

# Are Increased Weight and Appetite Useful Indicators of Depression in Children and Adolescents?

David A. Cole, Sun-Joo Cho, and Nina C. Martin  
Vanderbilt University

Eric A. Youngstrom  
University of North Carolina at Chapel Hill

John S. March  
Duke University Medical Center

Robert L. Findling  
Case Western Reserve University

Bruce E. Compas  
Vanderbilt University

Ian M. Goodyer  
University of Cambridge

Paul Rohde  
Oregon Research Institute

Myrna Weissman  
Columbia University College of Physicians and Surgeons

Marilyn J. Essex  
University of Wisconsin School of Medicine and Public Health

Janet S. Hyde  
University of Wisconsin–Madison

John F. Curry  
Duke University Medical Center

Rex Forehand  
University of Vermont

Marcia J. Slattery  
University of Wisconsin School of Medicine and Public Health

Julia W. Felton and Melissa A. Maxwell  
Vanderbilt University

This article was published Online First June 11, 2012.

David A. Cole, Sun-Joo Cho, Nina C. Martin, Bruce E. Compas, Julia W. Felton, and Melissa A. Maxwell, Department of Psychology and Human Development, Vanderbilt University; Eric A. Youngstrom, Departments of Psychology and Psychiatry, University of North Carolina at Chapel Hill; John S. March and John F. Curry, Department of Psychiatry and Behavioral Sciences, Duke University Medical Center; Robert L. Findling, Department of Psychiatry, Case Western Reserve University; Ian M. Goodyer, Department of Psychiatry, University of Cambridge, Cambridge, England; Paul Rohde, Oregon Research Institute, Eugene, Oregon; Myrna Weissman, Department of Epidemiology and Psychiatry, Columbia University College of Physicians and Surgeons; Marilyn J. Essex and Marcia J. Slattery, Department of Psychiatry, University of Wisconsin School of Medicine and Public Health; Janet S. Hyde, Department of Psychology, University of Wisconsin–Madison; Rex Forehand, Department of Psychology, University of Vermont.

This research was supported in part by the following grants. David Cole: Gifts from Patricia and Rodes Hart and from the Warren family; Bruce Compas: NIMH Grants R01MH069940 and R01MH069928 and a gift from Patricia and Rodes Hart; Robert Findling and Eric Youngstrom: NIMH Grants R01MH066647 and P20-MH066054 and a Clinical Research Center Grant from the Stanley Medical Research Institute; Rex Forehand: NIMH Grants R01MH069940 and R01MH069928 and a gift from the Heinz and Rowena Ansbacher Professorship; Marilyn J. Essex: John D. and Catherine T. MacArthur Foundation Research Network on Psychopathology and Development and NIMH Grants R01MH44340, P50-MH52354, P50-MH69315, and P50-MH84051; Janet S. Hyde: NIMH Grant R01MH44340; Ian Goodyer: National Health Service (NHS) Health Technology Assessment Programme, Central Manchester and Manchester Children's University Hospitals NHS Trust, and the Cambridge and Peterborough Mental Health Trust; John S. March: NIMH 98-DS-0008 (Treatment for Adolescents With Depression Study [TADS]); John F. Curry was a co-investigator who collaborated on this project.; Paul Rohde: NIMH Grants MH56238, MH67183, and MH 56238; Marcia J. Slattery: Grant 1UL1RR025011 from the Clinical and Translational Science Award Program of the National Center for Research Resources in NIH and NIMH grant P50-MH69315; Myrna Weissman: NIMH Grant R01MH063852 and

NIMH Contract N01 MH90003. Eric A. Youngstrom receives or has received travel support or acted as a consultant for Bristol-Myers Squibb and Lundbeck. John S. March has served as a consultant or scientific advisor to Pfizer, Lilly, GSK, BMS, Johnson and Johnson, Psymetrix, Atentiv, Avanir, Alkermes, Translational Venture Partners, Vivus and MedAvante; received study drug for an NIMH-funded study from Eli Lilly and from Pfizer; serves on a DSMB for NIDA, Lilly and Otsuka; is an equity holder in MedAvante; receives royalties from Guilford Press, Oxford University Press and MultiHealth Systems. Dr. March receives research support from Pfizer, NIMH, NIDA, and NARSAD. Dr. March has not engaged in promotional work, e.g., speakers bureau or training, for over 15 years. Dr. March's conflict of interest is fully reported to the University, viewable at <http://www.dcri.duke.edu/research/coi.jsp>, and a conflict of interest management plan has been established. Robert Findling receives or has received research support, acted as a consultant, received royalties from, and/or served on a speaker's bureau for Abbott, Adrenex, Alexza, American Psychiatric Press, AstraZeneca, Biovail, Bristol-Myers Squibb, Daiippon Sumitomo Pharma, Forest, GlaxoSmithKline, Guilford Press, Johns Hopkins University Press, Johnson & Johnson, KemPharm Lilly, Lundbeck, Merck, National Institutes of Health, Neuropharm, Novartis, Noven, Organon, Otsuka, Pfizer, Physicians' Post-Graduate Press, Rhodes Pharmaceuticals, Roche, Sage, Sanofi-Aventis, Schering-Plough, Seaside Therapeutics, Sepracore, Shionogi, Shire, Solvay, Stanley Medical Research Institute, Sunovion, Supernus Pharmaceuticals, Transcept Pharmaceuticals, Validus, WebMD, and Wyeth. In the past two years, Myrna Weissman received funding from the National Institute of Mental Health (NIMH), the National Institute on Drug Abuse (NIDA), the National Alliance for Research on Schizophrenia and Depression (NARSAD), the Sackler Foundation, the Templeton Foundation, and the Interstitial Cystitis Association; and receives royalties from the Oxford University Press, Perseus Press, the American Psychiatric Association Press, and Multi-Health Systems.

Correspondence concerning this article should be addressed to David A. Cole, Department of Psychology and Human Development, Vanderbilt University, GPC 0552, 230 Appleton Place, Nashville, TN 37203. E-mail: david.cole@vanderbilt.edu

During childhood and adolescence, physiological, psychological, and behavioral processes strongly promote weight gain and increased appetite while also inhibiting weight loss and decreased appetite. The Diagnostic and Statistical Manual-IV (*DSM-IV*) treats both weight-gain/increased-appetite and weight-loss/decreased-appetite as symptoms of major depression during these developmental periods, despite the fact that one complements typical development and the other opposes it. To disentangle the developmental versus pathological correlates of weight and appetite disturbance in younger age groups, the current study examined symptoms of depression in an aggregated sample of 2307 children and adolescents, 47.25% of whom met criteria for major depressive disorder. A multigroup, multidimensional item response theory model generated three key results. First, weight loss and decreased appetite loaded strongly onto a general depression dimension; in contrast, weight gain and increased appetite did not. Instead, weight gain and increased appetite loaded onto a separate dimension that did not correlate strongly with general depression. Second, inclusion or exclusion of weight gain and increased appetite affected neither the nature of the general depression dimension nor the fidelity of major depressive disorder diagnosis. Third, the general depression dimension and the weight-gain/increased-appetite dimension showed different patterns across age and gender. In child and adolescent populations, these results call into question the utility of weight gain and increased appetite as indicators of depression. This has serious implications for the diagnostic criteria of depression in children and adolescents. These findings inform a revision of the DSM, with implications for the diagnosis of depression in this age group and for research on depression.

*Keywords:* depression, K-SADS, IRT, weight, appetite

Appetite increase and decrease as well as weight gain and loss are listed by the *Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition (DSM-IV)* (American Psychiatric Association, 2000) as symptoms of depression for all ages, despite the fact that during childhood and adolescence weight gain and increased appetite are normative and weight loss and appetite decrease are not. Treating increased appetite and weight as indicators of depression in youths may be problematic, as so many physiological and psychological processes affect weight and appetite during this developmental period that the effect of depression on these symptoms may be negligible. The primary goal of this article is to examine depression symptoms (as defined in DSM) in 2307 children and adolescents, assessed via semistructured clinical interview (i.e., the Kiddie Schedule of Affective Disorders and Schizophrenia for School-Aged Children; KSADS), to test whether or not changes in appetite and weight provide useful information in the diagnosis of depression or in the assessment of its severity. Answers to these questions could inform efforts to refine research and clinical diagnostic criteria for depression in youths. Further, the revision of the DSM that is underway should be informed by the best possible empirical evidence. The overarching goal of this study is to provide such information about depression in children and adolescents.

Recent research provides theoretical reasons why increased appetite and weight may not be useful indicators of child and adolescent depression (Felton, Cole, Tilghman-Osborne, & Maxwell, 2010; Maxwell & Cole, 2009). Taken together, these articles argue that at least two broad classes of factors affect weight and appetite in childhood and adolescence to such an extent that depression may not be associated with sufficient increases in weight or appetite to warrant regarding them as symptoms of the disorder. One set of factors involves physiological and metabolic changes: hypothalamic neuropeptides (Fehm et al., 2001; Jéquier & Tappy, 1999; Terasawa & Fernandez, 2001), endogenous reward systems (Neary, Goldstone, & Bloom, 2004; Stanley, Wynne, McGowan, & Bloom, 2005), and puberty-related hormonal levels (Ahmed et al., 1999; Bornstein, Schuppenies, Wong, & Licinio, 2006; Neary

et al., 2004; Romeo & McEwen, 2006). During childhood and adolescence, such hormonal and physiological changes profoundly affect weight and appetite.

A second set of factors includes psychological and behavioral variables, which also affect appetite and weight during this developmental period. One such behavior is dieting. Up to 60% of adolescent girls and more than 10% of adolescent boys are “on a diet” at any given time (Patton et al., 1997). Ironically, adolescent dieting is actually predictive of weight *gain*, after controlling for initial body mass (Field et al., 2003; Neumark-Sztainer et al., 2006; Stice, Cameron, Killen, Hayward, & Taylor, 1999). A second behavioral factor is exercise (Goldberg & King, 2007; Ross et al., 2000). Steep reductions in physical activity occur during adolescence, largely because of declining participation in nonorganized sports (Sallis, 2000). A third psychological factor is stress. Self-reported levels of social stress reach their peak during adolescence and early adulthood (Turner, Wheaton, & Lloyd, 1995). During periods of stress, individuals exhibit poorer impulse control (Tice, Bratslavsky, & Baumeister, 2001) and show increased inclination to use food to alleviate distress (Gallup & Castelli, 1989; Markowitz, Friedman, & Arent, 2008).

