Age, Gender, Race, Socioeconomic Status, and Birth Cohort Differences on the Children’s Depression Inventory: A Meta-Analysis

Jean M. Twenge
San Diego State University

Susan Nolen-Hoeksema
University of Michigan

A within-scale meta-analysis was performed on 310 samples of children (ages 8–16; \( N = 61,424 \)) responding to the Children’s Depression Inventory (CDI). Girls’ depression scores stayed steady from ages 8 to 11 and then increased between ages 12 and 16. Boys’ CDI scores were stable from ages 8 to 16 except for a high CDI score at age 12. Girls’ scores were slightly lower than boys’ during childhood, but girls scored higher beginning at age 13. There were no socioeconomic status effects and no differences between White and Black samples. However, Hispanic samples scored significantly higher on the CDI.

Analyses for birth cohort showed a slight decrease in boys’ CDI scores over time and no change for girls. Longitudinal studies demonstrated a marked testing effect.

Depression in children and adolescents has been a frequent focus of research in clinical and developmental psychology in the last 3 decades (see reviews by Cicchetti & Toth, 1998; Kovacs & Devlin, 1998; Nolen-Hoeksema & Girgus, 1994; Petersen et al., 1993; and Stavroukaki & Gaudet, 1989). Much of this research has sought to document changes with age in the prevalence of depressive symptoms, particularly as children transition into adolescence. Other work has focused on differences in the prevalence of depression associated with gender, race, or class. Most studies have been cross-sectional, but several longitudinal studies of depression in children and adolescents now exist as well. In this article, we report a meta-analysis of 310 studies of depressive symptoms in children and adolescents, examining age, gender, birth cohort, race, and class differences in these symptoms. We analyzed children’s scores on the Children’s Depression Inventory (CDI; Kovacs, 1985, 1992) using a technique called within-scale meta-analysis (see, e.g., Twenge, 1997b, 2000; Twenge & Campbell, 2001). This unique method allows the comparison of scores across many groups by summarizing and analyzing mean scores reported in the literature for a particular scale. Overall, these studies collected data from 61,424 children.

Measuring Depressive Symptoms

Over the years, the most frequently used scale with children has been the CDI (Kovacs, 1985, 1992). The CDI was designed to measure symptoms of depression from several domains, including emotional, cognitive, psychomotor, and motivational (see Shaver & Brennan, 1991). This scale presents children with groups of three sentences that describe various levels of a given symptom of depression, as in the following:

I am sad once in a while.
I am sad much of the time.
I am sad all the time.¹

Children are asked to indicate which sentence best describes the way they have been feeling for the past 1 or 2 weeks. Item choices are assigned a numerical value from 0 to 2. High scores on the CDI indicate higher levels of depressive symptoms. The CDI has shown good test–retest reliability, internal consistency, and construct validity, especially in nonclinic populations (for a review, see Sitrenios & Kovacs, 1999). Scores on the CDI are moderately correlated with psychiatrists’ ratings of children’s levels of depression (Kazdin, 1989) and differentiate children with mood disorders from children with anorexia nervosa or behavioral disturbances (Carlson & Cantwell, 1980). The CDI does not appear to distinguish well between symptoms of depression and symptoms of anxiety, however, and is often considered a measure of general distress rather than depression alone (Saylor, Finch, Spirito, & Bennett, 1984). Yet, it should be noted that depressive and anxiety symptoms are highly comorbid in children and adolescents, whereas pure depression is relatively rare (Kovacs & Devlin, 1998).

The CDI has been used most often to measure depressive symptoms among normal populations of children; thus, it usually measures depression at a subclinical level. Yet, there is evidence that even moderate levels of depression can have a negative impact on children, adolescents, and adults. Moderate levels of depression have been found to persist for months, and even years, in some children (Nolen-Hoeksema, Girgus, & Seligman, 1992) and some adults (Aneshensel, 1985; Barrett, Hurst, DiScala, & Rose, 1978). In addition, chronic, moderate depression is associated with significant impairment in school and peer functioning in children.

¹ Copyright 1982 by Maria Kovacs, PhD. Copyright 1991, 1992 by Multi-Health Systems Inc. All rights reserved. Reproduced with permission.
Age and Gender Differences in Depressive Symptoms

Qualitative reviews have suggested that the prevalence of depressive symptoms increases from childhood into adolescence but only among girls and not among boys (e.g., Hankin & Abramson, 1999, 2001; Nolen-Hoeksema & Girgsa, 1994). Indeed, some studies suggest that, in preadolescence, boys have higher levels of self-reported depressive symptoms than girls (Angold & Rutter, 1992; Nolen-Hoeksema, Girgsa, & Seligman, 1991). Girls’ levels of self-reported depressive symptoms then increase rapidly in early adolescence, but boys’ levels either increase only slightly or remain stable. Not all studies of self-reported depressive symptoms in community samples of children and adolescents have found these Age × Gender trends, however (e.g., Compas et al., 1997; Lefkowitz & Tesiny, 1985; Schoenback, Kaplan, & Wagner, 1983). For example, Compas et al. (1997) found that girls’ depressive symptom scores were fairly stable until age 17, when they increased; boys’ symptoms were steady until age 15 and then increased. Meta-analysis is an appropriate technique to quantitatively summarize this body of literature and determine the trajectories of girls’ and boys’ depressive symptoms from childhood to adolescence.

In addition, in studies that do find the emergence of a gender difference in depressive symptoms in early adolescence, the timing of this emergence has been somewhat unclear (Hankin & Abramson, 1999). The timing of the emergence of gender differences holds important implications for the various theories of why girls would show increases in depressive symptoms in adolescence (Hankin & Abramson, 2001; Nolen-Hoeksema & Girgsa, 1994). For example, several theorists have argued that the physiological changes of puberty increase girls’ risk for depression (e.g., Angold, Costello, & Worthman, 1998). The precise timing of the emergence of more depression in girls would suggest which physiological changes might be increasing girls’ risk of depression. Alternatively, social changes such as school transitions could be the crucial turning point for gender differences in depression.

In this article, we report analyses of the Age × Gender trends in levels of depressive symptoms across 310 community samples of children and adolescents and attempt to specify the period when girls begin to show increases in depressive symptoms relative to boys, if indeed this consistently happens across studies. We also compute an effect size for gender differences at each age level to determine the magnitude of the gender difference in depressive symptoms at each age level.

