Is Depression Best Viewed as a Continuum or Discrete Category?
A Taxometric Analysis of Childhood and Adolescent Depression in a Population-Based Sample

Benjamin L. Hankin and R. Chris Fraley
University of Illinois at Chicago

Benjamin B. Lahey
University of Chicago

Irwin D. Waldman
Emory University

The authors examined the latent structure of depression in a population-based sample of children and adolescents. Youth’s self-reports and parents’ reports of the youth’s Diagnostic and Statistical Manual of Mental Disorders (4th ed.; DSM–IV; American Psychiatric Association, 1994) major depressive symptoms were assessed via a structured clinical interview. The authors used Meehl’s (1995) taxometric procedures to discern whether youth depression is dimensional or categorical. Taxometric analyses that explicitly took into account the skewness of depressive symptoms suggested that depression is a dimensional, not categorical, construct. The dimensional structure of depression was obtained for all of the DSM–IV major depressive symptoms as well as for different domains of depression (emotional distress symptoms and vegetative, involuntary defeat symptoms), youth and parent reports, and different subsamples (i.e., boys vs. girls and younger vs. older youth).

Since the publication of the Diagnostic and Statistical Manual of Mental Disorders (3rd ed.; DSM–III; American Psychiatric Association, 1980), the manner in which psychopathology is conceptualized, described, assessed, studied, and treated has changed radically. The introduction of the DSM–III, and its subsequent revisions, has revolutionized the classification and study of psychopathology. Many agree that the current DSM system and nomenclature has had a positive impact on the study of psychopathology. Despite these gains, recent writers (e.g., Beutler & Malik, 2002; Widiger & Clark, 2000) have criticized the DSM on numerous fronts, including its lack of a sufficient scientific basis for a psychopathological nomenclature, its potential to stifle innovative research, and its atheoretical grounding. One of the most significant criticisms of the DSM as a scientific organizer of psychopathology is its assumption that people should be classified as having or not having particular psychopathological disorders (Widiger & Clark, 2000).

The “types versus dimensions” issue—the question of whether psychopathology is a matter of degree or kind—is of considerable importance in developing and evaluating a psychopathological nosological system (e.g., DSM). This issue has been particularly significant in the study of depression (Solomon, Haaga, & Arnow, 2001). On the one hand, the current psychiatric diagnostic structure (e.g., DSM–IV; American Psychiatric Association, 1994) portrays depression as a qualitatively distinct disorder. On the other hand, several researchers have argued that depression may best be viewed as a quantitative deviation from “normal” affective experience (e.g., Solomon, Haaga, & Arnow, 2001). Unfortunately, previous research has not been able to delineate the latent structure of depression in an unambiguous manner. A major impediment to resolving this controversy has been that traditional methods (e.g., cluster analysis) used to address this debate are incapable of uncovering the latent structure of a construct (Fraley & Waller, 1998). To answer the types versus dimensions question satisfactorily, structure uncovering methods (i.e., those that reveal a construct’s actual structure) are needed, rather than structure imposing ones (i.e., those that force a structure onto the data; Waller & Meehl, 1998).

In recent years, Meehl and his colleagues (Meehl, 1973, 1992, 1995; Meehl & Yonce, 1996; Waller & Meehl, 1998) have developed a suite of taxometric procedures that are well-suited for addressing the types versus dimensions question empirically. The primary goal of taxometric analysis is to determine whether a latent variable represents a “naturally occurring type” or “taxon” (Meehl, 1992). These structure-uncovering techniques have been applied successfully in various psychological domains (see Haslam & Kim, 2002), including dissociation (Waller, Putnam, & Carlson, 1996), personality disorders (Trull, Widiger, & Guthrie,
Our primary objective in the current article is to evaluate the extent to which major depression among youth, as defined by the DSM–IV, is categorical or dimensional by using Meehl’s (Waller & Meehl, 1998) taxometric procedures. We begin by reviewing extant taxometric investigations of depression using adult and adolescent samples. Next, we discuss some of the limitations of previous work on this topic. To help overcome these limitations, we present a taxometric investigation of youth depression using interview ratings of all DSM–IV-defined depressive symptoms among children and adolescents from a population-based sample. Additionally, we examine youth’s self-reports and parents’ reports of their child’s depressive symptoms for convergence of findings across informant sources.

Previous Taxometric Studies of Depression

A number of published studies have applied Meehl’s taxometric procedures to depression among adults. In an early study, Haslam and Beck (1994) used taxometric procedures in a large sample of depressed clinical outpatients to examine categorical subtypes of depression (endogenous, hopelessness depression, sociotropic, and autonomous depression). Endogenous depression (a biological-based subtype) was found to be categorical (see also Grove et al., 1987), whereas variation with respect to nonendogenous subtypes of depression was dimensional.

In another innovative study, Ruscio and Ruscio (2000) used taxometric methods on two large clinical samples of adults and found evidence that depression was dimensional, not taxonic. Still, A. M. Ruscio and Ruscio (2002) noted several limitations of their study, including the limited generalizability of their samples (i.e., adult clinical outpatients and inpatients) and the exclusive use of self-report questionnaires of depression. Franklin, Strong, and Greene (2002) showed that depression, as assessed by the Minnesota Multiphasic Personality Inventory—II Depression scale, was dimensional in a large sample of psychiatric patients.

More recently, Beach and Amir (2003) applied taxometric techniques to examine the taxonicity of potentially different domains of depression using items selected from the Beck Depression Inventory (BDI) in a sample of college undergraduates. They found that the latent variable of emotional distress, as assessed by predominantly affective BDI items, was dimensional, whereas a different latent variable, as assessed by BDI somatic symptoms (e.g., sleep, weight disturbance), was categorical. They referred to these latter symptoms of depression as an involuntary defeat syndrome (IDS) on the basis of an evolutionary psychological model of depression (Gilbert, 1992, 2000) in which certain depressive symptoms (particularly somatic, vegetative symptoms) are hypothesized to reflect the disruption of adaptive homeostatic processes.