The net result of these relatively typical physiological and psychological factors is that people typically gain 50% of adult body weight during adolescence (Lerner & Steinberg, 2004). With so many systems vying for control of weight and appetite, the question becomes, does any residual opportunity exist for depression to precipitate further increases in appetite and weight during this developmental period? Conversely, symptoms of *decreased* appetite and weight emerge despite these normative developmental forces and therefore may serve as strong indicators of depression during this time period.

Weiss and Garber (2003) and Felton et al. (2010) reviewed studies that examined weight and appetite disruption in depressed children and adolescents. Many of these studies bundled weight/appetite increase together with weight/appetite decrease, perhaps because *DSM-IV* describes them as a single symptom (e.g., Flament, Cohen, Choquet, Jeammet, & LeDoux, 2001). Of studies

that have separated this symptom into its components, most did not actually test the relation of symptoms to the disorder (e.g., Borchardt & Meller, 1996; Friedman, Hurt, Clarkin, Corn, & Aronoff, 1983; Strober, Green, & Carlson, 1981). Nevertheless, these studies often showed that weight gain and appetite increase were among the least prevalent symptoms in youths who were depressed (e.g., Mitchell, McCauley, Burke, & Moss, 1988; Yorbik, Birmaher, Axelson, Williamson, & Ryan, 2004). Yorbik et al.'s (2004) study is especially interesting in that they factor analyzed symptoms of depression based on KSADS data in samples of depressed children and adolescents. In both age groups, evidence emerged for a separate weight-gain/appetite-increase factor, suggesting that weight and appetite measured something qualitatively different from other depressive symptoms. Unfortunately, several methodological issues make it difficult to use their results to address our questions. Varimax rotation prevented examination of correlations between the factors. Use of Kaiser's criterion may have overestimated the number of factors (Zwick & Velicer, 1986). Use of only depressed individuals in their analyses could restrict range on key variables and attenuate parameter estimates.

We did find two studies that (a) separately examined increased appetite and weight gain and (b) formally tested whether the likelihood of having these symptoms was conditional on having the disorder. One was Mitchell et al.'s (1988) study of 125 children and adolescents who were in psychiatric treatment, 95 of whom met criteria for major depression. In this sample, not only were weight gain and appetite increase the two least prevalent symptoms, but their occurrence was not statistically related to the diagnosis of major depression. Furthermore, weight loss and appetite decrease were much more prevalent. The second was Roberts, Lewinsohn, and Seeley's (1995) community-based study of 1709 high school students, comparing 44 depressed with 1665 nondepressed individuals. Odds ratios revealed that both increased appetite and weight gain had statistically higher prevalence estimates among depressed than nondepressed participants. Within the depressed group, however, both symptoms had low prevalence estimates, placing them among the bottom four of 27 symptoms. Conversely, moderately to substantially higher prevalence estimates emerged for weight and appetite loss.

Despite the strengths of these studies, two issues qualify their implications for the current question. First, the criterion with which the symptoms were compared was the presence or absence of a major depressive episode. Such diagnoses were based, in part, upon the presence or absence of each particular symptom. This represents a part-whole problem that has the potential to create an upward bias in estimates of the relation between disorder and symptoms. Second, in the Roberts et al.'s study, the total number of depressed cases was relatively small (only 44 of 1709), and the sample was predominantly White (91.1%) and represented a relatively narrow age range (14–18 years). In the Mitchell et al. (1988) study, the number of depressed cases was much larger but the comparison group was relatively small ( $n = 30$ ) and consisted of both in- and outpatient youths without major depression.

Several aspects of the current study address these concerns. First, regarding the sample, our goal was to obtain a relatively large sample of youths that spanned a wide age range, was ethnically diverse, and contained large numbers of individuals with and without major depressive disorder (MDD). To meet this goal, we used a subset of Cole et al.'s (2011) composite data set, consisting

of KSADS depression data on children and adolescents provided by eight clinical research groups in the United States and Great Britain. For the current study, this subsample contained data from community samples, high-risk samples, and clinical treatment samples, so that collectively they represented all levels of depression severity (47.2% met criterion for MDD). The sample was diverse with regard to gender, age, and ethnicity. This method represents an example of what Curran and Hussong (2009) call integrative data analysis.

Second, to resolve the part-whole problem, we adopted a latent-variable approach in which each symptom is compared with the underlying dimension (or factor)<sup>1</sup> rather than with a manifest variable algorithmic combination of observed symptoms. More specifically, we conceptualized symptoms of depression as indicators of one or more underlying dimensions. On the one hand, if all depressive symptoms represent a single underlying dimension of psychopathology, we would expect them all to load positively onto a single, general depression dimension. On the other hand, if weight gain and appetite increase are not good indicators of depression in children and adolescents, we would not expect KSADS items that represent these symptoms to load onto such a general depression dimension. Further, if weight gain and increased appetite were to load onto a second dimension, evidence begins to accrue that these symptoms represent something different from depression in this age group. In this case, important follow-up tests should address specific issues: (a) how highly correlated the two dimensions are; (b) how much the depression dimension changes when weight gain and appetite increase are statistically controlled; (c) how much the relation of a depression dimension to the diagnosis of MDD changes when weight gain and appetite increase are statistically controlled; and (d) whether or not the two dimensions relate differentially to other variables in accordance with theory and previous research on depression.

This last point deserves elaboration because it speaks to the issue of construct validity. Cronbach and Meehl (1955) described validation of a measure as residing, in part, in the empirical support for its correlation with measures of other constructs to which it is theoretically related. Previous theoretical and empirical reports led us to anticipate that measures of depression would reflect the interaction effects of gender and age (Angold & Costello, 2006; Kessler, McGonagle, Swartz, Blazer & Nelson, 1993; Nolen-Hoeksema, 1990; Rutter, 1986). Specifically, we expected that gender difference on depression would be stronger among adolescents than children (with girls evincing higher scores than boys).

## Method

### Data Set Selection

The data set used in the current study represents a subset of the data set described by Cole et al. (2011). For data from the Cole et al. study to be eligible for inclusion in the current study, five criteria were required. First, participants had to be from 4 to 18 years old. Second, the KSADS data must have been collected

<sup>1</sup> We use the terms *dimension* and *factor* interchangeably to refer to the latent variables extracted using either factor analysis or IRT models.

before any treatment or preventive intervention. Third, the study must have included participants for whom KSADS screening items (e.g., depression, irritability, anhedonia, and, in some cases, suicide) were not used to skip the majority of depression questions. Fourth, participants could not have missing data for either gender or age. Fifth, samples from each contributing dataset had to include sufficient response variation across the variables of interest. Inclusion criteria 3 and 5 reduced the number of contributing studies relative to the Cole et al. (2011) study. Eight studies met these criteria. Of the 2576 participants in these studies, 2307 (90%) met all criteria for inclusion. Excluded participants did not differ from included participants on any study variable. Before data acquisition, we obtained institutional review board approval, arranged for complete de-identification of data sets, made explicit the limitations on our use of the data, conferred with the principal investigator (PI) and other study collaborators to ensure that no conflicts of interest existed between our research agenda and those of the original investigator(s), discussed authorship, and obtained signed letters of agreement from the PI or co-PI of each project.

We refer to studies by the name of the investigator who was our key collaborator on this project. Contributors included the following: Compas and Forehand (Compas et al., 2009; 2010), Curry (The TADS Team, 2003, 2005), Findling (Findling et al., 2005), Goodyer (Goodyer et al., 2007, 2008), Hyde and Essex (Essex et al., 2006; Essex et al., 2009; Grabe, Hyde, & Lindberg, 2007; Mezulis, Priess, & Hyde, 2010; Priess, Lindberg, & Hyde, 2009), Rohde (Kaufman, Rohde, Seeley, Clarke, & Stice, 2005; Rohde, Clarke, Mace, Jorgensen, & Seeley, 2004; Rohde, Seeley, Kaufman, Clarke, & Stice, 2006), Weissman (Pilowsky et al., 2008; Weissman et al., 2006), and Youngstrom (Youngstrom et al., 2005). Key characteristics of the sample appear in Table 1. In total, 1088 (47.2%) met *DSM-IV* MDD criteria. For the majority of these individuals, the current episode was their first.

## Measures

The contributing studies used slightly different versions of the KSADS: the KSADS–Present and Lifetime Version (KSADS–PL; J. Kaufman et al., 1997; J. Kaufman, Birmaher, Brent, Rao, & Ryan, 1996), the KSADS–Epidemiological Version (K–SADS–E; Orvaschel, 1994), and the Washington University KSADS (KSADS–WASHU; Geller, Zimmerman, Williams, et al., 2001). As in Cole et al. (2011), when the lifetime KSADS was used to assess multiple episodes of major depression, we used either the current or most recent episode. All versions of the KSADS contain example questions for interviewers to use with children about their own symptoms and with parents about their child's symptoms. Although example questions differ slightly from version to version, no version requires that interviewers adhere precisely to the questions that are listed. All versions instruct interviewers to inquire about symptoms in ways that the participants can best understand.<sup>2</sup> In the current study, KSADS symptoms were scored such that 1 = not present, 2 = present at a subclinical level, and 3 = present at a clinical level. Interrater reliabilities across studies ranged from 0.71 to 0.91 (median = 0.82).

## Missing Data

Some missing data occurred in the composite data set. Of the total 2307 cases, 78.6% had no missing data. On average, participants had

Table 1  
Sample Characteristics (n = 2307)

Variable	Mean (SD)	Percent
Demographics		
Gender		
Female	—	47.3
Male	—	52.7
Age	12.5 (3.2)	
4 to 8	—	13.7
9 to 10	—	12.75
11 to 12	—	15.95
13 to 14	—	25.53
15 to 16	—	24.44
17 to 18	—	7.63
Race/ethnicity		
Black	—	30.5
Caucasian	—	58.8
Hispanic	—	3.6
Other or unknown	—	7.2
Population		
High-risk	—	37.9
Targeted	—	44.7
Community	—	17.4
Meeting criteria for MDD		47.3
MDD symptoms (on 3-point KSADS scale)		
Affective disturbance	2.33 (0.85)	—
Depressed mood	1.97 (0.87)	—
Irritability	1.96 (0.88)	—
Anhedonia	1.93 (0.93)	—
Weight/appetite disturbance	1.77 (0.91)	—
Appetite decrease	1.46 (0.78)	—
Weight loss	1.17 (0.52)	—
Appetite increase	1.29 (0.66)	—
Weight gain	1.21 (0.58)	—
Sleep disturbance	2.04 (0.94)	—
Initial insomnia	1.62 (0.84)	—
Middle insomnia	1.38 (0.72)	—
Terminal insomnia	1.23 (0.60)	—
Hypersomnia	1.41 (0.76)	—
Psychomotor disturbance	1.93 (0.90)	—
Agitation	1.56 (0.77)	—
Retardation	1.66 (0.83)	—
Fatigue, low energy, tired	1.97 (0.94)	—
Worthlessness/guilt	2.08 (0.91)	—
Worthlessness	1.90 (0.89)	—
Guilt	1.45 (0.75)	—
Cognitive disturbance	2.06 (0.94)	—
Concentration	1.82 (0.91)	—
Indecision	1.56 (0.82)	—
Suicide	1.67 (0.84)	—

*Note.* MDD = major depressive disorder. For high-risk (selected for some risk factor but not the disorder), targeted (selected for the disorder), and community (from general community but not selected for risk factor or disorder), numbers are approximations because some studies used multiple recruitment methods. Because of their low base rates, we did not examine the symptoms subsumed under Suicide in the DSM (i.e., recurrent thoughts of death, suicidal ideation, suicidal acts, and suicide plan).

missing values for three of the 23 variables used in the current study. No significant psychometric differences distinguished participants

<sup>2</sup> Lower-order suicide symptoms (i.e., recurrent thoughts of death, suicidal ideation, suicidal acts, suicide plan) were not included in the analysis because of their low response frequencies.



with and without missing data. Furthermore, neither the pattern of missingness nor the number of missing values was significantly related to the scores on any variable in the study. Consequently, we included all 2307 participants in the analyses.