Birth Cohort Effects

If a consistent change in CDI scores with children’s age is found, this would not necessarily represent an age effect. It could also represent a birth cohort effect. Birth cohort is an effective proxy for the larger sociocultural environment (Casp, 1987; Stewart & Healy, 1989; Twenge, 2001, 2002). Children’s depressive symptoms may have been affected by the changes of the last 20 years. Some of these changes have been positive: The economy has been robust, the divorce rate has decreased since the early 1980s, and worries about nuclear war are down (see Twenge, 2000; Twenge & Campbell, 2001). On the other hand, concern about AIDS grew throughout most of this period, and level of trust in other people declined steadily (see Fukuyama, 1999; Twenge, 2000). Other changes have been mixed: The violent crime rate increased significantly from the early 1980s to the early 1990s but decreased after 1993 (U.S. Bureau of the Census, 1999). Unlike the 1970s, which showed unprecedented declines in almost every social indicator, the 1980s and 1990s are characterized by forces of opposite valence. These changes suggest that children’s depression could decrease, increase, or stay steady over time.

Studies of depressive disorders in adults have suggested a substantial birth cohort effect such that more recent generations appear much more prone to depression than more distant generations (Klerman & Weissman, 1989; Lewinsohn, Rohde, Seeley, & Fischer, 1995). Unfortunately, however, most of these previous studies have all relied on retrospective accounts of depressive episodes (e.g., Klerman & Weissman, 1989; Lewinsohn, Rohde, Seeley, & Fischer, 1995). As these authors have acknowledged, this is a major limitation: Memories could be skewed or selective, and different birth cohorts could have different norms for what constitutes a depressive episode and when to report it. Two studies of depressive disorders in children have found a birth cohort effect over a relatively short number of years (Kovacs & Gatzonis, 1994; Ryan, Williamson, Iyengar, & Orvaschel, 1992). The present analysis extends these findings, examining self-report data collected at different points in time for birth cohorts spanning over 2 decades, a method that sidesteps the confounds of retrospective accounts. In addition, the children in these studies were born between the mid-1960s and the late 1980s, birth cohorts that have not been studied extensively. (They were not included in the Klerman and Weissman, 1989, or the Lewinsohn, Rohde, et al., 1993). As these authors have acknowledged, this is a major limitation: Memories could be skewed or selective, and different birth cohorts could have different norms for what constitutes a depressive episode and when to report it. Two studies of depressive disorders in children have found a birth cohort effect over a relatively short number of years (Kovacs & Gatzonis, 1994; Ryan, Williamson, Iyengar, & Orvaschel, 1992). The present analysis extends these findings, examining self-report data collected at different points in time for birth cohorts spanning over 2 decades, a method that sidesteps the confounds of retrospective accounts. In addition, the children in these studies were born between the mid-1960s and the late 1980s, birth cohorts that have not been studied extensively. (They were not included in the Klerman and Weissman, 1989, or the Lewinsohn, Rohde, et al., 1993). As these authors have acknowledged, this is a major limitation: Memories could be skewed or selective, and different birth cohorts could have different norms for what constitutes a depressive episode and when to report it. Two studies of depressive disorders in children have found a birth cohort effect over a relatively short number of years (Kovacs & Gatzonis, 1994; Ryan, Williamso
when the children were age 9, girls had a mean CDI score of 9.87, whereas at the twelfth administration, when they were age 14, they had a mean CDI score of 5.50. Another study found a CDI mean score of 10.10 at the first administration, and progressively lower scores at administrations 2 weeks apart, until the mean score at 6 weeks was 8.57 (Finch, Saylor, Edwards, & McIntosh, 1987). Other longitudinal studies of depressive symptoms have shown similar effects (e.g., Angold, Erkanli, Loebel, & Costello, 1996; Burks, Dodge, & Price, 1995; Cole, Peeke, Martin, Truglio, & Seroczynski, 1998; Devine, Kempton, & Forehand, 1994). (Note that these effects have not been due to attrition.) Even two administrations is enough to create a decrease in scores (Slavin & Rainer, 1990).

In contrast, a cross-sectional study of the same age groups (Smucker, 1982) found a mean CDI score for 8-year-old girls of 8.68 versus a mean of 11.13 for 14-year-old girls. Other cross-sectional studies have found similar results (e.g., Craighead, Smucker, Craighead, & Ilardi, 1998; Finch, Saylor, & Edwards, 1985; Hecht, Inderbitzen, & Bukowski, 1998). In addition, longitudinal studies of clinical depression assessed by diagnosis and interview have not found decreases with age (e.g., Hankin et al., 1998). Thus, the self-report method may be the cause of the decrease.

The decrease in mean scores over time in longitudinal studies could represent a testing or measurement effect. Sharpe and Gilbert (1998) found that college students’ responses to the Beck Depression Inventory decreased 25% after three administrations of the measure. These decreases may have occurred for several reasons (Sharpe & Gilbert, 1998), including social desirability, habituation or boredom, decreases in test anxiety, and the initiation of coping mechanisms. There has been debate over the social desirability explanation in particular. Choquette and Hesselbrock (1987) argued that response biases may appear as respondents realize what the scales measure; however, Sharpe and Gilbert (1998) questioned this explanation given that measures of positive affect do not demonstrate increases over repeated administrations. Sharpe and Gilbert favored a coping explanation, in which participants who respond affirmatively to depression items might subsequently seek help, talk out their problems and, in general, use coping mechanisms once they realize that they feel depressed.

Of course, these testing effects do not necessarily mean that the most valid way to examine age changes in depression scores is through cross-sectional studies because these studies confound age and birth cohort. In the present analyses, we examined depression scores from samples collected over a range of 20 years, and from every region of the United States, so we could better tease apart birth cohort, testing, and age effects.

Race/Ethnicity and Social Class Differences in Depressive Symptoms

Very few individual studies have had sufficient numbers of children and adolescents from different racial/ethnic backgrounds and different social classes to compare their levels of depression (see Stavrakaki & Gaudet, 1989). One exception is a large study by Roberts, Roberts, and Chen (1997), in which an ethnically diverse sample of over 5,000 students in Grades 6 through 8 were assessed on major depression. Mexican American youth had higher rates of depression than other groups. In addition, youths who reported that their socioeconomic status (SES) was somewhat or much worse off than their peers had a higher prevalence of depression. These results parallel findings from the most recent nationwide community sample of adults in the United States, the National Comorbidity Survey, which found people of Hispanic ethnic background to have higher levels of current depressive disorders than African Americans or Whites and that a low SES was correlated with a greater prevalence of depression (Blazer, Kessler, McGonagle, & Swartz, 1994). However, another large study found the opposite, with White and Asian American youth reporting more depressive symptoms than Black and Hispanic youth (Dornbusch, Mont-Reynaud, Ritter, Chen, & Steinberg, 1991).