Although these studies have relied on adult samples, at least two studies have examined the taxonicity of depression among adolescents. In a moderately sized sample of psychiatric outpatients, Ambrosini and colleagues (Ambrosini, Bennett, Cleland, & Haslam, 2002) found that symptoms of melancholic depression were taxonic, whereas Whisman and Pinto (1997) found that hopelessness depression was dimensional in a small sample of adolescent inpatients.

In summary, the evidence of taxonicity in depression is somewhat mixed. A few studies suggest that the latent structure of depression is dimensional, particularly when emotional distress symptoms are considered (e.g., Beach & Amir, 2003; Franklin et al., 2002; nonendogenous subtypes in Haslam & Beck, 1994), whereas depression appears taxonic when examining somatic symptoms (e.g., IDS, Beach & Amir, 2003; or endogenous depression, Ambrosini et al., 2002; Grove et al., 1987).

Limitations of and Unanswered Questions Concerning Existing Taxometric Studies

There are several limitations of past research that make it difficult to draw strong inferences about the latent structure of depression. First, all of the samples were drawn from selected populations (i.e., clinical patients or college students). It may be that focusing only on college students introduces systematic sampling bias, such that those individuals who are functioning well enough to attend classes are studied. Alternatively, focusing only on clinical patients may limit variability in the construct under study. For example, clinical outpatient and inpatient participants exhibit more severe symptoms and greater comorbidities than population-based samples (e.g., Newman et al., 1998). As Waller and Meehl (1998) explained, if true categories exist, taxometric techniques will be able to identify them most effectively when there is variability in the latent entity (i.e., when the sample is composed of both people who belong to the latent class and people who do not). A population-based sample would not share these problems and would have the potential to provide a more accurate picture of the latent structure of depression.

Second, existing studies have not explicitly examined the latent structure of DSM–IV’s conceptualization of depression. All of the published taxometric studies of depression have used questionnaires (e.g., BDI, Minnesota Multiphasic Personality Inventory—II) to assess depressive symptoms, but none of these assess all of the DSM–IV major depression symptoms. To evaluate DSM–IV’s conceptualization and classification of major depression as a qualitatively distinct category, one must assess all of DSM–IV defined depressive symptoms.

Third, the problem of skewness has generally not been taken into account in most taxometric investigations. When the indicators of a dimensional construct are positively skewed, the output from commonly used taxometric procedures will tend to suggest erroneously that there is a low-base-rate taxon (e.g., J. Ruscio, Ruscio, & Meron, 2003). This occurs because the skewness of the responses distorts the relationship between the observed scores and the latent variable. Such distortions commonly occur when assessment techniques are designed to assess the high end of a latent variable (e.g., severe or clinical depression) as opposed to the full range of the variable (Fraley, Waller, & Brennan, 2000).

Although some investigators have begun to recognize the significance of this problem for taxometric research on clinical syndromes (Waldman & Lilienfeld, 2001), only a few have begun to take the skewness problem seriously in their taxometric research. A. M. Ruscio and Ruscio (2002) recently presented a method that allows researchers to address the types versus dimensions question.
in the presence of skewed indicators (see also J. Ruscio et al., 2003). Their method involves generating simulated data under taxonic and dimensional conditions but scaling each set of variables to conform to the same distributional properties of the empirical items. As a consequence, the manifest distribution of items can be made to match the empirical distributions while the latent structure of the items is varied. By comparing the empirical taxometric results against those expected under these two conditions, one is able to test taxonic conjectures in a rigorous manner. In this article, we adapt the Ruscio and Ruscio approach but extend it by generating sampling distributions to quantify the amount of variation in taxometric results that might be expected under taxonic and dimensional situations. In the absence of such considerations, taxometric results are incomplete at best and potentially misleading at worst.

Present Investigation

We sought to improve on the foundation provided by previous studies by collecting data using structured diagnostic interviews from a population-based sample of youth. Our study advances knowledge by remedying limitations of past research (e.g., we assess all DSM–IV depressive symptoms and account for item skew) and by focusing explicitly on youth depression. The primary aim of this study is to address the continuity of depression among children and adolescents in a population-based sample. Although the taxonicity of adolescent depression has been explored previously (Ambrosini et al., 2002; Whisman & Pinto, 1997), both studies had limitations (e.g., relatively small sample sizes for taxometric research, nonrepresentative patient samples, indicator skewness). It is necessary to replicate these initial studies with adolescents in a population-based study. It is also important to extend this investigation to preadolescent children because the mechanisms underlying depression may differ for children and adolescents (see Garber, 2000, for reviews). Such developmental differences raise the possibility that the mechanisms contributing to the latent structure of depression may differ over the life span.

Whereas the mechanisms underlying depression may differ in youth and adults, the presentation and symptoms of depression appear to be mostly similar in children, adolescents, and adults (Kolvin & Sadowski, 2001; Kovacs et al., 1984), with slight differences in the level of particular symptoms seen in different ages (e.g., weight loss and suicidality being less common in younger children; Garber, 2000). Although there is ongoing research examining whether the symptomatic expression and manifestation of depression differs for youth and adults (e.g., Weiss & Garber, 2003), DSM–IV asserts that the same symptom criteria list can and should be used for depression among youth as adults, with the exception that irritability among youth is listed as a mood symptom along with depressed, sad mood (APA, 1994). Our primary goal to evaluate the latent structure of DSM–IV’s conceptualization and classification of depression among youth, we use the same set of DSM-defined depressive symptoms because the DSM asserts that the manifestation of depression is similar across age.

Although the presentation of depressive symptoms may be similar across age, research has shown that parent and youth reports of the youth’s depression levels are not interchangeable because they tend to correlate only moderately at best (Kolvin & Sadowski, 2001). Further, these different sources provide nonredundant perspectives on depression (Achenbach, McConaughy, & Howell, 1987). Examination of informant sources is a novel approach in taxometrics and is consistent conceptually with an emphasis on consistency of results from multiple, nonredundant tests in taxometric investigations. Prior taxometric study on depression has used different informant sources to evaluate the convergence of findings.