## Results

### Number of Dimensions

Of the methods available for determining the number of dimensions to extract in factor analysis, parallel analysis is generally the most accurate (Horn, 1965; Humphreys & Montanelli, 1975; Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986). Parallel analysis is a Monte-Carlo-based simulation method that compares observed eigenvalues with those obtained from uncorrelated normal variables. A dimension is retained if an eigenvalue from the observed data is larger than the corresponding value from the random data. Because the current data contained ordinal responses, we conducted parallel analysis with polychoric correlations (Cho, Li, & Bandalos, 2009). The results depicted in Figure 1 support extraction of two dimensions.

Using *Mplus*, version 5.21 (Muthén & Muthén, 1998–2006), we conducted a series of exploratory factor analyses using polychoric correlations (specifically, limited information robust weighted least square estimation with Oblimin rotation and Oblique type), extracting 1, 2, 3, and 4 factors. In Table 2, we compare these models using standardized root-mean-square residual (SRMR), root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis index (TLI). According to empirically supported guidelines, a model fits well if the SRMR is smaller than .08, the RMSEA is smaller

than .06, and the CFI and TLI are larger than .95 (Hu & Bentler, 1999; Yu, 2002). The unidimensional model did not provide a good fit to the data according to these criteria. All multidimensional models fit the data reasonably well. We then examined residual variances (and factor correlations) for the 1-, 2-, 3-, and 4-factor solutions. Shifting from the 1- to 2-factor model produced noteworthy reductions in residual variances for four items, all pertaining to weight and appetite (in boldface). Furthermore, factors 1 and 2 correlated only 0.19 ( $SE = 0.02$ ) with each other. Shifting to a 3- or 4-factor model produced no large reductions in residual variances (none larger than .20). Furthermore, larger factor correlations emerged (some  $> .70$ ). Taken together, these results suggest that a bidimensional model provided the most parsimonious yet statistically compelling fit.

### Study Comparability

Previous research with this data set supports the basic psychometric equivalence of measure across the contributing studies (Cole et al., 2011). In the current study, we further examined comparability across three subgroups that used slightly different versions of the KSADS (KSADS-PL, KSADS-E, and WASH-U). First, we conducted parallel analysis on polychoric correlations for each group. In all three groups, these analyses clearly supported the extraction of two dimensions. We then conducted differential item functioning (DIF) analysis across subgroups, using the multidimensional version of Samejima's (1969) graded response model (Muraki & Carlson, 1995). We chose this method in part because of our evidence of multidimensionality and in part because it later serves as our main data analytic method. We used a compensatory multidimensional model in which each item was allowed to load onto two dimensions. We chose an exploratory over a confirmatory approach because misspecification of zero discriminations in a confirmatory approach would bias estimates of item characteristics and correlations between dimensions (Asparouhov & Muthén, 2009; Browne, 2001).

This model was estimated using limited information robust weighted least square estimation (WLSM in *Mplus*) and Geomin rotation.<sup>3</sup> No estimation problems occurred, and the model fit the data well in each of the three groups. We then conducted three two-group comparisons, using two DIF detection procedures based on WLSM estimates: (1) scaled  $\Delta\chi^2$  test (Satorra & Bentler, 2001) of the different measurement models (Asparouhov & Muthén, 2009) and (2) differential functioning of items and tests (DFIT) for the multidimensional model (Oshima, Raju, & Flowers, 1997). Under a DFIT framework, the noncompensatory DIF index (NCDIF) is an item level index similar to Raju's (1988) unsigned area index (Raju, van der Linden, & Fleer, 1995). DIF testing based on the chi-square statistic is known to be highly sensitive to sample size (Kim, Cohen, Alagoz, & Kim, 2007). When the

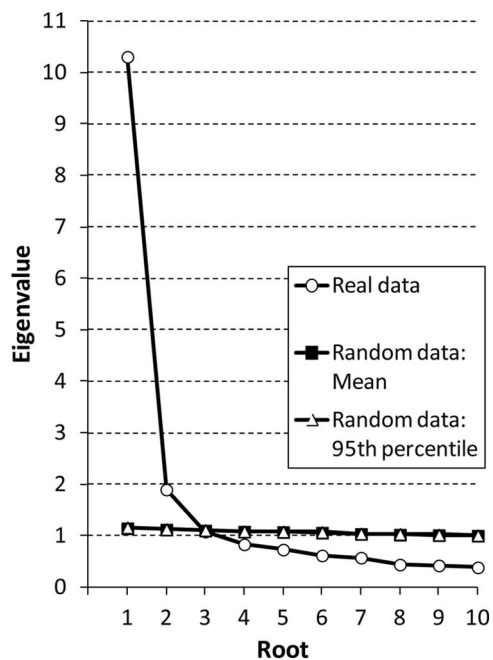


Figure 1. Eigenvalues from polychoric-based parallel analysis for the real data, mean eigenvalues from random data, and 95th percentile of eigenvalues from random data.

<sup>3</sup> Limited information and full information estimation produced similar parameter estimates and standard errors in models without group invariance constraints. To test equivalence across groups, however, cross-group constraints must be imposed, limiting the use of full information estimation. Therefore, limited information estimation was used for the DIF analyses. Also, we tried Quartimin, Geomin, and Target rotation methods and found that the patterns of item discriminations were similar across all methods. We report the results of Geomin rotation here.

Table 2  
Fit Indices and Residual Variances From Exploratory Factor Analyses Extracting 1, 2, 3, and 4 Factors

Index/variable	1-factor	2-factor	3-factor	4-factor
Goodness-of-fit indices				
SRMS	0.102	0.053	0.041	0.030
RMSEA	0.095	0.057	0.047	0.036
CFI	0.907	0.967	0.979	0.989
TLI	0.940	0.986	0.990	0.994
Residual variances				
Depressed mood	0.118	0.115	0.113	0.107
Irritability	0.286	0.282	0.281	0.278
Anhedonia	0.231	0.226	0.208	0.202
Appetite decrease	<b>0.506</b>	<b>0.127</b>	0.110	0.084
Weight loss	<b>0.602</b>	<b>0.379</b>	0.378	0.379
Appetite increase	<b>0.591</b>	<b>0.101</b>	0.080	0.085
Weight gain	<b>0.632</b>	<b>0.222</b>	0.224	0.214
Initial insomnia	0.534	0.527	0.402	0.377
Middle insomnia	0.523	0.495	0.319	0.207
Terminal insomnia	0.627	0.605	0.466	0.450
Hypersomnia	0.553	0.536	0.508	0.509
Agitation	0.536	0.531	0.529	0.529
Retardation	0.458	0.452	0.441	0.438
Fatigue	0.221	0.213	0.195	0.175
Worthlessness	0.346	0.341	0.338	0.297
Guilt	0.500	0.494	0.486	0.458
Concentration	0.365	0.360	0.331	0.303
Indecision	0.464	0.455	0.377	0.354
Suicide	0.578	0.573	0.537	0.332

Note. Bold type indicates items for which the shift from 1 to 2 factors produced large reductions in residual variance.

sample size is large, statistical significance can emerge even when DIF is actually quite small. Typically, examination of DIF effect sizes addresses this concern. In such cases, Raju et al. (1995) recommended using cutoff values. Although statistically significant DIF was detected using our two procedures, the NCDIF values for the DIF items are less than 0.05, indicating that the magnitude of DIF is not large (Bolt, 2002; Flowers, Oshima, & Raju, 1999). Given these results, we concluded that DIF was negligible and that the subgroups of data could be combined.

### IRT Analyses: Exploratory Multidimensional Graded Response Model

Item parameter estimates and *SEs* for an exploratory bidimensional graded response model using the aggregated dataset are listed in Table 3. Most important for our purposes are the item discriminations, which can be interpreted as (and transformed into) factor loadings for categorical responses. The scale for these estimates was standardized in probit values. All but two of the 19 items had large discriminations relative to Dimension 1. Because all items represented symptoms of MDD, we regarded Dimension 1 as general Depression. Weight loss and appetite decrease were among these 17 symptoms. The two remaining items, appetite increase and weight gain, had large discriminations on Dimension 2. Weight loss and decreased appetite also contributed to Dimension 2; however, their discriminations were in the opposite direction relative to weight gain and increased appetite. Because discriminations of all other variables relative to this dimension were relatively small, we regarded Dimension 2 as a specific Weight-

gain/Increased-appetite dimension. Item thresholds are also presented in Table 3, which represent symptom severity information for each item on the sum of Dimensions 1 and 2 (because a compensatory multidimensional structure was used). Threshold 1 reflects the transition points from a score of 1 (symptom is absent) to a score of 2 or 3 (symptom is present at a subclinical or clinical level) on a given item. Threshold 2 reflects the transition points from a score of 1 or 2 to a score of 3 on a given item.

Figure 2 depicts Test Information Functions (TIFs) for both dimensions. These curves show how much information derives from each dimension, holding the other dimension constant at its mean. A higher curve implies greater measurement fidelity. A high curve that is also wide implies good measurement fidelity across a wide range on the underlying dimension. Clearly, the TIF for Dimension 1 is substantially elevated relative to the TIF for Dimension 2. Much less information can be expected from Dimension 2 compared with Dimension 1 (in part because so few items made strong contributions to this dimension).