In this article, we took advantage of the fact that a meta-analysis allows the examination of racial/ethnic differences in depressive symptoms across studies. In this meta-analysis, we compared studies with samples that were 90% or more from one racial/ethnic background. We also compiled data across studies to examine the relationship of SES to depression. These techniques allowed the examination of mean scores among large numbers of respondents.

Why Perform a Within-Scale Meta-Analysis?

To address the questions raised here, we performed a within-scale meta-analysis rather than a traditional meta-analysis. For our purposes, a within-scale meta-analysis is superior to a traditional meta-analysis for several reasons (see also Twenge & Campbell, 2001). A conventional meta-analysis computes an effect size (usually, a difference in terms of standard deviations) between two groups. As we examined it here, age included a span of 9 years. A conventional meta-analysis would have to compute 36 effect sizes to compare all of the groups; even then, the comparisons would only reflect differences between two groups without reference to the others. Gender differences could be computed within studies, but the developmental change within samples of boys and girls would be lost (for example, an increase in the gender difference during adolescence could mean several things: that boys’ scores were decreasing and girls’ were increasing; that both genders’ scores were increasing but girls’ scores increased at a faster rate; that boys’ scores were unchanged whereas girls’ scores increased, etc.). In addition, a conventional meta-analysis can only include studies that tested at least two distinct age groups, whereas a within-scale analysis can include studies with only one age group. Furthermore, cohort differences must be studied using means because most studies do not include samples of different birth cohorts. Using the mean scores of samples allows the isolation of birth cohort differences in a time-lag design. Time-lag is an underutilized methodology that examines same-age groups at different points in time (Schaie, 1965). The CDI is by far the most popular measure of children’s depression. Thus, there are many samples available to analyze, and the vast majority of useful information is included. Overall, such a within-scale analysis allows us to generalize over many domains, gathering datapoints collected at many different locations and times. It also allows us to examine differences that few individual studies can (for example, across many age groups, racial groups, and birth cohorts). We can draw conclusions that go beyond individual studies gathering data from a specific population. Performing a within-scale meta-analysis is thus an innovative and effective way to examine individual differences in CDI scores. This method has the necessary
limitation that it can examine only one measure; however, the CDI is so widely used and trusted that this analysis should provide a fairly accurate general picture of self-reported distress in children.

**Method**

**Locating Studies**

Datapoints were located using the Web of Science Citation Index (CI; for published datapoints) and Dissertation Abstracts International (DAI; for unpublished dissertations and master’s theses). The CI includes the Social Sciences Citation Index, the Science Citation Index, and the Arts and Humanities Citation Index. Thus, it spans a wide range of fields (for example, it includes medical journals as well as psychology and sociology journals). It also includes both major and minor journals. We searched the CI for all articles that cited the original sources for the CDI (Kovacs, 1980/1981, 1983, 1985; Kovacs & Beck, 1977).

Because the DAI has no citation index, we searched it using the keyword “Children’s Depression Inventory.” Although mean scores are unlikely to show a publication bias, searching for dissertations ensures that unpublished studies are included. In addition, including dissertations widens the pool of possible datapoints (in particular because dissertations are more likely to report extensive tables of means).

**Inclusion Rules**

Possible studies for the meta-analysis were included or excluded on the basis of specific rules. To be included, a study had to meet the following criteria: (a) Samples were collected in the United States or Canada. (Thus, samples from all other countries were eliminated; it is expected that the differing cultural and political climates in these countries would confound the results); (b) the study included at least 15 participants; (c) participants were not psychiatric patients, delinquents, hospital patients, sufferers from a particular disease, or any other group singled out for maladjustment; (d) means were reported for unselected groups of students, not groups that were extremely high or extremely low on the CDI, an anxiety measure, or any other scale likely to be correlated with depression scores; and (e) means for the 27- or 26-item CDI were reported. A number of studies did not provide the mean for their sample on the CDI; most studies excluded from the analysis were eliminated for not including means rather than for violating any other criterion. Some studies were eliminated because the data they reported were identical to data sets from other sources.

**Coding of Variables**

Age, gender, SES, and race/ethnicity were coded from the information provided in the articles. Given the importance of age as a variable, we eliminated studies that reported only summary means for samples with an age range of 4 years or greater. We included samples testing ages 8 to 16; only a few samples were ages 7 or younger, or 17 or older. SES was coded using four categories: 1 (lower class), 2 (lower to middle class), 3 (middle class), and 4 (middle to upper class) using the information provided in the article (usually an explicit labeling of the sample into one of these categories). Not all studies provided information on SES. As for the racial/ethnic breakdown of the samples, the percentage of participants of each race/ethnic group (White, Black, Hispanic, or Asian) was recorded. Samples reporting CDI means broken down by racial group were entered as separate datapoints (for example, an article reporting separate CDI means for Black and White respondents would be entered as two separate datapoints). For the analyses here, we considered a sample uniracial if 90% of more of the participants were a particular race/ethnic group. There were very few samples that were more than 90% Asian. Thus, the race analyses compare White, Black, and Hispanic samples. Year of data collection was coded as 2 years prior to publication unless the year was otherwise noted in the article (Oliver & Hyde, 1993). Longitudinal studies were given special consideration, with the time elapsed since the first administration also subtracted from the year of publication (thus, a study published in 1988 that collected data 1 year apart over 3 years would receive a year of 1984 for Wave 1, 1985 for Wave 2, and 1986 for Wave 3). The years of data collection available for our analysis ranged from 1980 to 1998.

**Adjusting Means**

In some cases, it was necessary to adjust the means. Some studies used only 26 items of the CDI because school officials or review boards would not allow them to include the item about suicide. These datapoints were recalculated to conform to the usual 27-item mean (the 26-item mean divided by 26, and the result added to the 26-item mean). Controls for number of items (26 vs. 27) did not produce any changes in the results. Studies using less than 26 items were not included in the analyses. No other adjustments were made to mean scores.