In summary, the present investigation uses taxometric analyses to evaluate whether depression is viewed better as a qualitatively discrete structural category or the extreme end along a continuum of affective processes. Analyses were performed on data from a population-based sample of children and adolescents. Youth’s self-reports and parents’ reports of the youth’s DSM–IV-defined major depressive symptoms were assessed using a structured clinical interview. Analyses were conducted using all of the DSM–IV-defined major depressive symptoms, as well as the selected domains of emotional distress and IDS symptoms. We also sought to determine whether our results varied as a function of age (childhood vs. adolescence) and gender.

Method

Participants and Measures

A total of 845 9–17-year-old youth (51% girls) constituted the present sample. The sample was drawn from the Georgia Health and Behavior Study (GHBS), a population-based study that recruited youth from Georgia (for full details of the sampling and response rate, see Lahey, Applegate, Waldman, Loft, Hankin, & Rick, 2004). In addition to the youth interview data, the GHBS study also included parent reports of the youth’s (ages 4–17) symptoms for various psychopathological disorders. Only the youth respondent-based interview data and parent reported data for the youth from ages 9–17 are used in the present report. The age distribution of the sample was essentially uniform across the 9–17 years. Sixty-eight percent of the sample was Caucasian, 24% was Black, 2% was Hispanic, 2% was bi- or multiracial, and 1% was Asian.

The GHBS sample was selected from a frame of residential mailing addresses in the Atlanta, GA, metropolitan statistical area (MSA). To assure that the GHBS sample was selected across all socioeconomic strata, researchers included in the frame households with and without telephones. A stratified simple random sample of households was selected from the population of addresses.

Following receipt of an advance letter, sampled households were screened in person for the presence of eligible children. Children and adolescents were eligible for participation if (a) they were 4–17 years of age on the date of screening, (b) they had co-resided with the consenting adult caretaker for at least 6 of the last 12 months, and (c) both the youth and the adult caretaker spoke English. In families with several eligible youth, one child was randomly selected. Parents and guardians who agreed to participate in the study gave written informed consent, and youth who were old enough to be interviewed (≥ 9 years of age) gave oral assent to participate. All interviews of adult caretakers and youth were conducted in person in the family’s home. Adult caretakers and youth were paid for their participation. Twenty-two families were classified as ineligible because of the absence of an English-speaking caretaker in the household.

Interviews began in January of 2000. The effective response rate was 75.9%. In March of 2001, recruitment and data collection on a second independent cohort was initiated to increase the sample size. The response rate for this second cohort was 70.6%. In total, the combination from both
cohorts provided data on 845 youth reports and the parents’ report of their youth (9–17 year olds) that were used for the present analyses.

Clinical interview. The youth and one of their caretakers (82% biological mothers; 14% biological fathers; 1% step-mothers; 3% grandmothers) were interviewed with a respondent-based clinical interview, the investigational version of the Child and Adolescent Psychopathology Scale (I-CAPS; Lahey et al., in press) to assess all of the symptoms of major depression as defined according to DSM–IV. The I-CAPS contains items that are based on and similar to those used in Version IV of the Diagnostic Interview Schedule for Children (DISC; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000) but also contains new items written in the DISC format to describe nonoverlapping emotions and behaviors referenced in widely used rating scales (e.g., Achenbach, 1991). The youth and caretakers were interviewed separately in private settings. Youth and parent respondents rated the I-CAPS items on a 4-point response scale that ranged from 1 (not at all) to 4 (very much). The youth were asked to rate each symptom item by thinking how well it described their emotional behavior, how often such symptoms occurred, and how serious the symptom was for the past 12 months. Parents rated the symptoms for how the items described their child’s emotion or behavior.

The I-CAPS follows the DISC strategy of asking multiple questions to address particular symptoms. To obtain the nine DSM–IV major depressive symptoms, we averaged the ratings from the multiple questions that assess the same symptom to create a composite score for that particular symptom. For example, for the symptom of fatigue or loss of energy, we averaged the following I-CAPS items: (a) sluggish and tired, and (b) tired out by little things, (c) sluggish and not energetic, and (d) had less energy than usual.

The general I-CAPS interview has demonstrated good internal consistency and test–retest reliability (Lahey et al., in press). In the present sample, internal consistency (coefficient alpha) for the youth’s reports of DSM–IV-defined depressive symptoms items was .92. Internal consistency of parents’ reports of youth’s depressive symptoms was .85. The means, standard deviations, and skewness of the DSM–IV major depressive symptoms are reported in Table 1. The test–retest reliability (Pearson correlation) of the depressive symptoms over 7–14 days was .81 among a subsample of 196 youth (out of 234 selected) who completed retest interviews. Validity of the I-CAPS depression items is provided by associations with measures of impairment and distress as reported by the youth and parents (Lahey et al., in press).

Taxometric procedures. To address the types versus dimensions question, we used two taxometric procedures developed by Meehl and his colleagues. The first, Maximum Covariance-Hitmax (MAXCOV or MAXCOV-HITMAX; Meehl, 1973; Meehl & Yonce, 1996), is one of the most widely used taxometric procedures for addressing questions about taxonicity (for a detailed overview of MAXCOV, see Meehl, 1973, or Waller & Meehl, 1998.). In MAXCOV, one studies the covariance between two variables as a function of a third variable. The function characterizing these conditional covariances is called a MAXCOV function and its shape depends on the taxonic status of the latent variable under study. If the latent variable is categorical with a base rate of .50, the MAXCOV curves tend to have a mountain-like peak. In samples in which the base rate is less than .50, the peak will be shifted to the right, whereas in samples in which the base rate is larger than .50, the peak will be shifted to the left. If the latent variable is continuous, the MAXCOV curves tend to resemble a flat line.

In our MAXCOV analyses, we studied the conditional covariances between all pairwise variables as a function of the composite of the remaining variables. Because multiple MAXCOV curves can be generated from the same set of variables, the MAXCOV procedure provides multiple tests of taxonicity for a given set of indicators or symptoms. Meehl’s approach emphasizes the consistency of results derived from multiple, nonredundant tests, rather than the statistical significance of a single test. If a latent class truly exists, then the various MAXCOV functions observed for a set of indicators should have a similar form. Furthermore, the taxon base-rate estimates derived from each MAXCOV analysis should converge on a single value (i.e., the true base rate of the latent class). Support for a taxonic interpretation of a construct is strengthened when these consistency tests are passed.