### Interpretation and Utility of the Dimensions

One practical question is whether the assessment of youths on a general Depression dimension is seriously affected when we also extract (and thereby control for) the Increased Weight/Appetite dimension. First, we compared scores on Dimension 1 as derived from a unidimensional model to scores on Dimension 1 as derived from a bidimensional model. The regression of one set of scores on the other revealed a correlation of 0.999. Examination of the residuals revealed that more than 98% of the residuals were within 0.5 units of their predicted values. These results suggest that we would make very similar inferences about level of general depression without information from Dimension 2.

A second way to address this question is to compare scores on the extracted dimensions to the presence or absence of MDD. Figure 3 shows the relation of MDD to five different levels of the various extracted dimensions (D): Level 1 ( $D < -1$ ), Level 2 ( $-1 \leq D < 0$ ), Level 3 ( $0 \leq D < 1$ ), Level 4 ( $1 \leq D < 2$ ), and Level 5 ( $2 \leq D$ ). The diagnostic frequencies at different levels of Dimension 1 derived from a unidimensional model are very similar to those for Dimension 1 derived from a bidimensional model (compare the first and second set of bars in Figure 3). Clearly, the inclusion of information about weight gain and increased appetite did not enhance the diagnostic utility of a general Depression dimension. Reasons for this finding become evident when examining the relation of Dimension 2 to MDD (right side of Figure 3), where the correspondence with MDD is much weaker.

### Gender and Age Differences

Another study goal was to examine Age and Gender differences on the two extracted dimensions. Group mean differences are not meaningful, however, if DIF exists across the groups. Therefore, we began by testing gender- and age-related DIF, using the same two procedures described above, the  $\Delta\chi^2$  comparison of different measurement models and DFIT for the multidimensional model.

**Gender DIF.** We began by comparing three nested models. First, we fit a configural invariance model in which all item

Table 3  
Item Parameter Estimates (and Standard Errors) for Bidimensional Graded Response Models

Items	Discriminations		Thresholds	
	Dimension 1 Est. (SE)	Dimension 2 Est. (SE)	Threshold 1 (1 vs. 2 or 3) Est. (SE)	Threshold 2 (1 or 2 vs. 3) Est. (SE)
Depressed mood	<b>0.907</b> (0.011)	0.004 (0.016)	-0.273 (0.029)	0.348 (0.030)
Irritability	<b>0.794</b> (0.017)	0.015 (0.028)	-0.226 (0.029)	0.323 (0.030)
Anhedonia	<b>0.824</b> (0.016)	0.035 (0.030)	-0.081 (0.026)	0.252 (0.026)
Appetite decrease	<b>0.860</b> (0.078)	-0.649 (0.030)	0.597 (0.031)	0.904 (0.034)
Weight loss	<b>0.741</b> (0.077)	-0.606 (0.034)	1.246 (0.039)	1.531 (0.046)
Appetite increase	0.229 (0.118)	<b>0.832</b> (0.054)	0.905 (0.034)	1.218 (0.038)
Weight gain	0.136 (0.123)	<b>0.841</b> (0.056)	1.171 (0.038)	1.377 (0.042)
Initial insomnia	<b>0.689</b> (0.020)	-0.073 (0.034)	0.291 (0.029)	0.718 (0.032)
Middle insomnia	<b>0.723</b> (0.027)	-0.179 (0.036)	0.717 (0.032)	1.070 (0.036)
Terminal insomnia	<b>0.638</b> (0.031)	-0.158 (0.044)	1.080 (0.036)	1.337 (0.041)
Hypersomnia	<b>0.594</b> (0.029)	0.106 (0.041)	0.692 (0.032)	0.963 (0.035)
Agitation	<b>0.653</b> (0.021)	-0.013 (0.033)	0.297 (0.030)	0.935 (0.034)
Retardation	<b>0.670</b> (0.021)	0.018 (0.033)	0.186 (0.029)	0.717 (0.032)
Fatigue	<b>0.826</b> (0.017)	0.059 (0.029)	-0.111 (0.026)	0.180 (0.026)
Worthlessness	<b>0.735</b> (0.020)	0.034 (0.032)	-0.118 (0.029)	0.388 (0.030)
Guilt	<b>0.660</b> (0.022)	-0.034 (0.036)	0.535 (0.031)	1.014 (0.035)
Concentration	<b>0.847</b> (0.014)	0.005 (0.021)	0.043 (0.029)	0.409 (0.030)
Indecision	<b>0.729</b> (0.021)	0.059 (0.033)	0.381 (0.030)	0.807 (0.033)
Suicide	<b>0.581</b> (0.023)	-0.028 (0.035)	0.181 (0.026)	0.705 (0.029)

Note. Large discriminations appear in bold type.

parameters were estimated simultaneously in each gender group. Factor means were fixed to 0, factor variances were fixed to 1, and residual variances were constrained to 1 in both groups to identify the model. This model fit the data well (CFI = 0.988, TLI = 0.985, and RMSEA = 0.057). Second, we fit a weak invariance model, in which the factor means were

fixed to 0, factor variances were fixed to 1 for males but were freely estimated for females, factor loadings were equal across groups, all item thresholds were estimated, and all item residual variances were set to 1 across groups. This model also fit the data well (CFI = 0.990, TLI = 0.988, RMSEA = 0.050). Furthermore, the weak invariance model did not fit significantly worse than the configural invariance model (scaled  $\chi^2_{(df=34, n=2307)} = 2.627$  with a  $p$  value of approximately 1.000). In addition, the local model fit did not suggest removing any loading constraints across groups. Third, we fit a strong invariance model, which was identical to the weak invariance model except that all item thresholds were constrained to be equal across groups. This model provided a good fit to the data (CFI = 0.988, TLI = 0.988, RMSEA = 0.051) and did not fit significantly worse than the weak invariance model (scaled  $\chi^2_{(df=36, n=2307)} = 6.337$ ,  $p$  value of approximately 1.000). Collectively, these analyses supported measurement invariance across gender. Finally, we examined gender-related DIF using the DFIT framework. No evidence of DIF existed, even using NCDIF cutoffs of either 0.05 or 0.009.

**Age DIF.** We constructed two age groups for DIF analysis (younger: age <12 years; older: age  $\geq$ 12 years).<sup>4</sup> We treated Group 1 as the reference group. The same three measurement models were used as described in the Gender DIF section (above).

<sup>4</sup> Typically we would not dichotomize a continuous variable like age. Unfortunately, statistical methods for examining DIF as a function of a continuous variable have not been developed. We tried several cutoffs, with very similar results. We concentrate here on results for the 12-year-old cutoff because it creates balanced sample sizes (therefore maximizing power) and represents a reasonable separation of children versus adolescents. Use of lower age cutoffs generated similar results, but the smaller  $n$  for the younger sample created relatively large standard errors.

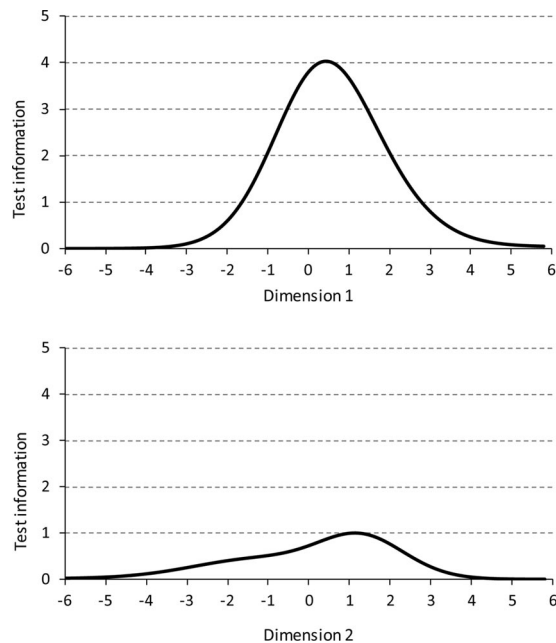


Figure 2. Test information function for Dimensions 1 and 2, holding the other dimension constant at its mean.

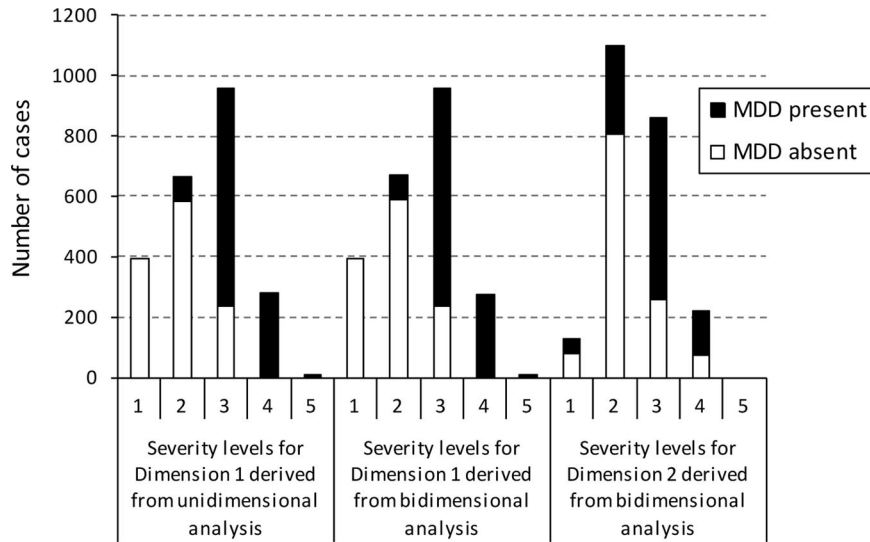


Figure 3. Relation of severity levels on dimensions derived from uni- and bidimensional IRT Graded Response Models to the presence/absence of an MDD diagnosis.

The configural invariance model fit well (CFI = 0.987, TLI = 0.983, RMSEA = 0.059). Similarly, the weak invariance model fit the data well (CFI = 0.988, TLI = 0.986, RMSEA = 0.053) and was not significantly different from the configural model (scaled  $\chi^2_{(df=34, n=2307)} = 3.584$ ,  $p$  value of approximately 1.000). The strong invariance model also provided a good fit (CFI = 0.933, TLI = 0.983, RMSEA = 0.059) and was not significantly different from the weak invariance model (scaled  $\chi^2_{(df=36, n=2307)} = 20.350$ ,  $p$  value < 0.983). No evidence of DIF emerged based on NCDIF with a cutoff of 0.05 or 0.028. These tests showed measurement invariance across the two age groups.