**Resulting Sample Size**

These data collection and inclusion strategies yielded 310 datapoints/data sets. In total, these samples included 61,424 children between the ages of 8 and 16 (29,637 boys and 31,787 girls). Because some studies yielded more than one datapoint, the number of studies is fewer than the number of datapoints. (A reference list of the studies included in the analyses is available on request from Jean M. Twenge.) The studies were relatively well distributed among the ages studied here (8 to 16), with somewhat fewer studies at the ends of the distribution (e.g., age 8, age 16). About 50% of the samples were of mixed race (with no one group exceeding 90% of the sample); about 40% were 90% or more White, and the remaining 10% were composed of 90% or more of another racial group (Black, Hispanic, or Asian).

**Within-Scale Meta-Analysis Method**

When used previously by one of us (Twenge, 1997a, 1997b, 2000, 2001a, 2001b), within-scale meta-analysis has been called cross-temporal meta-analysis (because the main variable of interest in previous studies was birth cohort or change over time). However, the method can also be used to study other moderator variables such as age, race, and SES (see Twenge & Campbell, 2001). Because this method differs from traditional meta-analysis, we explain it in some detail. Procedures for identifying and collecting data for studies are essentially the same as in a traditional meta-analysis. However, instead of computing an effect size for each study, a within-scale meta-analysis records the sample means (for example, the mean CDI score for a sample of male 5th and 6th graders). Sample characteristics are coded (e.g., year of data collection, age, sample size, race, region).

The data are then analyzed using conventional statistical procedures but weighted by sample size. The weights differ from those used in traditional meta-analysis, which relies on a formula involving the effect size (d) and sample size (e.g., Hedges & Becker, 1986). Because this analysis uses means (which do not have the same bias as d), sample size alone is used as a weight (w). (In practice, weighting by sample size is very similar to weights used for d because the weights for d correct sample size for a very small bias in d; see Twenge & Campbell, 2001, for a discussion.) Weighting by sample size allows larger studies (which are better estimates of the population mean) to carry more weight in the correlations or averages. For example, a study collecting data on 200 14-year-olds would carry twice as much weight as a study with 100 14-year-olds when computing the average CDI score of 14-year-olds across studies. One exception to the weighting scheme in our study was the case of gender differences, in which we used conventional meta-analytic techniques to compute an effect size (d) for each sample reporting CDI means for both boys and girls. (We were able
to compute $d$ in these cases because a sufficient number of studies provided means broken down by sex, allowing the computation of the within-study effect size. In this case, we did rely on $w$ as a weight, as is conventional when analyzing traditional effect sizes (Hedges & Becker, 1986).

Results

Testing Effects in Longitudinal Studies

Several studies included here were longitudinal analyses; we entered the means for each wave as separate datapoints. However, these are not technically independent observations because they involve the same individuals at different points in time. There is also the possibility of testing effects, as discussed earlier. We analyzed directly for testing effects by computing the correlation between CDI scores and wave/time (i.e., Wave 1, Wave 2, etc.), weighted by sample size. These correlations were all negative and significant, meaning that CDI scores decreased with each administration of the scale. For male samples, $r(45) = -0.34, p < .01$; for female samples, $r(45) = -0.58, p < .001$; and for total samples (both sexes combined), $r(87) = -0.44, p < .001$. These correlations were very similar when controlled for age and birth cohort, suggesting that this is not a period effect. The correlations were also very similar when only the second and later waves of the studies were analyzed; thus, the effect is not caused only by a drop between the first and second administrations (see a further analysis of decreases between specific waves in the following paragraph). A regression including both linear and quadratic trends showed that quadratic trends were not significant for any of the groups. As we noted in the introduction, cross-sectional studies do not reveal decreases with children’s age in CDI scores. Thus, the decreases are most likely to be measurement effects.

We examined data from the total samples (which had the most studies available) to determine if the decrease in scores was steady, accelerating, or slowing with successive waves. Overall, the changes were gradual and fairly steady (see Table 1). Using the average standard deviations from the individual studies for each wave, we computed effect sizes ($d$, or the difference in terms of standard deviations) for the decreases between waves. Scores decreased 0.08 standard deviations from Wave 1 to Wave 2; 0.03 from Wave 2 to Wave 3; 0.07 from Wave 3 to Wave 4; 0.04 from Wave 4 to Wave 5; and 0.01 from Wave 5 to Wave 6 (only one study had more than 6 waves). Thus, from Wave 1 to Wave 6, average CDI scores decreased 0.23 standard deviations. This qualifies as a small effect size (around .20) according to Cohen (1977); a medium effect is around .50, and a large effect is around .80). This effect size equals or exceeds the increase in girls’ depression from childhood to adolescence (see the section on Age and Gender later). Thus, it represents a serious confound in the study of age differences in depression in longitudinal studies.

Are testing effects caused by the number of times the respondent completes the measure or by the amount of time that has passed between administrations? Using the total sample data, we selected studies that had a 6-month lag between administrations with those that had a 1-year lag between administrations. Specifically, we compared the third wave of studies with a 6-month lag to the second wave of studies with a 1-year lag; thus, all measurements were taken 1 year after the original testing, but the session was either the second or third time the respondents had completed the measure. A t-test showed that the scores of the third-wave respondents were significantly lower ($M = 7.18$) than those of the second-wave respondents ($M = 8.24$), $t(37) = 2.23, p < .05$ (even though the same amount of chronological time had passed for both groups). In contrast, the second waves of each study type were not significantly different from each other ($M = 7.99$ for 6-month lag vs. $M = 8.24$ for 1-year lag). Thus, the number of times the respondents have completed the measure seems to be the crucial variable rather than the amount of time that has passed. This also suggests that a period effect (in which scores are influenced by the year the testing was done) is not operating here (in addition, a period effect is unlikely over the span of just 1 year). Overall, these results suggest that scores decrease slowly but steadily every time a group of respondents completes a measure and that the resulting decrease in scores is a noticeable effect.

Given this, we decided that the later waves of longitudinal studies should not be included in the remaining analyses. Thus, the remainder of the analyses include only nonlongitudinal studies and the first wave only of longitudinal studies. These totaled 251 samples, including 43,916 children (21,106 boys and 22,900 girls).