The second taxometric procedure we used was L-Mode (Waller & Meehl, 1998). L-Mode involves examining the latent score distribution, as estimated using traditional factor analytic techniques, for signs of bimodality. When the latent variable is taxonic, the distribution of factor score estimates should exhibit signs of bimodality. (It is important to note that the distribution of observed scores for a single indicator may not be bimodal because a single indicator will not approximate the latent distribution as well as a weighted composite of indicators.) Moreover, it is possible to use the location of those modes to derive two estimates of the base rate that should converge when the taxonic hypothesis is correct (see Waller & Meehl, 1998, Chapter 5, for mathematical details). When the latent variable is continuous, however, the distribution of factor score estimates will not exhibit clear signs of bimodality, and estimates of the base rate may or may not converge. In the analyses reported here, we estimate the factor score distributions and study those distributions for signs of bimodality.

1 Because previous work has highlighted the difficulty in obtaining sufficient sample sizes in the extreme ends of the MAXCOV sorting variable (see J. Ruscio & Ruscio, 2000), we computed the conditional covariances in ordered cuts of 80 cases rather than ordered cuts of specific scale scores (e.g., every half standard deviation). This approach appears to yield more stable estimates than the traditional approach. In our subsample analyses, we computed the conditional covariances for ordered cuts of 40 cases.

<table>
<thead>
<tr>
<th>Depressive symptom</th>
<th>M</th>
<th>SD</th>
<th>Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>Youth reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressed/irritable mood</td>
<td>1.85</td>
<td>0.69</td>
<td>0.82</td>
</tr>
<tr>
<td>Anhedonia</td>
<td>1.01</td>
<td>0.15</td>
<td>6.54</td>
</tr>
<tr>
<td>Weight loss/gain</td>
<td>1.50</td>
<td>0.48</td>
<td>1.00</td>
</tr>
<tr>
<td>Insomnia/hypersomnia</td>
<td>1.87</td>
<td>0.71</td>
<td>0.73</td>
</tr>
<tr>
<td>Psychomotor retardation/agitation</td>
<td>1.46</td>
<td>0.50</td>
<td>1.11</td>
</tr>
<tr>
<td>Fatigue</td>
<td>1.59</td>
<td>0.57</td>
<td>0.96</td>
</tr>
<tr>
<td>Worthlessness/guilt</td>
<td>1.56</td>
<td>0.63</td>
<td>1.27</td>
</tr>
<tr>
<td>Reduced concentration</td>
<td>1.71</td>
<td>0.65</td>
<td>0.73</td>
</tr>
<tr>
<td>Suicidal thoughts/behaviors</td>
<td>1.74</td>
<td>0.29</td>
<td>1.24</td>
</tr>
<tr>
<td>Parent reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressed/irritable mood</td>
<td>1.61</td>
<td>0.57</td>
<td>1.03</td>
</tr>
<tr>
<td>Anhedonia</td>
<td>1.0</td>
<td>0.12</td>
<td>3.88</td>
</tr>
<tr>
<td>Weight loss/gain</td>
<td>1.23</td>
<td>0.31</td>
<td>1.76</td>
</tr>
<tr>
<td>Insomnia/hypersomnia</td>
<td>1.35</td>
<td>0.46</td>
<td>1.76</td>
</tr>
<tr>
<td>Psychomotor retardation/agitation</td>
<td>1.16</td>
<td>0.30</td>
<td>2.83</td>
</tr>
<tr>
<td>Fatigue</td>
<td>1.31</td>
<td>0.44</td>
<td>1.95</td>
</tr>
<tr>
<td>Worthlessness/guilt</td>
<td>1.35</td>
<td>0.48</td>
<td>1.86</td>
</tr>
<tr>
<td>Reduced concentration</td>
<td>1.39</td>
<td>0.50</td>
<td>1.33</td>
</tr>
<tr>
<td>Suicidal thoughts/behaviors</td>
<td>1.82</td>
<td>0.34</td>
<td>2.03</td>
</tr>
</tbody>
</table>

Note. Total N = 845 for both youth’s self-reports and parents’ reports of youth’s depression. DSM–IV = Diagnostic and Statistical Manual of Mental Disorders (4th ed.).
though we discuss results using both MAXCOV and L-Mode, we focus on the MAXCOV results because, of the various taxometric procedures that have been developed by Meehl and his colleagues, MAXCOV is the most widely used and understood (Haslam & Kim, 2002).

Simulation of Taxonic and Dimensional Comparison Data

As discussed in the introduction, the interpretation of taxometric results is not clear cut when indicators are skewed. Specifically, when indicators are highly skewed, the resulting MAXCOV curves will be consistent with those expected when there is a low-base-rate taxon—even if the data were generated under a dimensional model. To draw valid taxometric inferences, one must evaluate MAXCOV curves with respect to those that would be expected in both taxonic and dimensional situations in which skew is present.

To do so, we simulated skewed data under taxonic and dimensional models following an iterative procedure highly similar to that developed by A. M. Ruscio and Ruscio (2002; see also J. Ruscio et al., 2003). Specifically, we simulated data for hypothetical subjects by generating scores according to a one-factor model in which the latent variable was either normally distributed (i.e., dimensional) or taxonic with a base rate of .10. We began each simulation by generating observed scores under a dimensional or taxonic model. Next, the distribution of each of the simulated indicators was skewed and scaled to conform to the distribution of empirical indicators by sorting the values and replacing them using the same values and frequencies observed in the empirical data (see, Ruscio & Ruscio, 2002 for more information). Next, the discrepancy between the interitem correlation matrices based on the simulated indicators and the empirical ones was calculated, and the vector of loadings was adjusted to minimize this discrepancy. The iterations proceeded until the average squared discrepancy, quantified as the root mean squared error (RMSE), was .10 or less. In short, this method allows us to capture the surface-level statistical properties of the observed variables (i.e., their means, standard deviations, skew, and interitem correlations) while allowing us to vary the latent structure that generated them (Ruscio & Ruscio, 2002).3

As might be expected, the simulated MAXCOV curves generated under each model varied from one simulation to the next because of random sampling errors. To quantify this variation, we simulated data under each kind of model (dimensional and taxonic) 50 times to approximate sampling distributions of MAXCOV curves expected under each model. In the analyses that follow, we discuss both the average simulated MAXCOV function (denoted as a solid line in figures) observed under each model (i.e., dimensional or taxonic), as well as the region corresponding to the standard error (estimated as the standard deviation of the simulated sampling distribution; denoted as the gray region in figures).