**Mean differences across gender and age groups on the two dimensions.** We tested mean differences on the two dimensions across gender and age using the explanatory multidimensional graded response model (De Boeck & Wilson, 2004) using WLSMV in *Mplus*. We used effect-coding of Gender (male = -1, female = 1) and age (Group 1 = -1 and Group 2 = 1). On Dimension 1, tests were significant for the Gender main effect,  $t = 6.18$ ,  $p < .001$ , Age main effect,  $t = 14.33$ ,  $p < .001$ , and Gender by Age interaction,  $t = 2.70$ ,  $p < .007$ . We followed the interaction with four pairwise comparisons ( $\alpha = .05/4$ ). As depicted in the upper graph in Figure 4, significant differences emerged for younger boys versus younger girls, older boys versus older girls, younger girls versus older girls, and younger boys versus older boys ( $ps < .001$ ). On Dimension 2, only the Gender main effect was significant,  $t = 3.90$ ,  $p < .001$ . As shown in the lower graph in Figure 4, girls had higher scores than boys.

We also examined the effects of Age and Gender on MDD. In a logistic regression, the interaction effect of Age and Gender on MDD was not significant. The odds ratio was 1.41, with a 95% confidence interval (CI) of 0.90 to 2.21; however, given the strong a priori literature support for the existence of gender differences in adolescents but not children, we then conducted planned pairwise comparisons of boys versus girls on MDD for younger and older participants. Among younger participants, the odds of having MDD for females were 1.26 times the odds for males; however,

the 95% confidence interval (CI = 0.84 – 1.88) contained 1.0, and the likelihood ratio chi-square test of association between Gender and MDD was not significant:  $\chi^2_{(df=1, n=750)} = 1.26$ ,  $p < .261$ . Among older participants, the odds of having MDD for females were 1.77 times those for males, with a 95% CI of 1.44 – 2.18

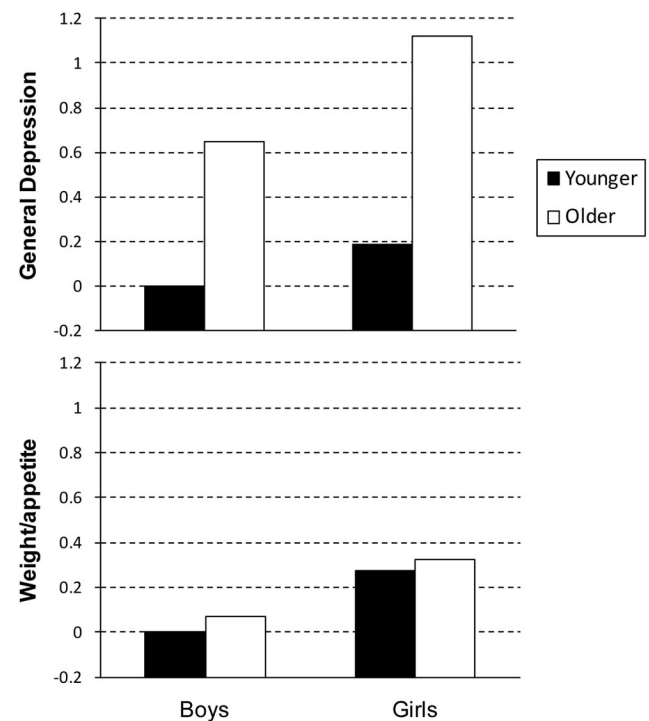


Figure 4. Group means for younger (<12) and older ( $\geq 12$ ) boys and girls on Dimension 1 (General Depression) and Dimension 2 (Weight-gain/increased-appetite). Note: Younger boys were regarded as the reference group and their means were set equal to zero.



and a significant likelihood ratio chi-square test of association:  $\chi^2_{(df=1, n=1557)} = 29.75, p < .001$ . Using *DSM-IV* diagnostic standards for depression, the gender difference in MDD was significant for older but not younger participants.

### Discussion

Three major results emerged from this study that, in concert with prior investigations, suggest that increased appetite and weight gain should not be treated as indicators of depression in children or adolescents (although weight loss and appetite decrease should). First, multidimensional factor analysis and IRT analysis of 19 depression symptoms, measured by the KSADS, revealed that weight *loss* and appetite *decrease* loaded onto a general depression factor, but weight *gain* and appetite *increase* did not; instead, these latter two symptoms represented a separate factor, which related weakly to the general depression factor. Second, excluding weight gain and increased appetite had virtually no effect on the nature of the general depression factor, nor did it affect the relation of the general depression factor to MDD. And third, different patterns of age and gender differences emerged for the general depression factor versus the specific weight gain/increased appetite factor. We elaborate on each of these findings and their implications below.

Our first major result was that, in children and adolescents, two distinct dimensions (not just one) emerged from a set of 19 symptoms implicated by *DSM-IV* as indicators of major depression. The first dimension was characterized by 17 of these 19 symptoms, clearly identifying it as a general depression dimension. Decreases in weight and appetite contributed to this dimension; however, weight gain and increased appetite did not. Instead, weight gain and increased appetite contributed so strongly to the second dimension as to identify it as a specific weight-gain/increased-appetite dimension. At first blush, one might be tempted to interpret this second dimension as evidence that depression is bidimensional; however, at least two findings argue against this interpretation. One is that very few other symptoms (and none of the core depressive symptoms, such as dysphoria and anhedonia) contributed to the second dimension. Second, the weight-gain/increased-appetite dimension was only weakly correlated ( $r = .30$ ) with the general depression dimension. To put this in perspective, the *latent* weight-gain/increased-appetite dimension correlated far less with a general depression dimension than did any of the other *manifest* variable indicators of depression.

This finding reinforces Yorbik et al.'s (2004) similar discovery of a separate weight/appetite dimension in depressed children and adolescents; however, our interpretation differs from theirs. They interpreted this factor to be a component or dimension of depression; however, given aspects of their data analytic method, this interpretation is difficult to justify. For example, they used a factor extraction method that yields orthogonal factors, preventing the examination of correlations between their dimensions. A large correlation with their "endogenous depression" factor would support their interpretation that weight/appetite is a component of depression. A small correlation (as we found), however, motivates the interpretation that weight/appetite is tangential to the symptomatology of depression. One could also use strong cross-loadings of core depressive symptoms of depression onto the weight/appetite factor as evidence that the second factor is an

aspect of depression. Such cross-loadings are not reported in Yorbik et al.'s paper, probably indicating that they were less than some (unspecified) criterion. Without evidence that a weight/appetite factor is related to something depressive, we find it difficult to regard it as an aspect of depression.

Our second major finding was that the exclusion of weight gain and increased appetite had no effect on the nature of the general depression factor and did not affect its relation to MDD. Four specific findings support these claims: (1) the correlation between the general depression factor scores with and without weight gain and appetite increase was 0.999; (2) inclusion of information about weight gain and appetite increase contributed almost nothing to the proper classification of participants as having MDD or not; (3) how well a general depression factor corresponded with MDD was essentially unaffected by the inclusion or exclusion of weight gain and increased appetite; and (4) the correspondence of a weight-gain/increased-appetite dimension to MDD was quite weak. Taken together, these results suggest that little or no practical value derives from using weight gain and increased appetite as indicators in the diagnosis of child and adolescent depression.

In fact, we would take this one step further. In most clinical situations, diagnosticians do not have access to physical measures of weight gain. Instead they rely on retrospective self- or parent reports of weight change, despite the fact that such information is known to be inaccurate and is often affected by mood, memory bias, and/or body dissatisfaction (Felton et al., 2010; Vartanian, Herman & Polivy, 2004). Furthermore, most diagnosticians lack the technical resources to take age, gender, and racial normative weight gain into consideration when assessing the degree to which weight gain might be due to depression. Taken together, these difficulties can generate considerable inaccuracy in the assessment of this symptom and consequently in the diagnosis of depression.

Third, comparing children with adolescents and comparing males with females revealed a different pattern of mean differences for the general depression dimension versus the specific weight-gain/increased-appetite dimension. For general depression, the expected Age by Gender interaction emerged, showing that depression increased with age for both boys and girls but more so for girls. This finding reflects a classic pattern for depression (see Angold & Costello, 2006; Kessler et al., 1993; Nolen-Hoeksema, 1990; Rutter, 1986, for reviews). For the weight-gain/increased-appetite dimension, a very different pattern of mean differences emerged in which girls had higher levels than boys but neither the Age main effect nor the Age  $\times$  Gender interaction was significant. The fact that the pattern of group means was different for the two dimensions suggests that the weight-gain/increased-appetite dimension is embedded in a qualitatively different nomological net than is the general depression dimension (Cronbach & Meehl, 1955). Indeed, evidence of discriminant validity derives, in part, out of the discovery that two dimensions relate differentially to a common set of variables. This finding further reinforces our conclusion that the two dimensions represent different constructs, one of which has little to do with depression in child and adolescent populations.

A closer look at the Age by Gender effect on depression reveals unexpected insights into two somewhat controversial issues raised in previous research. One is that evidence of an Age by Gender interaction has been stronger for categorical/diagnosis-based measures of depression than for continuous/severity-based measures:

compare Compas et al. (1997) with Hankin et al. (1998), leading to implications about the superiority of categorical/diagnostic operationalizations of depression (Hankin et al., 1998). Interestingly, most studies of dimensional depression rely on paper-and-pencil inventories, whereas most studies of categorical depression rely on clinical interview data. In the current study, our measure of both *dimensional* and *categorical* depression derived from KSADS *clinical interviews*. Our results revealed the opposite pattern: the Age by Gender effect on our continuous general depression dimension was significant, whereas the interaction effect on our categorical index of MDD diagnosis was not. Statistically, interactions are notoriously difficult to detect.<sup>5</sup> Cautiously, we suggest that detection of a significant Age by Gender effect on depression hinges on the reliability and validity of the depression measure, and that (a) a continuous measure of depression derived from clinical interviews with children and parents contains more information than (b) a dichotomous diagnosis of depression based on the same information, which contains more information than (c) a depression inventory administered only to children. (Similar points derived from the TADS project; March et al., 2004, 2007.)

The second controversial issue pertains to whether a gender difference in depression exists in preadolescent children. Most studies have suggested that the gender difference in depression does not exist until adolescence (see Angold & Costello, 2006; Kessler et al., 1993; Nolen-Hoeksema, 1990; Rutter, 1986). A few studies have suggested that it does (Kazdin, French & Unis, 1983; Liss, Phares, Liljequist, 2001).<sup>6</sup> In the current study, we found no evidence of a gender effect on depression in younger participants when depression was represented by presence/absence of MDD. Conversely, we found a significant gender difference among younger participants (with girls > boys) when depression was represented by our continuous general depression dimension. Again, we speculate that the discrepancy between these findings is attributable to the fact that more information is lost than gained when shifting from our dimensional representation of depression to our categorical/diagnostic representation of depression (Haslam & Beck, 1994; Ruscio, Zimmerman, McGlinchey, Chelminski, & Young, 2007).

