicator set constructed in parallel with the R8 and QR8 indicator sets. The MAXEIG and MAMBAC results were ambiguous. Though we do not have independent ratings from our judges for these supplementary analyses, the CCFI values of .481 and .644 (for MAXEIG and MAMBAC, respectively) underscore the ambiguity of these results. These graphs are available on request.

## References

- Ambrosini, P., Bennett, D. S., Cleland, C. M., & Haslam, N. (2002). Taxonicity of adolescent melancholia: A categorical or dimensional construct? *Journal of Psychiatric Research, 36*, 247-256.
- Andrews, G., Slade, T., Sunderland, M., & Anderson, T. (2007). Issues for *DSM-V*: Simplifying *DSM-IV* to enhance utility: The case of major depressive disorder. *American Journal of Psychiatry, 164*, 1784-1785.
- Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment, 10*, 176-181.
- Beach, S. R. H., & Amir, N. (2003). Is depression taxonic, dimensional, or both? *Journal of Abnormal Psychology, 112*, 228-236.
- Beauchaine, T. P., & Beauchaine, R. J. (2002). A comparison of maximum covariance and k-means cluster analysis in classifying cases into known taxon groups. *Psychological Methods, 7*, 245-261.
- Beauchaine, T. P., & Waters, E. (2003). Pseudotaxonicity in MAMBAC and MAXCOV analyses of rating-scale data: Turning continua into classes by manipulating observer's expectations. *Psychological Methods, 8*, 3-15.
- Beck, A. T., & Steer, R. A. (1993). *Beck Depression Inventory manual*. Orlando, FL: Harcourt Brace.
- Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8*, 77-100.
- Brown, T. A. (1996). Validity of the *DSM-III-R* and *DSM-IV* classification systems for anxiety disorders. In R. M. Rapee (Ed.), *Current controversies in the anxiety disorders* (pp. 21-45). New York: Guilford Press.
- Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour Research and Therapy, 35*, 79-89.
- Brown, T. A., Di Nardo, P. A., Lehman, C. L., & Campbell, L. A. (2001). Reliability of *DSM-IV* anxiety and mood disorders: Implications for the classification of emotional disorders. *Journal of Abnormal Psychology, 110*, 49-58.
- Coyne, J. C. (1994). Self-reported distress: Analog or ersatz depression? *Psychological Bulletin, 116*, 29-45.
- Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology, 42*, 111-131.
- Deardorff, W. W., & Funabiki, D. (1985). A diagnostic caution in screening for depressed college students. *Cognitive Therapy and Research, 9*, 277-284.
- Di Nardo, P. A., Brown, T. A., & Barlow, D. H. (1994). *Anxiety Disorders Interview Schedule for DSM-IV: Lifetime Version (ADIS-IV-L)*. San Antonio, TX: Psychological Corporation.
- Flett, G. L., Vredenburg, K., & Krames, L. (1997). The continuity of depression in clinical and nonclinical samples. *Psychological Bulletin, 121*, 395-416.
- Gibb, B. E., Alloy, L. B., Abramson, L. Y., Beevers, C. G., & Miller, I. W. (2004). Cognitive vulnerability to depression: A taxometric analysis. *Journal of Abnormal Psychology, 113*, 81-89.
- Gilbert, P. (1992). *Depression: The evolution of powerlessness*. New York: Guilford Press.
- Gilbert, P. (2000). Varieties of submissive behavior as forms of social defense: Their evolution and role in depression. In L. Sloman & P. Gilbert (Eds.), *Subordination and defeat: An evolutionary approach to mood disorders and their therapy* (pp. 3-45). Mahwah, NJ: Erlbaum.
- Grove, W. M., Andreasen, N. C., Young, M., Endicott, J., Keller, M. B., Hirschfeld, R. M. A., et al. (1987). Isolation and characterization of a nuclear depressive syndrome. *Psychological Medicine, 17*, 471-484.
- Hankin, B. L., Fraley, R. C., Lahey, B. B., & Waldman, I. D. (2005). Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. *Journal of Abnormal Psychology, 114*, 96-110.
- Haslam, N., & Beck, A. T. (1994). Subtyping major depression: A taxometric analysis. *Journal of Abnormal Psychology, 103*, 686-692.
- Haslam, N., & Kim, H. C. (2002). Categories and continua: A review of taxometric research. *Genetic, Social, and General Psychology Monographs, 128*, 271-320.
- Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory* (Rev. ed.). Minneapolis: University of Minnesota Press.
- Kollman, D. M., Brown, T. A., Liverant, G. I., & Hofmann, S. G. (2006). A taxometric investigation of the latent structure of *DSM-IV* social anxiety disorder in outpatients with anxiety and mood disorders. *Depression and Anxiety, 23*, 190-199.
- Lips, H., & Ng, M. (1985). Use of the Beck Depression Inventory with three nonclinical samples. *Canadian Journal of Behavioural Science, 18*, 62-74.
- Lovibond, S. H., & Lovibond, P. F. (1995a). *Manual for the Depression Anxiety Stress Scales* (2nd ed.). Sydney, Australia: Psychology Foundation.
- Lovibond, S. H., & Lovibond, P. F. (1995b). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy, 33*, 335-343.