Results

MAXCOV Analyses

DSM–IV indicators of depression. MAXCOV analyses were conducted on the nine symptoms of DSM–IV major depression. Overall, the empirical MAXCOV curves were most similar to those expected under a dimensional model as opposed to a taxonic one. In Figure 1, we have illustrated the average MAXCOV curves for data based on youth reports (top row) and parent reports (bottom row). The shaded regions of the graphs illustrate the range of MAXCOV functions that should be observed under dimensional (left-most panels) and taxonic (right-most panels) situations. Notice that the empirical curves closely follow the ones expected under a dimensional model. Also notice that the empirical curves are well within the region that one might expect if the dimensional model was correct but sampling errors were contributing to deviations from the predicted curve. In contrast, the empirical curves

2 The base rate of 10% was chosen because it is consistent with 1-year prevalence rates of depression among youth samples (Garber, 2000). For example, a prospective community study of clinical depression, using the DISC to interview youth and applying DSM depression symptom criteria without impairment, found that children (ages 11–15) exhibited approximately a 2%–5% prevalence rate, and adolescents (ages 16–18) exhibited a 17% prevalence rate of depression prevalence rate in the past year (Hankin et al., 1998). Results from the MECA study of 9–17-year-olds (n = 1285), based on combined parent and youth DISC interviews, indicated a 8.8% 6-month prevalence rate of DSM–III–R depression based on the DSM symptom criteria only and a 2.6% prevalence rate based on the DSM criteria plus impairment (Shaffer et al., 1996). In addition to these prevalence estimates from published studies of community samples, the empirical 1-year prevalence rate of depression in the GHBS sample, based on half of the sample being interviewed with the DISC–IV and applying stringent DSM–IV symptom criteria plus impairment, was 2% for the entire sample of 9–17-year-olds, 3% for girls, 1% for boys, 1.5% for children (ages 9–13), and 2.5% for adolescents (ages 14–18). These empirical prevalence estimates from the GHBS appear very similar to the rates found in the MEGA study when DSM–IV depression symptom and impairment criteria were applied. It was not possible to estimate 12-month depression prevalence rates for the GHBS sample using only DSM–IV depression symptom criteria without impairment. Given this discrepancy in prevalence rates for the GHBS sample compared with other published prevalence rates from general population samples, we conducted taxometric simulations using both a 2% and 10% base-rate estimate. The findings were comparable regardless of which base rate was assumed. It is important to note that when we adjusted downward the assumed base rate of depression in these simulations, we still had the power to detect a latent taxon, if one existed. Varying the assumed base rate has the effect of shifting the peak illustrated in the figures horizontally and, as will be seen, there are no obvious peaks to be shifted in the empirical curves. Readers interested in the details of the simulations can view the S-Plus code at the following Web page: http://www.psych.uiuc.edu/~rcfraley/hflwnotes.htm.

3 It should be noted that we did not model “nuisance covariance” (i.e., the covariation that may exist between indicators within a group, because of continuous sources of variation that are common to the indicators) in our simulations for two reasons. First, to the extent to which indicator covariation exists within a group, it is because (a) the “group” is not exclusively composed of taxon (or nontaxon) members or (b) there is an additional factor common to the indicators—presumably, one that does not covary with the taxonic variable—that generates indicator covariation. In this latter case, the appropriate model is not a pure taxonic one and does not map well onto the categorical assumptions made by the DSM–IV. If, for example, the covariation among indicators is a weighted function of both taxonic and dimensional sources of variation, then a continuous model is necessary to conceptualize and assess that variation appropriately (see Fraley & Spieker, 2003b). A second reason that we did not model nuisance covariation is that doing so necessarily makes the taxonic model more flexible than a dimensional one. In other words, it is possible to reproduce any empirical MAXCOV function by adding just the right combination of taxonic and dimensional sources of influence. Thus, to avoid this kind of looseness, we chose to evaluate strong versions of both models: a dimensional model that assumes no taxonic variation and a taxonic model that assumes no dimensional variation (i.e., variation that would give rise to nuisance covariation). Although we believe the strong form of the taxonic hypothesis is reasonable in this context, we thought it would be valuable to investigate just how much nuisance covariation would be necessary to
deviate considerably from those expected under a taxonic model (see the right-most panels).

In Table 2, we summarize the base rate estimates derived from our empirical analyses. Overall, there was quite a bit of variation in these estimates ($SD = .20$ for both youth and parent reports), with a mean estimate of .28 based on youth reports and a mean estimate of .28 based on parent reports. To determine the estimates that should be expected on the basis of dimensional and taxonic models, we examined the average base-rate estimates obtained in our simulations. The empirical base-rate estimates were more similar to those expected if the underlying structure of depression is dimensional than taxonic (see Table 2). Moreover, the amount of variability observed in the empirical base-rate estimates (.20) was more consistent with that expected under a dimensional (.16) than a taxonic (.01) model.

**Distress versus involuntary defeat.** Beach and Amir (2003) recently suggested that depressive symptoms of distress (e.g., sadness) might tap a different kind of latent variable than those that reflect feelings of IDS (e.g., loss of appetite). Beach and Amir presented taxometric analyses that corroborated this conjecture, showing that the MAXCOV curves for BDI items indicative of IDS were more taxonic than BDI items indicative of distress. To conceptually replicate this finding, we identified four DSM–IV depressive symptoms that were indicative of distress (i.e., depressed/irritable mood; worthlessness; poor concentration; and suicidal thoughts/behaviors) and five indicators indicative of IDS (weight loss/gain; insomnia/hypersomnia; psychomotor agitation/retardation; fatigue; and anhedonia) and analyzed those sets of items separately. These distress and IDS symptoms were chosen on the basis of the BDI symptoms that Beach and Amir used in their distress and IDS taxometric analyses. The resulting averaged MAXCOV curves for distress (left side) and defeat items (right side) are presented in Figure 2.