Despite the strengths of the current study (a relatively large and diverse sample, the use of KSADS interviews, a high proportion of clinically depressed youths, and utilization of multigroup and multidimensional factor analytic and IRT data analytic methods), various shortcomings also exist that suggest avenues for continued research. First, the current article focuses on data assessed by semistructured clinical interview. For most symptoms of psychopathology, clinical interviews are about as close to a gold standard as is possible for diagnosticians; however, the symptom of weight-change is the exception. Self-reports of weight (let alone retrospective self-reports of weight change) are known to be biased (Goodman, Hinden & Khandelwal, 2000; Himes & Faricy, 2001; Vartanian et al., 2004; Wang, Patterson & Hills, 2002). Despite this fact, only one study has attempted to estimate the relation between physical measures of weight change and depression symptoms in children or adolescents (Felton et al., 2010), and it was based on a community sample of relatively nondepressed youths. Despite the current result and all the physiological, psychological, and behavioral arguments presented in our introduction, research is still needed on the relation between actual weight gain and major depression in children and adolescents.

Second, given the paucity of factor analytic and IRT work on the dimensionality of clinical-interview measures of depressive symptoms in young people, we elected to take an exploratory, not a confirmatory, data analytic approach. The fact that the results of our exploratory approaches supported our hypothesis represents a kind of confirmation; however, more work is needed. The current results pave the way for a confirmatory factor analytic or IRT work. That said, a word of caution is in order. For a variety of reasons, problems with confirmatory analyses can produce results that appear not to “confirm” exploratory results, especially when individual items represent the units of analysis (Asparouhov & Muthén, 2009; Floyd & Widaman, 1995; Reise, Waller & Comray, 2000; van Prooijen & Van der Kloot, 2001). Applications of confirmatory analytic methods to these questions must avoid these pitfalls.

Third, a factor that could complicate interpretation of our results is the use of psychotropic medication by some study participants. Some psychotropic medications can affect physical growth in children and adolescents, with one of the stronger effects being the association of amphetamines with slowed physical growth (Correll & Carlson, 2006; Faraone, Biederman, Morley, & Spencer, 2008). Despite the many advantages of our aggregate data set, a disadvantage was the diversity of approaches to collecting information about current medications. For some of the contributing studies, participants were excluded if they were on psychotropic medication. In other studies, psychotropic medications were allowed and medication data were systematically gathered. In still other studies, data on psychotropic use were simply not available. Taken together, this methodological diversity prevented us from testing or controlling systematically for psychotropic use by some of the study participants. Better measures of medication use would be important in future research.

Fourth, among our participants who had clinically significant depression, an unknown subset may well have gone on to develop bipolar disorder (Angst, Felder, Frey, & Stassen, 1978; Angst, Gamma, Sellaro, Lavori, & Zhang, 2003; Beesdo et al., 2009; Fiedorowicz et al., 2011). Some research suggests that such cases are more likely to have the “atypical” presentation: hypersomnia, increased appetite, interpersonal sensitivity, leaden paralysis, and so forth (Akiskal & Benazzi, 2005; Angst, Gamma, Sellaro, Zhang, & Merikangas, 2002; Benazzi, 2005). These cases could be responsible for the migration of increased appetite and weight gain (atypical features) into a separate factor. Although future studies should examine this possibility, our analyses revealed no clustering of weight gain and appetite increase with symptoms such as hypersomnia or psychomotor retardation (data on leaden paralysis and interpersonal sensitivity were not available), as one might expect if atypical depression were responsible for our results.

Finally, various technical caveats are in order. First, one must guard against the overinterpretation of our weight gain/appetite increase dimension. We are confident that it does not represent

<sup>5</sup> For a variety of reasons (including reduced power), interaction effects are often difficult to detect. Increasing measurement reliability and validity can diminish this problem (e.g., McClelland & Judd, 1993).

<sup>6</sup> Two studies have even suggested the opposite—that before adolescence, boys are more depressed than girls (Anderson, Williams, McGee, & Silva, 1987; Angold, Costello, & Worthman, 1998).

depression, but as the dimension is anchored by only two symptoms assessed by the KSADS interview and not by any physical measures, we caution against more elaborate interpretations. Second, in exploratory analyses such as ours results can vary depending upon the choice of method. Third, as no practical guidelines yet exist for the interpretation of DIF effect sizes in multidimensional IRT, we used cutoff values recommended for DFIT procedures. Fourth, we examined results from both a unidimensional and a bidimensional graded response model even though the unidimensional model did not fit the data well. In circumstances like ours, however, where one dimension is dominant and other dimensions are minor, unidimensional parameter estimates and latent variable scores are little affected by the other dimensions (Ansley & Forsyth, 1985; Dragow & Parsons, 1983; Reckase, 1979; Way, Ansley, & Forsyth, 1988).

## Conclusion

In conclusion, results of the current study (with support from prior investigations) suggest that increased appetite and weight gain should not be treated as indicators of depression in children and adolescents. We speculate that during developmental periods of rapid physical growth and psychological change, weight gain and increased appetite are under so many other physiological and psychological controls as to reduce greatly the sensitivity of these variables as indicators of depression. Importantly, this conclusion does not extend to the opposites of these symptoms, weight loss and decreased appetite, which do seem to be valid indicators of depression during these developmental periods. As such, these results carry major implications for the revision of the DSM. Whether or not weight gain and increased appetite are valid indicators of depression in adults (and some studies suggest they are not: e.g., Zimmerman, McGlinchey, Young, & Chelminski, 2006), our results indicate that they are not valid indicators for children and adolescents. Empirical support for DSM diagnostic criteria will be improved if these findings are taken into account.

## References

- Ahmed, M. L., Ong, K. K. L., Morrell, D. J., Cox, L., Drayer, N., Perry, L., . . . Dunger, D. B. (1999). Longitudinal study of leptin concentrations during puberty: Sex differences and relationship to changes in body composition. *Journal of Clinical Endocrinology and Metabolism*, *84*, 899–905. doi:10.1210/jc.84.3.899
- Akiskal, H. S., & Benazzi, F. (2005). Atypical depression: A variant of bipolar II or a bridge between unipolar and bipolar II? *Journal of Affective Disorders Special Issue: Bipolar Depression: Focus on Phenomenology*, *84*, 209–217.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Anderson, J. C., Williams, S. M., McGee, R., & Silva, P. A. (1987). DSM-III disorders in preadolescent children: Prevalence in a large sample from the general population. *Archives of General Psychiatry*, *44*, 69–76. doi:10.1001/archpsyc.1987.01800130081010
- Angold, A., Costello, E. J., & Worthman, C. M. (1998). Puberty and depression: The roles of age, pubertal status and pubertal timing. *Psychological Medicine: A Journal of Research in Psychiatry and the Allied Sciences*, *28*, 51–61.
- Angold, A., & Costello, E. J. (2006). Puberty and Depression. *Child and Adolescent Psychiatric Clinics of North America*, *15*, 919–937. doi:10.1016/j.chc.2006.05.013
- Angst, J., Felder, W., Frey, R., & Stassen, H. H. (1978). The course of affective disorders: I. change of diagnosis of monopolar, unipolar, and bipolar illness. *Archiv für Psychiatrie und Nervenkrankheiten*, *226*, 57–64. doi:10.1007/BF00344124
- Angst, J., Gamma, A., Sellaro, R., Lavori, P. W., & Zhang, H. (2003). Recurrence of bipolar disorders and major depression: A life-long perspective. *European Archives of Psychiatry and Clinical Neuroscience*, *253*, 236–240. doi:10.1007/s00406-003-0437-2
- Angst, J., Gamma, A., Sellaro, R., Zhang, H., & Merikangas, K. (2002). Toward validation of atypical depression in the community: Results of the Zurich cohort study. *Journal of Affective Disorders*, *72*, 125–138. doi:10.1016/S0165-0327(02)00169-6
- Ansley, T. M., & Forsyth, R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement*, *9*, 37–48. doi:10.1177/014662168500900104
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397–438. doi:10.1080/10705510903008204
- Beesdo, K., Höfler, M., Leibenluft, E., Lieb, R., Bauer, M., & Pfennig, A. (2009). Mood episodes and mood disorders: Patterns of incidence and conversion in the first three decades of life. *Bipolar Disorders*, *11*, 637–649. doi:10.1111/j.1399-5618.2009.00738.x
- Benazzi, F. (2005). Atypical depression and its relation to bipolar spectrum. In A. Marneros & F. K. Goodwin (Eds.), *Bipolar disorders: Mixed states, rapid cycling and atypical forms* (pp. 131–156). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511544019.007
- Bolt, D. M. (2002). A Monte Carlo comparison of parametric and non-parametric polytomous DIF detection methods. *Applied Measurement in Education*, *15*, 113–141. doi:10.1207/S15324818AME1502\_01
- Borchardt, C. M., & Meller, W. H. (1996). Symptoms of affective disorder in pre-adolescent versus adolescent inpatients. *Journal of Adolescence*, *19*, 155–161. doi:10.1006/jado.1996.0015
- Bornstein, S. R., Schuppenies, A., Wong, M. L., & Licinio, J. (2006). Approaching the shared biology of obesity and depression: The stress axis as the locus of gene–environment interactions. *Molecular Psychiatry*, *11*, 892–902. doi:10.1038/sj.mp.4001873
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*, 111–150. doi:10.1207/S15327906MBR3601\_05
- Cho, S.-J., Li, F., & Bandalos, D. L. (2009). Accuracy of the parallel analysis procedure using polychoric correlations. *Educational and Psychological Measurement*, *69*, 748–759. doi:10.1177/0013164409332229
- Cole, D. A., Cai, L., Martin, N. C., Findling, R. L., Youngstrom, E. A., Garber, J., . . . Forehand, R. (2011). Structure and measurement of depression in youths: Applying Item Response Theory to clinical data. *Psychological Assessment*, *23*, 819–833.
- Compas, B. E., Champion, J. E., Forehand, R., Cole, D. A., Reeslund, K. L., Fear, J., . . . Roberts, L. (2010). Coping and parenting: Mediators of 12-month outcomes of a family group cognitive-behavioral preventive intervention with families of depressed parents. *Journal of Consulting and Clinical Psychology*, *78*, 623–634. doi:10.1037/a0020459
- Compas, B. E., Forehand, R., Keller, G., Champion, J. E., Rakow, A., Reeslund, K. L., . . . Cole, D. A. (2009). Randomized controlled trial of a family cognitive-behavioral preventive intervention for children of depressed parents. *Journal of Consulting and Clinical Psychology*, *77*, 1007–1020. doi:10.1037/a0016930
- Compas, B. E., Oppedisano, G., Connor, J. K., Gerhardt, C. A., Hinden, B. R., Achenbach, T. M., & Hammen, C. (1997). Gender differences in depressive symptoms in adolescence: Comparison of national samples of clinically referred and nonreferred youths. *Journal of Consulting and Clinical Psychology*, *65*, 617–626. doi:10.1037/0022-006X.65.4.617
- Correll, C. U., & Carlson, H. E. (2006). Endocrine and metabolic adverse effects of psychotropic medications in children and adolescents. *Journal*