Age and Gender

Developmental changes within gender. How do depression scores change over development, and are these changes different for boys and girls? When do girls begin to exhibit more depression compared with boys? Because previous evidence suggests that age changes vary considerably between boys and girls, we analyzed the data separately by gender. To begin the analyses, we weighted the means by sample size. We then introduced year of data collection as a covariate in an analysis of covariance (ANCOVA) and obtained means adjusted for year (this effectively controls for birth cohort). The means for the age and gender groups are displayed in Table 2 and Figure 1. For boys, depression scores were basically flat with the exception of comparatively high depression scores at age 12. Overall, age did not correlate with depression scores for boys, $r(137) = -.03$ (with year of data collection controlled, $r = -.01$). A regression equation including both linear and quadratic trends showed that neither was significant. Girls, however, showed a significant increase in depression scores from childhood to adolescence, $r(139) = .35, p < .001$ (with year of data collection controlled, $r = .36$). A regression equation including the linear and quadratic effects showed no curvilinear trends. Girls also demonstrated elevated depression scores at age 12, but their CDI scores at 14, 15, and 16 were even higher.

Table 1

<table>
<thead>
<tr>
<th>Wave</th>
<th>$k$</th>
<th>$M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>8.68</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>8.08</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>7.84</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>7.32</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>7.01</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>6.96</td>
</tr>
</tbody>
</table>

Note. $k$ = number of samples.
We computed the magnitude of the increase in girls' depression scores by using the average standard deviation of girls' CDI scores (7.12, obtained from the studies in the analysis). The change in depression from age 8 \( (M = 8.81) \) to age 16 \( (M = 10.06) \) represents an increase of 0.18 standard deviations. This nearly explains a significant portion of the variance, and age seemed the most obvious candidate.

From age 8 to age 12, boys scored slightly higher than girls on the CDI (thus the negative \( d \)s in Table 2 at those ages). However, these gender differences were not statistically significant, as the 95% confidence intervals (CIs) for \( d \) include zero; for a meta-analytic effect size, the test of statistical significance is whether the 95% CI includes zero or not (see Table 2). If all of the samples of ages 8–12 were combined (\( d = -0.04; k = 86, \) where \( k \) is the number of samples), the 95% CI barely eludes zero \( (CI = -0.07 \) to \(-0.01) \). Thus, strictly speaking, there were no gender differences in CDI scores from age 8 to age 12.

Starting at age 13, however, girls reported significantly higher levels of depressive symptoms, with the CI not including zero \( (d = .08 \) at age 13; see Table 2). By age 14, the difference reached \( .21 \), satisfying Cohen's (1977) criteria for a small effect \( (d = .20) \). All of the gender differences favoring girls after age 13 were statistically significant, as the 95% CI for \( d \) did not include zero (see Table 2). This was also true when the samples of ages 13–16 were combined: \( d = .16, k = 49, CI = .12 \) to .20. The gender difference during childhood \( (d = -0.04) \) was significantly different from the gender difference during adolescence \( (d = .16) \), \( \chi^2 = 53.76, k = 129, p < .001 \). These shifts in scores were reflected in the correlation between \( d \) and age (weighted by \( w \), with the CI calculated using the procedure in Hedges & Becker, 1986), \( r(129) = .42, p < .001 \), meaning here that \( d \) began negative (with boys scoring higher) and was then positive and larger (with girls scoring higher) over increasing age. These results were unchanged when year of data collection was entered into the regression equation. There was no significant quadratic effect when both linear and quadratic trends were entered into a regression equation.

<table>
<thead>
<tr>
<th>Age</th>
<th>Boys</th>
<th>Girls</th>
<th>( d )</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>8.98</td>
<td>8.81</td>
<td>-0.05</td>
<td>-0.14 to .04</td>
</tr>
<tr>
<td>9</td>
<td>9.02</td>
<td>8.73</td>
<td>-0.02</td>
<td>-0.09 to .05</td>
</tr>
<tr>
<td>10</td>
<td>9.17</td>
<td>8.75</td>
<td>-0.05</td>
<td>-0.11 to .01</td>
</tr>
<tr>
<td>11</td>
<td>8.50</td>
<td>8.36</td>
<td>-0.04</td>
<td>-0.10 to .02</td>
</tr>
<tr>
<td>12</td>
<td>9.93</td>
<td>9.42</td>
<td>-0.06</td>
<td>-0.14 to .02</td>
</tr>
<tr>
<td>13</td>
<td>8.66</td>
<td>9.05</td>
<td>-0.08</td>
<td>-0.01 to .15</td>
</tr>
<tr>
<td>14</td>
<td>8.99</td>
<td>10.42</td>
<td>.22</td>
<td>.13 to .31</td>
</tr>
<tr>
<td>15</td>
<td>8.74</td>
<td>10.51</td>
<td>.22</td>
<td>.13 to .31</td>
</tr>
<tr>
<td>16</td>
<td>9.06</td>
<td>10.06</td>
<td>.18</td>
<td>.05 to .31</td>
</tr>
</tbody>
</table>

\( r \) with age \( .03, .35*, .42* \)

**Table 2**

**Boys’ and Girls’ Scores by Age on the Children’s Depression Inventory**

Note. Negative \( d \) = boys score higher; positive \( d \) = girls score higher. 95% CI = 95% confidence interval for effect size \( (d) \).

\* \( p < .001 \).

**Race/Ethnicity**

Our analyses for race compared samples that were more than 90% White, Black, or Hispanic in composition. We were able to compare only mixed-sex samples because very few male-only or female-only samples of a particular race were available. Overall, Hispanic samples scored higher on the CDI than Whites or Blacks. The 109 mixed-sex samples of Whites averaged 8.84 on the CDI compared with 8.67 for fourteen samples of Blacks and 10.34 for eleven samples of Hispanics (these means were weighted by sample size). Hispanics scored significantly higher on the CDI compared with Whites, \( r(118) = 3.37, p < .001 \). This represented an effect size of .62. Hispanics also scored significantly higher than Blacks, \( r(23) = 3.13, p < .01, d = 1.31 \). There was not a significant difference between the CDI scores of White and Black samples, \( r(121) = 0.42 \). Thus, Hispanic children, on average, reported more depressive symptoms than White or Black children. Given the cutoff for a large effect size of .80, these race/ethnicity differences represent moderate to large effect sizes (Cohen, 1977). This is also the largest effect size among the individual differences studied here (see Table 3).