- McGrath, R. E., Neubauer, J., Meyer, G. J., & Tung, K. (in press). Instructional set and the structure of responses to rating scales. *Personality and Individual Differences*.
- Meehl, P. E. (1973). MAXCOV-HITMAX: A taxonomic search method for loose genetic syndromes. In P. E. Meehl (Ed.), *Psychodiagnosis: Selected papers* (pp. 200-224). Minneapolis: University of Minnesota Press.
- Meehl, P. E. (1995). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist*, *50*, 266-274.
- Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports*, *74*, 1059-1274.
- Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports*, *78*, 1091-1227.
- Ruscio, A. M., & Ruscio, J. (2002). The latent structure of analogue depression: Should the BDI be used to classify groups? *Psychological Assessment*, *14*, 135-145.
- Ruscio, A. M., Ruscio, J., & Keane, T. M. (2002). The latent structure of post-traumatic stress disorder: A taxometric investigation of reactions to extreme stress. *Journal of Abnormal Psychology*, *111*, 290-301.
- Ruscio, J. (2007). Taxometric analysis: An empirically-grounded approach to implementing the method. *Criminal Justice and Behavior*, *24*, 1588-1622.
- Ruscio, J., Haslam, N., & Ruscio, A. M. (2006). *Introduction to the taxometric method: A practical guide*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ruscio, J., & Kacetow, W. (2008). Simulating multivariate non-normal data using an iterative algorithm. *Multivariate Behavioral Research*, *43*, 355-381.
- Ruscio, J., & Marcus, D. K. (2007). Detecting small taxa using simulated comparison data: A reanalysis of Beach, Amir, and Bau's (2005) data. *Psychological Assessment*, *19*, 241-246.
- Ruscio, J., & Ruscio, A. M. (2000). Informing the continuity controversy: A taxometric analysis of depression. *Journal of Abnormal Psychology*, *109*, 473-487.
- Ruscio, J., & Ruscio, A. M. (2002). A structure-based approach to psychological assessment: Matching measurement models to latent structure. *Assessment*, *9*, 5-17.
- Ruscio, J., & Ruscio, A. M. (2004a). Clarifying boundary issues in psychopathology: The role of taxometrics in a comprehensive program of structural research. *Journal of Abnormal Psychology*, *113*, 24-38.
- Ruscio, J., & Ruscio, A. M. (2004b). A conceptual and methodological checklist for conducting a taxometric investigation. *Behavior Therapy*, *35*, 403-447.
- Ruscio, J., Ruscio, A. M., & Keane, T. M. (2004). Using taxometric analysis to distinguish a small latent taxon from a latent dimension with positively skewed indicators: The case of involuntary defeat syndrome. *Journal of Abnormal Psychology*, *113*, 145-154.
- Ruscio, J., Ruscio, A. M., & Meron, M. (2007). Applying the bootstrap to taxometric analysis: Generating empirical sampling distributions to help interpret results. *Multivariate Behavioral Research*, *42*, 349-386.
- Ruscio, J., Zimmerman, M., McGlinchey, J. B., Chelminski, I., & Young, D. (2007). Diagnosing major depressive disorder: XI. A taxometric investigation of the categorical-dimensional debate on the structure underlying *DSM-IV* symptoms. *Journal of Nervous and Mental Disease*, *195*, 10-19.
- Sanavio, E. (1988). Obsessions and compulsions: The Padua Inventory. *Behaviour Research and Therapy*, *26*, 169-177.
- Solomon, A., Ruscio, J., Seeley, J. R., & Lewinsohn, P. R. (2006). A taxometric investigation of unipolar depression in a large outpatient sample. *Psychological Medicine*, *36*, 973-986.
- Strong, D. R., Brown, R. A., Kahler, C. W., Lloyd-Richardson, E. E., & Niaura, R. (2004). Depression proneness in treatment-seeking smokers: A taxometric analysis. *Personality and Individual Differences*, *36*, 1155-1170.
- Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua*. Newbury Park, CA: Sage.
- Walters, G. D., Diamond, P. M., Magaletta, P. R., Geyer, M. D., & Duncan, S. A. (2007). Taxometric analysis of the Antisocial Features Scale of the Personality Assessment Inventory in federal prison inmates. *Assessment*, *14*, 351-360.
- Walters, G. D., Duncan, S. A., & Mitchell-Perez, K. (2007). A taxometric investigation of the Psychopathy Checklist-Revised in a heterogeneous sample of male prison inmates. *Assessment*, *14*, 270-278.
- Walters, G. D., Gray, N. S., Jackson, R. L., Sewell, K. W., Rogers, R., Taylor, J., et al. (2007). A taxometric analysis of the Psychopathy Checklist: Screening Version (PCL:SV): Further evidence of dimensionality. *Psychological Assessment*, *19*, 330-339.
- Weiten, W. (2001). *Psychology: Themes and variations* (5th ed.). Pacific Grove, CA: Wadsworth.
- Widiger, T. A., & Frances, A. J. (2002). Towards a dimensional model for the personality disorders. In P. T. Costa & T. A. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (2nd ed., pp. 23-44). Washington, DC: American Psychological Association.
- Widiger, T. A., Verheul, R., & van den Brink, W. (1999). Personality and psychopathology. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 347-366). New York: Guilford Press.
- Zimmerman, M., Chelminski, I., McGlinchey, J. B., & Young, D. (2006). Diagnosing major depressive disorder X: Can the utility of the DSM symptom criteria be improved? *Journal of Nervous and Mental Disease*, *194*, 893-897.