- of the American Academy of Child & Adolescent Psychiatry, 45, 771–791. doi:10.1097/01.chi.0000220851.94392.30
- Cronbach, L. J., & Meehl, P. C. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302. doi:10.1037/h0040957
- Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, 14, 81–100. doi:10.1037/a0015914
- De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York, NY: Springer.
- Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement*, 7, 189–199. doi:10.1177/014662168300700207
- Essex, M. J., Kraemer, H. C., Armstrong, J. M., Boyce, W. T., Goldsmith, H. H., Klein, M. H., . . . Kupfer, D. J. (2006). Exploring risk factors for the emergence of children's mental health problems. *Archives of General Psychiatry*, 63, 1246–1256. doi:10.1001/archpsyc.63.11.1246
- Essex, M. J., Kraemer, H. C., Slattery, M. J., Burk, L. R., Boyce, W. T., Woodward, H. R., . . . Kupfer, D. J. (2009). Screening for childhood mental health problems: Outcomes and early identification. *Journal of Child Psychology and Psychiatry*, 50, 562–570. doi:10.1111/j.1469-7610.2008.02015.x
- Faraone, S. V., Biederman, J., Morley, C. P., & Spencer, T. J. (2008). Effect of stimulants on height and weight: A review of the literature. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47, 994–1009.
- Fehm, H. L., Smolnik, R., Kern, W., McGregor, G. P., Bickel, U., & Born, J. (2001). The melanocortin melanocyte-stimulating hormone/adrenocorticotropin 4–10 decreases body fat in humans. *Journal of Clinical Endocrinology and Metabolism*, 86, 1144–1148. doi:10.1210/jc.86.3.1144
- Felton, J., Cole, D. A., Tilghman-Osborne, C., & Maxwell, M. (2010). The relation of weight change to depressive symptoms in adolescence. *Development and Psychopathology*, 22, 205–216. doi:10.1017/S0954579409990356
- Fiedorowicz, J. G., Endicott, J., Leon, A. C., Solomon, D. A., Keller, M. B., & Coryell, W. H. (2011). Subthreshold hypomanic symptoms in progression from unipolar major depression to bipolar disorder. *The American Journal of Psychiatry*, 168, 40–48. doi:10.1176/appi.ajp.2010.10030328
- Field, A. E., Austin, S. B., Taylor, C. B., Malspeis, S., Rosner, B., Rockett, H. R., . . . Colditz, G. A. (2003). Relation between dieting and weight change among preadolescents and adolescents. *Pediatrics*, 112, 900–906. doi:10.1542/peds.112.4.900
- Findling, R. L., Youngstrom, E. A., McNamara, N. K., Stansbrey, R. J., Demeter, C. A., Bedoya, D., . . . Calabrese, J. R. (2005). Early symptoms of mania and the role of parental risk. *Bipolar Disorders*, 7, 623–634. doi:10.1111/j.1399-5618.2005.00260.x
- Flament, M. F., Cohen, D., Choquet, M., Jeammet, P., & LeDoux, S. (2001). Phenomenology, psychosocial correlates, and treatment seeking in major depression and dysthymia of adolescence. *American Academy of Child & Adolescent Psychiatry*, 40, 1070–1078. doi:10.1097/00004583-200109000-00016
- Flowers, C. P., Oshima, T. C., & Raju, N. S. (1999). A description and demonstration of the polytomous-DFIT framework. *Applied Psychological Measurement*, 23, 309–326. doi:10.1177/01466219922031437
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7, 286–299. doi:10.1037/1040-3590.7.3.286
- Friedman, R. C., Hurt, S. W., Clarkin, J. F., Corn, R., & Aronoff, M. S. (1983). Symptoms of depression among adolescents and young adults. *Journal of Affective Disorders*, 5, 37–43. doi:10.1016/0165-0327(83)90034-4
- Gallup, G., & Castelli, J. (1989). *The people's religion: American faith in the 90's*. NY: Collier Macmillan.
- Geller, B., Zimerman, B., Williams, M., Bolhofner, K., Craney, J. L., Delbello, M. P., & Soutullo, C. (2001). Reliability of the Washington University in St. Louis Kiddie Schedule for Affective Disorders and Schizophrenia (WASH-U-KSADS) mania and rapid cycling sections. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40, 450–455. doi:10.1097/00004583-200104000-00014
- Goldberg, J. H., & King, A. C. (2007). Physical activity and weight management across the lifespan. *Annual Review of Public Health*, 28, 145–170. doi:10.1146/annurev.publhealth.28.021406.144105
- Goodman, E., Hinden, B., & Khandelwal, S. (2000). Accuracy of teen and parental reports of obesity and body mass index. *Pediatrics*, 106, 52–58. doi:10.1542/peds.106.1.52
- Goodyer, I., Dubicka, B., Wilkinson, P., Kelvin, R., Roberts, C., Byford, S., . . . Harrington, R. (2007). Selective serotonin reuptake inhibitors (SSRIs) and routine specialist care with and without cognitive behavior therapy in adolescents with major depression: Randomised controlled trial. *British Medical Journal*, 335, 142. doi:10.1136/bmj.39224.494340.55
- Goodyer, I., Dubicka, B., Wilkinson, P., Kelvin, R., Roberts, C., Byford, S., . . . Harrington, R. (2008). A randomized controlled trial of cognitive behavior therapy in adolescents with major depression treated by selective serotonin reuptake inhibitors. The ADAPT trial. *Health Technology Assessment*, 12, ix–60.
- Grabe, S., Hyde, J., & Lindberg, S. (2007). Body objectification and depression in adolescents: The role of gender, shame, and rumination. *Psychology of Women Quarterly*, 31, 164–175. doi:10.1111/j.1471-6402.2007.00350.x
- Hankin, B. L., Abramson, L. Y., Moffitt, T. E., Silva, P. A., McGee, R., & Angell, K. E. (1998). Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study. *Journal of Abnormal Psychology*, 107, 128–140. doi:10.1037/0021-843X.107.1.128
- Haslam, N., & Beck, A. T. (1994). Subtyping major depression: A taxometric analysis. *Journal of Abnormal Psychology*, 103, 686–692. doi:10.1037/0021-843X.103.4.686
- Himes, J. H., & Faricy, A. (2001). Validity and reliability of self-reported stature and weight of US adolescents. *American Journal of Human Biology*, 13, 255–260. doi:10.1002/1520-6300(200102/03)13:2<255::AID-AJHB1036>3.0.CO;2-E
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179–185. doi:10.1007/BF02289447
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. doi:10.1080/10705519909540118
- Humphreys, L. G., & Montanelli, R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, 10, 193–205. doi:10.1207/s15327906mbr1002\_5
- Jéquier, E., & Tappy, L. (1999). Regulation of body weight in humans. *Physiological Reviews*, 79, 451–480.
- Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., . . . Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School Age Children—Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36, 980–988. doi:10.1097/00004583-199707000-00021
- Kaufman, J., Birmaher, B., Brent, D., Rao, U., & Ryan, N. (1996). Kiddie SADS—Present and Lifetime Version (K-SADS-PL). Unpublished instrument, Western Psychiatric Institute and Clinics, University of Pittsburgh School of Medicine, PA.
- Kaufman, N. K., Rohde, P., Seeley, J. R., Clarke, G. N., & Stice, E. (2005). Potential mediators of cognitive-behavioral therapy for adolescents with comorbid major depressive and conduct disorder. *Journal of Con-*