As can be seen, neither set of items behaved in a way that would be consistent with a taxonic model. Rather, both sets of items were most consistent with a dimensional model of depression. Although the MAXCOV curves we obtained are highly similar in form to the ones obtained by Beach and Amir (2003), our simulations corroborate J. Ruscio, Ruscio, and Keane’s (2004) observation that the “taxonicity” of IDS may be due to item skew rather than true taxonicity. The base-rate estimates derived from our empirical analyses and our simulation analyses are reported in Table 2. Notice that there is considerable variation in the base-rate estimates obtained in the empirical data. Also, it is noteworthy that the average base-rate estimates (and the standard deviations in these estimates) tended to be closer to those expected under a dimensional model as opposed to a categorical model.

**Children versus adolescents.** As we introduced earlier, DSM–IV specifies that the manifest symptoms of DSM–IV-defined depression are largely the same across childhood, adolescence, and adulthood. The correlates of depressive symptoms, however, may differ from childhood to adolescence. This raises the possibility that the answer to the types versus dimensions question could be different, depending on the age group studied. To investigate this possibility, we conducted MAXCOV analyses on the nine DSM–IV depression symptoms separately for children between the ages of 9 and 13 ($n = 454$) and adolescents between the ages of 14 and 17 ($n = 391$). The results of those analyses are summarized in Figure 3. In short, the average empirical MAXCOV functions were more consistent with those that would be expected if a dimensional model were correct, regardless of the age in question. The base-rate estimates derived from our empirical analyses and our simulation analyses are reported in Table 2. Notice that there is considerable variation in the base-rate estimates obtained in the empirical data. Also, it is noteworthy that the average base-rate estimates (and their standard deviations) tended to be closer to those expected under a dimensional as opposed to a categorical model.

**Boys versus girls.** There are well-known sex differences in the prevalence of adolescent depression (Hankin & Abramson, 2001). Given that more girls exhibit depression than boys, it is possible that the mechanisms giving rise to depression may differ for the sexes (e.g., Cyranowski, Frank, Young, & Shear, 2000; Silberg et al., 1999) so that depression might be dimensional for one sex but taxonic for the other. To examine this possibility, we conducted MAXCOV analyses on the nine DSM–IV depression symptoms separately for boys ($n = 408$) and girls ($n = 437$). The results of those analyses are summarized in Figure 4. The empirical MAXCOV functions for both sexes were more consistent with what would be expected if a dimensional model was correct. Both boys (left side) and girls (right side) exhibited MAXCOV curves that closely paralleled the predictions of a dimensional model and diverged from those expected under a low-base-rate taxonic model. Notice that there is considerable variation in the base-rate estimates obtained in the empirical data (see Table 2). Also, it is

---

4 Technically, if it were the case that younger children, for example, could be characterized by a categorical model of depression, whereas adolescents could be characterized best by a dimensional one, the analyses that we presented previously would fully capture this state of affairs. Specifically, a low-base-rate taxon would emerge, and children classified into that taxon would be predominantly younger as opposed to older. Nonetheless, it may be valuable to study explicitly the taxonicity of depression separately within different age groups.
noteworthy that the average base-rate estimates (and their standard deviations) tended to be closer to those expected under a dimensional model as opposed to a categorical model.

L-Mode Analyses

To estimate the factor score distributions for each domain, we used the L-Mode programs written by John Ruscio for S-Plus.5 The estimated density plots for each domain are shown in Figure 5. For comparison, the upper-left panel of Figure 5 shows the distribution of factor scores that should be observed under a taxonic model (solid line) and a dimensional model (dotted line).

Figure 1. MAXCOV functions for DSM-IV symptoms of depression based on youth reports (top row) and parent reports (bottom row). The left-hand panels show the empirical MAXCOV functions (connected dots) as well as the range of MAXCOV functions, depicted by the shaded regions, that would be expected if a dimensional model were correct. The right-hand panels show the same empirical MAXCOV functions and the range of MAXCOV functions, depicted by the shaded region, that would be expected if a categorical model were correct. MAXCOV = Maximum Covariance-Hitmax.

5 These programs can be freely downloaded at http://www.etown.edu/psychology/Faculty/Ruscio.htm
Notice that the estimated distributions based on the empirical data are much more similar to those expected under a dimensional as opposed to a taxonic model. Although some of the distributions do show signs of bimodality, the overall form of the distributions deviates considerably from that expected under a taxonic model. In addition, the taxon base-rate estimates are (a) inconsistent with those derived from the MAXCOV analyses and (b), in many cases, inconsistent within the same analysis (see Table 2). The L-Mode estimates of the base rate for the full sample derived from the left and right modes were .77 and .48, respectively. These estimates were quite different, and both diverged from the estimates obtained from the MAXCOV analyses and (b), in many cases, inconsistent within the same analysis (see Table 2). The L-Mode estimates of the base rate for the full sample derived from the left and right modes were .77 and .48, respectively. These estimates were quite different, and both diverged from the estimates obtained from the MAXCOV analyses.