**Birth Cohort**

We computed the birth cohort of our samples by subtracting sample age from year of data collection. Contrary to expectations,
Gender difference in depression 

Decrease across six waves, in longitudinal studies

Gender difference in depression

During childhood (ages 8–12)

During adolescence (ages 13–16)

Increase in girls’ depression

From age 8 to age 16

From age 11 to age 15

Hispanic samples

Compared with White samples

Compared with Black samples

Birth cohort, boys only

(birth years 1964–1988)

Note. d = difference in terms of standard deviations.

a Not statistically different from zero.

Table 3

Summary of Effect Sizes

<table>
<thead>
<tr>
<th>Comparison</th>
<th>d</th>
<th>Scored higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decrease across six waves, in longitudinal studies</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>Gender difference in depression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>During childhood (ages 8–12)</td>
<td>.04*</td>
<td>Boys</td>
</tr>
<tr>
<td>During adolescence (ages 13–16)</td>
<td>.16</td>
<td>Girls</td>
</tr>
<tr>
<td>Increase in girls’ depression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>From age 8 to age 16</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>From age 11 to age 15</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>Hispanic samples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compared with White samples</td>
<td>.62</td>
<td>Hispanics</td>
</tr>
<tr>
<td>Compared with Black samples</td>
<td>1.31</td>
<td>Hispanics</td>
</tr>
<tr>
<td>Birth cohort, boys only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(birth years 1964–1988)</td>
<td>.19</td>
<td></td>
</tr>
</tbody>
</table>

CDI scores were negatively correlated with birth cohort, meaning that depression scores have decreased slightly with successive birth cohorts: for male samples, r(137) = −.22, p < .05; for female samples, r(139) = −.21, p < .05; for mixed-sex samples, r(234) = −.25, p < .001 (these correlations were weighted by sample size). This effect was not due to the use of the 26-item (vs. the 27-item) form of the CDI; entering number of items into a regression equation did not change the effect. Quadratic regressions showed no significant curvilinear effects.

It is possible that individuals from less recent birth cohorts were older when they completed the CDI. If depression increases with age (as it appears to for girls), this could create the false appearance of birth cohort effects. The correlation between year of data collection and CDI scores allowed us to examine the effect across all ages combined, guarding against this confound (this also allowed us to examine the period effect, or how the year of data collection alone had affected CDI scores). The correlations between year and CDI scores were as follows: for boys, r(137) = −.24, p < .01; for girls, r(139) = −.08, ns; for mixed-sex samples, r(234) = −.17, p < .05. The results were similar when year and age were entered simultaneously in a regression equation: for boys, the correlation between CDI scores and year was r(137) = −.24, p < .01; for girls, r(139) = −.12, ns; for mixed-sex samples, r(234) = −.17, p < .05. Thus, the birth cohort effect for girls disappeared when age was controlled, suggesting that the girls’ scores did not show birth cohort differences. On the other hand, the results consistently showed that boys’ CDI scores had decreased over time. Either a birth cohort or a period effect had occurred in boys’ CDI scores.

Using the regression equation and the average standard deviations from the samples in this study, we determined the birth cohort effect in terms of standard deviations for boys. (It is necessary to use this approach because the correlations between CDI means and year exaggerate the effect; they do not reflect the standard deviations of the individual samples.) The first birth year in the sample was 1964, when the average CDI score for boys was 9.70; the last was 1988, with an average CDI score of 8.36. With an average standard deviation of 7.20 (computed from the individual samples in the analysis), this represents a change of 0.19 standard deviations. Under Cohen’s (1977) guidelines, this is a small effect (defined as approximately .20). Overall, the results for age and birth cohort demonstrated an interaction with gender. Boys showed no overall age effect but a small birth cohort effect, whereas girls showed a small age effect but no birth cohort effect.

Discussion

We used meta-analytic techniques to analyze samples of children and adolescents responding to the CDI. We examined 310 samples for testing effects, gender, age, SES, race/ethnicity, and birth cohort. We now summarize and discuss these findings.

Testing Effects

One of the most wide-ranging implications of this analysis concerns testing effects on the CDI. These results demonstrate that longitudinal studies may be seriously confounded by significant decreases in scores over repeated administrations of the scale (see also Sharpe & Gilbert, 1998). In fact, the effect size for the decrease in longitudinal studies (d = .23) was just as strong as the increase in girls’ CDI scores at adolescence that was found using other methods. Thus, a longitudinal study examining girls would find no change in CDI scores, whereas other methods without this confound would suggest an increase in CDI scores over this time. This strongly suggests that a testing or measurement effect is at work.

We found that the drop in scores over time was relatively steady, with scores decreasing gradually over time. In addition, a comparison of studies with different lag times between waves suggests that repeated administrations are associated with the decrease in scores rather than the passage of time per se. Which explanation for testing effects do these results support? Habituation to the items (and/or boredom with the study) is consistent with the importance of repeated administrations rather than elapsed time. Habituation is also consistent with a gradual decline in scores over time, as respondents grow increasingly inured to the experience of completing the questionnaire. The results are less consistent with decreases in test anxiety, as this would be most likely to appear as a large decrease from the first administration to the second and little effect at further administrations. The results also do not support the coping mechanisms explanation (Sharpe & Gilbert, 1998). A realization that one is depressed and should therefore cope is most likely to occur after the first administration and appear by the second administration. However, we found that the decline in scores was gradual and not confined to the second administration. It seems unlikely that people make small changes to their lives as they gradually realize they are depressed. In addition, this explanation seems less plausible for children, who are less likely to realize their depression and take active steps to combat it. Finally, Sharpe and Gilbert (1998) argued that social desirability is unlikely to explain the effect, given that positive affect does not increase or decrease over repeated administrations. Our analysis concurs, given that a social desirability effect is also more likely to appear between the first and second administrations and then taper off. Thus, habituation emerges as the most likely explanation for the strong testing effect in longitudinal studies of CDI scores. It should be noted that these effects seemed to occur for many self-report measures of negative affect (Sharpe & Gil-
increase. Girls' symptoms were fairly steady until age 12, when they began to elevate at age 12. The CDI scores reached a peak at age 15 and remained high at age 16. Boys' depressive symptoms remained at about the same level regardless of age; the only exception was a slight elevation at age 12.