- sulting and Clinical Psychology, 73, 38–46. doi:10.1037/0022-006X.73.1.38
- Kazdin, A. E., French, N. H., & Unis, A. S. (1983). Child, mother, and father evaluations of depression in psychiatric inpatient children. *Journal of Abnormal Child Psychology: An official publication of the International Society for Research in Child and Adolescent Psychopathology*, 11, 167–179. doi:10.1007/BF00912083
- Kessler, R. C., McGonagle, K. A., Swartz, M., Blazer, D. G., & Nelson, C. B. (1993). Sex and depression in the National Comorbidity Survey. I: Lifetime prevalence, chronicity and recurrence. *Journal of Affective Disorders*, 29, 85–96. doi:10.1016/0165-0327(93)90026-G
- Kim, S.-H., Cohen, A. S., Alagoz, C., & Kim, S. (2007). DIF detection and effect size measures for polytomously scored items. *Journal of Educational Measurement*, 44, 93–116. doi:10.1111/j.1745-3984.2007.00029.x
- Lerner, R. M., & Steinberg, L. (2004). The scientific study of adolescent development. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (2nd ed, pp. 1–12). New York, NY: John Wiley & Sons.
- Liss, H., Phares, V., & Liljequist, L. (2001). Symptom endorsement differences on the children's depression inventory with children and adolescents on an inpatient unit. *Journal of Personality Assessment*, 76, 396–411. doi:10.1207/S15327752JPA7603\_03
- March, J., Silva, S., Petrycki, S., Curry, J., Wells, K., Fairbank, J., . . . the TADS Team. (2004). Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for adolescents with depression study (TADS) randomized controlled trial. *JAMA: Journal of the American Medical Association*, 292, 807–820. doi:10.1001/jama.292.7.807
- March, J., Silva, S., Petrycki, S., Curry, J., Wells, K., Fairbank, J., . . . Severe, J. (2007). The treatment for adolescents with depression study (TADS): Long-term effectiveness and safety outcomes. *Archives of General Psychiatry*, 64, 1132–1143. doi:10.1001/archpsyc.64.10.1132
- Markowitz, S., Friedman, M. A., & Arent, S. M. (2008). Understanding the relation between obesity and depression: Causal mechanisms and implications for treatment. *Clinical Psychology: Science and Practice*, 15, 1–20. doi:10.1111/j.1468-2850.2008.00106.x
- Maxwell, M. A., & Cole, D. A. (2009). Weight change and appetite disturbance as symptoms of adolescent depression: Toward an integrative biopsychosocial model. *Clinical Psychology Review*, 29, 260–273. doi:10.1016/j.cpr.2009.01.007
- McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, 114, 376–390. doi:10.1037/0033-2909.114.2.376
- Mezulis, A. H., Priess, H. A., & Hyde, J. S. (2010). Rumination mediates the relationship between infant temperament and adolescent depressive symptoms. *Depression Research and Treatment*, 2011 (online, open access, doi:10.1155/2011/487873).
- Mitchell, J., McCauley, E., Burke, P. M., & Moss, S. J. (1988). Phenomenology of depression in children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, 27, 12–20. doi:10.1097/00004583-198801000-00004
- Muraki, E., & Carlson, J. E. (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement*, 19, 73–90. doi:10.1177/014662169501900109
- Muthén, L. K., & Muthén, B. O. (1998–2006). *Mplus 5.21* [Computer program]. Los Angeles, CA: Muthén & Muthén.
- Neary, N. M., Goldstone, A. P., & Bloom, S. R. (2004). Appetite regulation: From the gut to the hypothalamus. *Clinical Endocrinology (Oxf)*, 60, 153–160.
- Neumark-Sztainer, D., Wall, M., Guo, J., Story, M., Haines, J., & Eisenberg, M. (2006). Obesity, disordered eating, and eating disorders in a longitudinal study of adolescents: How do dieters fare 5 years later? *Journal of the American Dietetic Association*, 106, 559–568. doi:10.1016/j.jada.2006.01.003
- Nolen-Hoeksema, S. (1990). *Sex differences in depression*. Stanford, CA: Stanford University Press.
- Orvaschel, H. (1994). *Schedule for Affective Disorders and Schizophrenia for School-Age Children—Epidemiological version* (5th ed.). Ft. Lauderdale, FL: Nova Southeastern University.
- Oshima, T. C., Raju, N. S., & Flowers, C. P. (1997). Development and demonstration of multidimensional IRT-based internal measures of differential functioning of items and tests. *Journal of Educational Measurement*, 34, 253–272. doi:10.1111/j.1745-3984.1997.tb00518.x
- Patton, G. C., Carlin, J. B., Shao, Q., Hibbert, M. E., Rosier, M., Selzer, R., & Bowes, G. (1997). Adolescent dieting: Healthy weight control or borderline eating disorder. *Journal of Child Psychology and Psychiatry*, 38, 299–306. doi:10.1111/j.1469-7610.1997.tb01514.x
- Pilowsky, D. J., Wickramaratne, P., Talati, A., Tang, M., Hughes, C. W., Garber, J., . . . Weissman, M. M. (2008). Children of depressed mothers 1 year after the initiation of maternal treatment: Findings from the STAR\*D-Child Study. *The American Journal of Psychiatry*, 165, 1136–1147. doi:10.1176/appi.ajp.2008.07081286
- Priess, H. A., Lindberg, S., & Hyde, J. S. (2009). Adolescent gender-role identity and mental health: Gender intensification revisited. *Child Development*, 80, 1531–1544. doi:10.1111/j.1467-8624.2009.01349.x
- Raju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement*, 19, 353–368. doi:10.1177/014662169501900405
- Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika*, 53, 495–502. doi:10.1007/BF02294403
- Reckase, M. D. (1979). Unidimensional latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 207–230. doi:10.2307/1164671
- Reise, S. P., Waller, N., & Comray, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, 12, 287–297. doi:10.1037/1040-3590.12.3.287
- Roberts, R. E., Lewinsohn, P. M., & Seeley, J. R. (1995). Symptoms of DSM-III-R major depression in adolescence: Evidence from an epidemiological survey. *Journal of the American Academy of Child & Adolescent Psychiatry*, 34, 1608–1617. doi:10.1097/00004583-199512000-00011
- Rohde, P., Clarke, G., Mace, D., Jorgensen, J., & Seeley, J. (2004). An efficacy/effectiveness study of cognitive-behavioral treatment for adolescents with comorbid major depression and conduct disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43, 660–668. doi:10.1097/01.chi.0000121067.29744.41
- Rohde, P., Seeley, J., Kaufman, N., Clarke, G., & Stice, E. (2006). Predicting time to recovery among depressed adolescents treated in two psychosocial group interventions. *Journal of Consulting and Clinical Psychology*, 74, 80–88. doi:10.1037/0022-006X.74.1.80
- Romeo, R. D., & McEwen, B. S. (2006). Stress and the adolescent brain. *Annals of the New York Academy of Sciences*, 1094, 202.
- Ross, R., Dagnone, D., Jones, P. J. H., Smith, H., Paddags, A., Hudson, R., & Janssen, I. (2000). Reduction in obesity and related comorbid conditions after diet-induced weight loss or exercise-induced weight loss in men: A randomized, controlled trial. *Annals of Internal Medicine*, 133, 92–103.
- Ruscio, J., Zimmerman, M., McGlinchey, J. B., Chelminski, I., & Young, D. (2007). Diagnosing major depressive disorder XI: A taxometric investigation of the structure underlying DSM-IV symptoms. *Journal of Nervous and Mental Disease*, 195, 10–19. doi:10.1097/01.nmd.0000252025.12014.c4
- Rutter, M. (1986). The developmental psychopathology of depression: Issues and perspectives. In M. Rutter, C. E. Izard, & P. B. Read (Eds.), *Depression in young people*. New York, NY: Guilford Press.

- Sallis, J. F. (2000). Age-related decline in physical activity: A synthesis of human and animal studies. *Medicine and Science in Sports and Exercise*, 32(9), 1598–1600. doi:10.1097/00005768-200009000-00012
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded response. *Psychometric Monograph Supplement*, 34 (Monograph No. 17).
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507–514. doi:10.1007/BF02296192
- Stanley, S., Wynne, K., McGowan, B., & Bloom, S. (2005). Hormonal regulation of food intake. *Physiological Reviews*, 85, 1131–1158. doi:10.1152/physrev.00015.2004
- Stice, E., Cameron, R. P., Killen, J. D., Hayward, C., & Taylor, C. B. (1999). Naturalistic weight-reduction efforts prospectively predict growth in relative weight and onset of obesity among female adolescents. *Journal of Consulting and Clinical Psychology*, 67, 967–974. doi:10.1037/0022-006X.67.6.967
- Strober, M., Green, J., & Carlson, G. (1981). Phenomenology and subtypes of major depressive disorder in adolescence. *Journal of Affective Disorders*, 3, 281–290. doi:10.1016/0165-0327(81)90029-X
- Terasawa, E., & Fernandez, D. L. (2001). Neurobiological mechanisms of the onset of puberty in primates. *Endocrine Reviews*, 22, 111–151. doi:10.1210/er.22.1.111
- The TADS Team. (2003). The Treatment for Adolescents With Depression Study (TADS): Rationale, design, and methods. *Journal of the American Academy of Child & Adolescent Psychiatry*, 42, 531–542. doi:10.1097/01.CHI.0000046839.90931.0D
- The TADS Team. (2005). The Treatment for Adolescents With Depression Study (TADS): Demographic and clinical characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44, 28–40. doi:10.1097/01.chi.0000145807.09027.82
- Tice, D. M., Bratslavsky, E., & Baumeister, R. F. (2001). Emotional distress regulation takes precedence over impulse control: If you feel bad, do it. *Journal of Personality and Social Psychology*, 80, 53–67. doi:10.1037/0022-3514.80.1.53
- Turner, R. J., Wheaton, B., & Lloyd, D. A. (1995). The epidemiology of social stress. *American Sociological Review*, 60, 104–125. doi:10.2307/2096348
- van Prooijen, J., & Van der Kloot, W. A. (2001). Confirmatory analysis of exploratively obtained factor structures. *Educational and Psychological Measurement*, 61, 777–792. doi:10.1177/00131640121971518
- Vartanian, L. R., Herman, C. P., & Polivy, J. (2004). Accuracy in the estimation of body weight: An alternate test of the Motivated-Distortion Hypothesis. *International Journal of Eating Disorders*, 36, 69–75. doi:10.1002/eat.20014
- Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41–71). Norwell, MA: Kluwer Academic.
- Wang, Z., Patterson, C. M., & Hills, A. P. (2002). A comparison of self-reported and measured height, weight and BMI in Australian adolescents. *Australian and New Zealand Journal of Public Health*, 26, 473–478. doi:10.1111/j.1467-842X.2002.tb00350.x
- Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). The comparative effects of compensatory and non-compensatory two dimensional data on unidimensional IRT estimates. *Applied Psychological Measurement*, 12, 239–252. doi:10.1177/014662168801200303
- Weiss, B., & Garber, J. (2003). Developmental differences in the phenomenology of depression. *Development and Psychopathology*, 15, 403–430. doi:10.1017/S0954579403000221
- Weissman, M. M., Pilowsky, D. J., Wickramaratne, P. J., Talati, A., Wisniewski, S. R., Fava, M., et al. for the STAR\*D Child Team. (2006). Remissions in maternal depression and child psychopathology: A STAR\*D-Child report. *JAMA: Journal of the American Medical Association*, 295, 1389–1398. doi:10.1001/jama.295.12.1389
- Yorbik, O., Birmaher, B., Axelson, D., Williamson, D. E., & Ryan, N. D. (2004). Clinical characteristics of depressive symptoms in children and adolescents with major depressive disorder. *Journal of Clinical Psychiatry*, 65, 1654–1659. doi:10.4088/JCP.v65n1210
- Youngstrom, E., Meyers, O., Demeter, C., Youngstrom, J., Morello, L., Piiparinen, R., . . . Findling, R. L. (2005). Comparing diagnostic checklists for pediatric bipolar disorder in academic and community mental health settings. *Bipolar Disorders*, 7, 507–517. doi:10.1111/j.1399-5618.2005.00269.x
- Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Unpublished doctoral dissertation, University of California, Los Angeles, CA.
- Zimmerman, M., McGlinchey, J. B., Young, D., & Chelminski, I. (2006). Diagnosing Major Depressive Disorder I: Psychometric evaluation of the DSM-IV symptom criteria. *Journal of Nervous and Mental Disease*, 194, 158–163. doi:10.1097/01.nmd.0000202239.20315.16
- Zwick, W. R., & Velicer, W. F. (1986). Factor influencing five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432–442. doi:10.1037/0033-2909.99.3.432

Received October 21, 2011

Revision received January 20, 2012

Accepted January 20, 2012 ■