**Table 2**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Empirical MAXCOV estimates</th>
<th>Simulated Dimensional</th>
<th>Simulated Taxonic</th>
<th>Empirical L-Mode estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>Range</td>
<td>M</td>
</tr>
<tr>
<td>DSM</td>
<td>.20</td>
<td>.20</td>
<td>.05–.64</td>
<td>.24</td>
</tr>
<tr>
<td>DSM (parent)</td>
<td>.28</td>
<td>.20</td>
<td>.10–.77</td>
<td>.21</td>
</tr>
<tr>
<td>IDS</td>
<td>.22</td>
<td>.19</td>
<td>.08–.65</td>
<td>.24</td>
</tr>
<tr>
<td>IDS (parent)</td>
<td>.32</td>
<td>.17</td>
<td>.11–.67</td>
<td>.17</td>
</tr>
<tr>
<td>Distress</td>
<td>.19</td>
<td>.23</td>
<td>.03–.61</td>
<td>.27</td>
</tr>
<tr>
<td>Distress (parent)</td>
<td>.41</td>
<td>.23</td>
<td>.11–.74</td>
<td>.27</td>
</tr>
<tr>
<td>Child</td>
<td>.23</td>
<td>.21</td>
<td>.07–.69</td>
<td>.25</td>
</tr>
<tr>
<td>Child (parent)</td>
<td>.23</td>
<td>.19</td>
<td>.10–.61</td>
<td>.25</td>
</tr>
<tr>
<td>Adolescent</td>
<td>.26</td>
<td>.20</td>
<td>.13–.74</td>
<td>.24</td>
</tr>
<tr>
<td>Adolescent (parent)</td>
<td>.31</td>
<td>.21</td>
<td>.10–.84</td>
<td>.28</td>
</tr>
<tr>
<td>Boys</td>
<td>.36</td>
<td>.20</td>
<td>.07–.87</td>
<td>.31</td>
</tr>
<tr>
<td>Girls</td>
<td>.39</td>
<td>.20</td>
<td>.13–.99</td>
<td>.34</td>
</tr>
<tr>
<td>Girls (parent)</td>
<td>.38</td>
<td>.22</td>
<td>.13–.99</td>
<td>.36</td>
</tr>
</tbody>
</table>

*Note.* On the basis of our simulations, the expected base-rate estimates for the left mode (LM) and right mode (RM) in L-Mode would be approximately .08–.12 if a categorical model were correct. MAXCOV = Maximum Covariance-Hitmax; DSM = *Diagnostic and Statistical Manual of Mental Disorders*; IDS = involuntary defeat syndrome.

Notice that the estimated distributions based on the empirical data are much more similar to those expected under a dimensional as opposed to a taxonic model. Although some of the distributions do show signs of bimodality, the overall form of the distributions deviates considerably from that expected under a taxonic model. In addition, the taxon base-rate estimates are (a) inconsistent with those derived from the MAXCOV analyses and (b), in many cases, inconsistent within the same analysis (see Table 2). The L-Mode estimates of the base rate for the full sample derived from the left and right modes were .77 and .48, respectively. These estimates were quite different, and both diverged from the estimates obtained from the MAXCOV analyses.

**Discussion**

**Latent Structure of Youth Depression**

To determine whether depression among youth is structured continuously or categorically, we applied Meehl’s taxometric techniques to youth’s self-reports and parents’ reports of youth depression in a population-based general sample of children and adolescents. Our results suggest that youth depression is continuously, not categorically, distributed. This finding held for both youth and parent reports, for all *DSM–IV* depressive symptoms, as well as for different domains of depression (emotional distress or somatic, IDS symptoms). Further, our data suggest that depression is dimensional for both children and adolescents and boys and girls.

Although previous studies have examined the taxonicity of depression, there are at least two important ways in which the current investigation advances the field in novel directions. First, we explicitly modeled the effects of skewness on taxonicity. We simulated data with the same statistical properties as the empirical data (e.g., means, variances, skewness, and average correlations) but generated under different latent models (i.e., continuous vs. categorical). By comparing the empirical taxometric results (e.g., MAXCOV curves) against those expected under these different situations, we were able to provide a more rigorous answer to the types versus dimensions question than has been possible with most research up until now.

Skewness raises a number of questions about the proper interpretation of previous taxometric investigations of psychopathology. As noted in the introduction, past taxometric findings regarding depression have been mixed. At face value, one might conclude from the existing literature that more extreme forms of depression may be categorical, whereas less severe, emotional distress forms of depression are not. We would like to offer an alternative explanation for previous findings of taxonicity. Specifically, we propose that what researchers typically construe as more severe forms of depression may be viewed best as the extreme end of a continuum of depressive experiences. As such, “severe” forms of depression—when treated as categories—should have lower base rates and, when treated as continua, should exhibit higher levels of indicator skew. In fact, when one examines the skewness of symptoms of depression, indicators of more severe forms of depression (e.g., IDS) do exhibit a higher degree of skew than the less severe forms of depression (e.g., emotional distress). In Beach and Amir’s (2003) study, for example, the distress items had an average skew of 1.75, whereas the IDS symptoms had an average skew of 2.43. Further, visual comparison of the MAXCOV curves for our study and Beach and Amir’s (2003) study suggests that the MAXCOV curves for the IDS symptoms are very similar. This suggests that, by not explicitly accounting for the skewness of the ([text continues on page 108](#))
Figure 2. MAXCOV functions for distress and involuntary defeat syndrome items based on youth reports (first row) and parent reports (second row). The empirical MAXCOV functions are represented by the dotted lines, the solid line represents the average for the simulated MAXCOV functions, and the shaded region depicts the range of MAXCOV functions that would be expected under a dimensional or taxonic model (i.e., the standard error). MAXCOV/H11005

Maximum Covariance-Hitmax.
Figure 3. MAXCOV functions for children and adolescents based on youth reports (first row) and parent reports (second row). The empirical MAXCOV functions are represented by the dotted lines, the solid line represents the average for the simulated MAXCOV functions, and the shaded region depicts the range of MAXCOV functions that would be expected under a dimensional or taxonic model (i.e., the standard error). MAXCOV = Maximum Covariance-Hitmax.
Figure 4. MAXCOV functions for girls and boys based on youth reports (first row) and parent reports (second row). The empirical MAXCOV functions are represented by the dotted lines, the solid line represents the average for the simulated MAXCOV functions, and the shaded region depicts the range of MAXCOV functions that would be expected under a dimensional or taxonic model (i.e., the standard error). MAXCOV = Maximum Covariance-Hitmax.
Figure 5. L-Mode estimated factor score distributions. The upper-left panel illustrates the expected distributions under a taxonic (solid line) and dimensional (dotted line) model. The remaining panels illustrate the estimated factor score distributions for youth (solid lines) and parent (dotted lines) reports. DSM = Diagnostic and Statistical Manual of Mental Disorders.
IDS symptoms, Beach and Amir may have inadvertently and incorrectly concluded that the more severe IDS symptoms are taxonic (see also J. Ruscio et al., 2004).