The pattern of gender differences in CDI scores is also noteworthy. Several authors have reported that boys score higher than girls on the CDI in the elementary school years (e.g., Angold & Rutter, 1992; Nolen-Hoeksema et al., 1991); however, this analysis suggests that this difference is very small ($d = -0.04$). Within each age group prior to age 13, the 95% CI for $d$ always included zero, suggesting that this was not a statistically significant gender difference. However, the gender differences during the adolescent years, when girls score higher, are statistically significant, beginning at age 13. This pattern of Age $\times$ Gender differences in levels of depressive symptoms confirms the results of qualitative reviews of the literature (e.g., Hankin & Abramson, 1999, 2001; Nolen-Hoeksema & Gigrus, 1994) but through an analytic technique that can produce much more reliable conclusions than a qualitative review. As noted earlier, this analysis also provided the unique contribution of a study without confounds from testing effects or birth cohort.

Many different biological and psychosocial explanations for the emergence of gender differences in depression have been proposed (Hankin & Abramson, 1999; Nolen-Hoeksema & Gigrus, 1994). Our results suggest that some of these explanations are more likely to be upheld than others. For example, some theorists have suggested that the transition to junior high school triggers depression in many girls (Simmons, Blyth, Van Cleave, & Bush, 1979). Our results suggest that the gender differences in depression emerge at age 13 (roughly between 7th and 8th grades) and thus more than a year after many students transition to junior high school. However, there may be other social influences that cause girls to experience higher levels of depression during that time. These results lend support to theories focusing on pubertal development. Many theorists have suggested that something about pubertal change contributes to more depression in girls (see Angold et al., 1999). Because pubertal changes occur over a span of several years, however, it has not been clear which of these many changes might contribute to more depression in girls. Our results suggest that mid- to late-pubertal change may be the most toxic to girls' emotional well-being and that changes during this period should be targeted for investigation. Consistent with this, Angold et al. (1999) found that mid- and late-pubertal changes in levels of testosterone and estrogen were most strongly related to increases in depressive symptoms in girls.

The gender differences in depressive symptoms are small effects, with the largest at .22 (Cohen, 1977, defines .20 as a small effect size). Some studies of adolescents have found larger gender differences in severe levels of depression than in mean scores that take into account the entire range of severity of depression (Angold, Costello, & Worthman, 1998; Lewinsohn, Hops, Roberts, Seeley, & Andrews, 1993; Silberg et al., 1999). In addition, most studies of adults have found gender differences in diagnosable depressive disorders on the order of 2 to 1, women to men (Nolen-Hoeksema, 2002). This suggests that the gender difference in depression is small (and unlikely to be clinically significant) in mild to moderate symptoms of depression but larger (and likely to be clinically significant) in more severe levels of depression. We could not break down the gender difference in CDI scores by different levels of severity because the necessary data were not available in most published studies. Clarifying differences in the magnitude of gender differences in depression at different levels of severity will be an important goal for future research.

It may be that the causes of gender differences in depression are different for different levels of severity. For example, gender differences in severe depression could be due to the experience of traumatic events, such as sexual abuse, which is both more likely to lead to severe depression than to mild depression and is over twice as likely to happen to adolescent girls than to boys (for reviews, see Nolen-Hoeksema, 2002; Weiss, Longhurst, & Mazure, 1999). In contrast, gender differences in milder depression could be due to gender differences in self-esteem, which might cause only moderate levels of depression and evidences gender differences of a similar magnitude to the gender differences in CDI scores found in this analysis (cf. Kling, Hyde, Showers, & Buswell, 1999).

Race/Ethnicity Differences

Examining race/ethnicity differences in CDI scores, we found that Hispanic children reported significantly more depressive symptoms than either White or Black children. This difference was the largest effect size in the meta-analysis ($d = .62$ comparing Hispanics and Whites, and $d = 1.31$ comparing Hispanics and Blacks). These results are consistent with the previous literature, which also found higher rates of depression in Hispanic populations (Blazer et al., 1994; Roberts et al., 1997). This meta-analytic treatment demonstrated that the ethnic differences are very strong: The CDI scores of Hispanic children were almost two thirds of a

Gender $\times$ Age Effects

One of the main goals of this analysis was to examine age differences in CDI scores without the confounds of testing effects or birth cohort. Because our analysis combined samples collected over 20 years and then controlled for year of data collection, it was effectively a cross-sectional study with the confound of birth cohort removed. It also summarized CDI data from a large number of sources (251 samples, including 43,916 children [21,016 boys and 22,900 girls], after excluding later waves of longitudinal studies).

What were the age differences in CDI scores when there was no testing effect and no confound for birth cohort? Girls' depressive symptoms were fairly steady until age 12, when they began to increase. Girls' CDI scores reached a peak at age 15 and remained high at age 16. Boys' depressive symptoms remained at about the same level regardless of age; the only exception was a slight elevation at age 12.

The pattern of gender differences in CDI scores is also noteworthy. Several authors have reported that boys score higher than girls on the CDI in the elementary school years (e.g., Angold & Rutter, 1992; Nolen-Hoeksema et al., 1991); however, this analysis suggests that this difference is very small ($d = -0.04$). Within each age group prior to age 13, the 95% CI for $d$ always included zero, suggesting that this was not a statistically significant gender difference. However, the gender differences during the adolescent years, when girls score higher, are statistically significant, beginning at age 13. This pattern of Age $\times$ Gender differences in levels of depressive symptoms confirms the results of qualitative reviews of the literature (e.g., Hankin & Abramson, 1999, 2001; Nolen-Hoeksema & Gigrus, 1994) but through an analytic technique that can produce much more reliable conclusions than a qualitative review. As noted earlier, this analysis also provided the unique contribution of a study without confounds from testing effects or birth cohort.

Many different biological and psychosocial explanations for the emergence of gender differences in depression have been proposed (Hankin & Abramson, 1999; Nolen-Hoeksema & Gigrus, 1994). Our results suggest that some of these explanations are more likely to be upheld than others. For example, some theorists have suggested that the transition to junior high school triggers depression in many girls (Simmons, Blyth, Van Cleave, & Bush, 1979). Our results suggest that the gender differences in depression emerge at age 13 (roughly between 7th and 8th grades) and thus more than a year after many students transition to junior high school. However, there may be other social influences that cause girls to experience higher levels of depression during that time. These results lend support to theories focusing on pubertal development. Many theorists have suggested that something about pubertal change contributes to more depression in girls (see Angold et al., 1999). Because pubertal changes occur over a span of several years, however, it has not been clear which of these many changes might contribute to more depression in girls. Our results suggest that mid- to late-pubertal change may be the most toxic to girls' emotional well-being and that changes during this period should be targeted for investigation. Consistent with this, Angold et al. (1999) found that mid- and late-pubertal changes in levels of testosterone and estrogen were most strongly related to increases in depressive symptoms in girls.