A second important feature of the present investigation was our focus on youth depression, including both children and adolescents. Given the limitations of the few adolescent studies (Ambrosini et al., 2002; Whisman & Pinto, 1997) and the fact that no study had evaluated the latent structure of depression in preadolescent children, it was important to extend taxometric investigation into the latent structure of depression in a population-based sample of children and adolescents. Our results suggest that there is no more evidence of taxonicity in children (ages 9–13) than in adolescents (ages 14–17). In light of this finding, it seems likely that depression behaves as a dimensional variable across childhood, adolescence, and, perhaps, into adulthood.

**Strengths and Limitations**

There are several methodological advantages to the present investigation. First, the sample was composed of children and adolescents drawn from the general population. The use of a general population-based sample is a substantial strength compared with previous taxometric studies that used more restricted samples, such as college students or clinical patients seeking inpatient or outpatient services. Further, our relatively large sample size (N = 845 for youth self-reports and parent reports) enabled us to conduct taxometric analyses separately as a function of sex and age. Still, larger sample sizes are recommended in future research examining the latent structure of depression by subtype domain, sex, and age because large samples are need for taxometric investigations, especially when the likely base rates of a putative depression taxon may be very low (e.g., less than 10% as is likely in the IDS domain and child-sample analyses). A second advantage of the present study is the use of both youth’s self-reports and parents’ reports of their youth’s DSM–IV-defined depressive symptoms. No prior taxometric study has used the strategy of multiple informants to examine consistency of results across reporters. Using multiple informants provides converging evidence that DSM–IV-defined depression may be continuous and that results are not specific to one source or easily explained away by informant bias. A third advantage is the use of structured clinical interviews to assess all DSM–IV depressive symptoms.

We should also note some limitations inherent with the use of taxometric methods, in general, and in this study, specifically (see Beauchaine, 2003, and Lenzenweger, 2004, for a review and discussion of assumptions, strengths, and weakness of taxometric research). First, conclusions about the latent structure of a construct based on taxometric output may yield misleading findings if the base rates are very low (e.g., below 10%), if the construct’s indicators are not valid, and if the separation between a putative taxon and its complement is not large. Second, we assumed that there was no “nuisance covariance” when simulating comparison data from categorical models (see Footnote 3). Although we believe this assumption allows us to best contrast alternative structural models of depression, it is possible that this is a strong assumption and may have biased the results toward a dimensional interpretation. Third, although the empirical results appear to fall closely within the values expected under the simulated dimensional model, a few of the values were outside of the dimensional sampling distribution. Fourth, we focused only on DSM symptoms of depression rather than additional potential indicators of depression, such as biological measures (e.g., hypothalamic-pituitary-adrenal axis [HPA] reactivity, sleep patterns, neurotransmitter levels) and psychosocial measures (e.g., cognitive vulnerabilities, interpersonal deficits). Using multiple indicators with established validity will enable more powerful taxometric tests in the future, especially for putative depression subtypes with low base rates, such as IDS (Beauchaine, 2003). Last, our simulations were based on A. M. Ruscio and Ruscio’s (2002) work, and this approach is relatively new and its validity has not been extensively evaluated. As recommended by Meehl (2004), we have based our interpretations on the totality and consistency of the taxometric evidence. Yet, given the limitations of taxometrics, generally, as well as the specific decisions we made in this study (e.g., not modeling nuisance covariance), the findings may not be conclusive. The limitations and cautions we have discussed should be kept in mind when evaluating the evidence reported in this article.

**Implications of the Present Study**

**Classification: Theory and practice.** The current diagnostic system and nomenclature, as embodied in DSM–IV, assumes that psychopathological disorders, such as depression, are categorical. Our results, taken together with other studies, argue against DSM’s categorical emphasis for depression in children, adolescents, and adults. One approach for organizing and classifying individual differences in depression is to conceptualize these differences as varying continuously (see also Hartman et al., 2001; Lahey et al., 2004; van den Oord, Pickles, & Waldman, 2003, for relevant factor analytic research).

**Methods and measurement.** Dimensional approaches may be valuable for studying and assessing depression. In fact, if individual differences in depression really are continuous, but researchers continue to use categorical measurement models, the empirical study of depression may suffer. For example, the common practice of forming groups by dichotomizing continuously distributed depression scores can lead to serious problems for measurement precision (Ruscio & Ruscio, 2002). This is equivalent to throwing away up to 36% of the “true score” variance (Cohen, 1983). Given the amount of attention that depression researchers have devoted to developing sensitive and precise measures of depression (e.g., questionnaires and interviews), it seems to be a disservice to throw away this information by categorizing people into artificial groups (MacCallum, Zhang, Preacher, & Rucker, 2002). When categories are used in lieu of linear continua, statistical power is severely compromised (Fraley & Spieker, 2003a). Conceptualizing, assessing, and analyzing depression as a continuous variable should improve statistical power and, ultimately, the ability of researchers to uncover correctly the causes and consequences of depression. Ultimately, to map the nomological network of depression as precisely as possible, researchers and clinicians must use a measurement model that allows for the full representation of the individual variation in depression.
Conclusion

In summary, the present investigation examined the latent structure of depression in a population-based sample of children and adolescents. Self and parents’ reports of the youth’s DSM-IV-defined major depressive symptoms were assessed using structured clinical interview. Taxometric techniques, accounting for symptom skew, suggested that the latent structure of depression was dimensional, not categorical. The dimensional structure of depression was found for all depressive symptoms, different domains of depression (i.e., emotional distress and IDS), boys and girls, and children and adolescents. We hope that these findings can be used to inform the classification debate and the study of depression across development.

References


Ruscio, A. M., & Ruscio, J. (2002). The latent structure of analogue...
depression: Should the Beck Depression Inventory be used to classify groups? Psychological Assessment, 14, 135–145.