The gender differences in depressive symptoms are small effects, with the largest at .22 (Cohen, 1977, defines .20 as a small effect size). Some studies of adolescents have found larger gender differences in severe levels of depression than in mean scores that take into account the entire range of severity of depression (Angold, Costello, & Worthman, 1998; Lewinsohn, Hops, Roberts, Seeley, & Andrews, 1993; Silberg et al., 1999). In addition, most studies of adults have found gender differences in diagnosable depressive disorders on the order of 2 to 1, women to men (Nolen-Hoeksema, 2002). This suggests that the gender difference in depression is small (and unlikely to be clinically significant) in mild to moderate symptoms of depression but larger (and likely to be clinically significant) in more severe levels of depression. We could not break down the gender difference in CDI scores by different levels of severity because the necessary data were not available in most published studies. Clarifying differences in the magnitude of gender differences in depression at different levels of severity will be an important goal for future research.

It may be that the causes of gender differences in depression are different for different levels of severity. For example, gender differences in severe depression could be due to the experience of traumatic events, such as sexual abuse, which is both more likely to lead to severe depression than to mild depression and is over twice as likely to happen to adolescent girls than to boys (for reviews, see Nolen-Hoeksema, 2002; Weiss, Longhurst, & Mazure, 1999). In contrast, gender differences in milder depression could be due to gender differences in self-esteem, which might cause only moderate levels of depression and evidences gender differences of a similar magnitude to the gender differences in CDI scores found in this analysis (cf. Kling, Hyde, Showers, & Buswell, 1999).

Race/Ethnicity Differences

Examining race/ethnicity differences in CDI scores, we found that Hispanic children reported significantly more depressive symptoms than either White or Black children. This difference was the largest effect size in the meta-analysis ($d = .62$ comparing Hispanics and Whites, and $d = 1.31$ comparing Hispanics and Blacks). These results are consistent with the previous literature, which also found higher rates of depression in Hispanic populations (Blazer et al., 1994; Roberts et al., 1997). This meta-analytic treatment demonstrated that the ethnic differences are very strong: The CDI scores of Hispanic children were almost two thirds of a
standard deviation higher than those of White children, and more than one standard deviation higher than those of Black children. This is most likely a clinically significant difference because a one-standard-deviation change shifts a score from the 50th percentile to the 84th percentile. It is also important to realize that this difference is probably based in culture and/or life situation, given that individuals of Hispanic origin may be of any race. The paucity of previous research on depressive symptoms in ethnic minority children means that researchers and mental health providers have little understanding of why Hispanic children would have elevated levels of symptoms. Investigating the sources of these group differences in depressive symptoms will be an important focus for future studies.

**Birth Cohort**

Our findings on birth cohort change were surprising because previous studies have suggested that depression would increase in successive birth cohorts (e.g., Klerman & Weissman, 1989). Instead, boys’ scores decreased from 1980 to 1998 and girls’ scores did not change. The shift in boys’ scores was 0.19, a smaller birth cohort effect than found in most studies (for example, the increase in anxiety from 1952 to 1993 was 0.99; Twenge, 2000). It is possible that levels of distress had reached a peak and had begun to level off or decrease, a conclusion consistent with a community sample study that found small decreases in depression from 1970 to 1992 (Murphy, Laird, Monson, Sobol, & Leighton, 2000). The results here are also consistent with changes in children’s self-esteem, which decreased from 1965 to 1979 but increased from 1980 to 1994 (Twenge & Campbell, 2001). The negative correlation between self-esteem and depression (Crandall, 1973; Pyszczynski & Greenberg, 1987; Tennen & Herzberger, 1987) suggests that children’s depression should have decreased after 1980. Given the large changes in depression and anxiety during the 1960s and 1970s, it could be that girls’ depression has plateaued at a high level. Boys’ depression may have decreased slightly from a high level in the late 1970s. This would be consistent with the pattern of social change for this era; most negative social influences peaked during the late 1970s and early 1980s (Twenge & Campbell, 2001), although a few others continued to climb during the 1980s and 1990s (e.g., crime).

The lack of change for girls suggests that more negative social forces balanced out positive forces. For example, perhaps positive changes in girls’ career choices and athletic participation have been countered by increases in body image concerns. For example, Feingold and Mazzella (1998) found that gender differences in body image increased over time, with more girls and women demonstrating negative body image since the 1970s. Future research should address possible explanations for the birth cohort trends found here. In addition, future research should examine birth cohort changes in children’s anxiety after 1980 to determine if anxiety decreased or plateaued as depression has.

**Limitations**

One limitation of this article is that it examined data only for the CDI, a scale that may measure negative affect as much as it measures depression (e.g., Saylor et al., 1984). Thus, many of the conclusions here may apply to negative affect more broadly and not as specifically to depression. Yet, as noted, depressive symptoms are highly comorbid with other symptoms of distress in children (Kovacs & Devlin, 1998). In addition, because the CDI is not designed to provide diagnoses of depression, we cannot know whether our results would generalize to diagnoses of depression. Throughout the discussion, however, we have noted where our results are similar or dissimilar to individual studies focusing on depressive diagnoses.

**Conclusion**

Meta-analyses provide information about reliable trends across many studies of different samples taken at different times in recent history. The analysis presented here helped to clarify the relative effects of age, gender, birth cohort, race/ethnicity, and repeated questionnaire administration on CDI scores in children and adolescents. This study summarizes data from a large number of respondents, providing normative scores for the CDI that are unconfounded by testing effects or birth cohort. Perhaps even more important, investigators can use these results when designing new studies that explore the sources of these effects. The norms and effect sizes provided here should serve as a springboard to better understand why Hispanic children suffer more depressive symptoms and why girls become more depressed around age 13. These analyses also strongly suggest that future research needs to address the substantial testing effect in longitudinal studies and find ways to minimize this apparent bias.

**References**


Received May 21, 2001
Revision received February 27, 2002
Accepted March 12, 2002