Children's Depression Inventory 2nd Edition (CDI 2) Test Critique

Attallah College of Educational Studies, Chapman University

Test Description

The Children's Depression Inventory 2nd Edition (CDI 2) by Maria Kovacs, Ph.D., was

published in 2011 by Multi-Health Systems Inc. The CDI 2 assesses the extent of depressive

symptoms in children and adolescents from ages 7 to 17. More specifically, the CDI 2 self-report

full-length (CDI 2:SR) form evaluates the symptoms of major depressive disorder (MDD) and

dysthymic disorder (DD) under the Diagnostic and Statistical Manual of Mental Disorders,

fourth edition (DSM-IV), criteria. The CDI 2 gathers self-rated data from many respondents

(student, parent/caregiver, and the teacher) about the student's depressive symptoms. This

comprehensive toolkit is helpful for identifying youth that are at-risk for depression and

evaluating response to treatment. The original CDI was published in 1977. The updated CDI 2

includes three new items on the CDI 2:SR, three revised items, two deleted items, and the CDI 2

is more demographically representative of the 7- to 17-year-old population of the United States.

Content and Use

The CDI 2 includes four different forms: the CDI 2:SR, which contains 28 items; the

Short CDI 2: Self-Report (CDI 2:SR[S]), which contains 12 items; the CDI 2: Parent (CDI 2:P),

which contains 17 items; and the CDI 2: Teacher (CDI 2:T), which contains 12 items. The CDI

2:SR has the longest administration time of only 15 minutes, followed by the CDI 2:P taking 10

minutes, and the CDI 2:SR[S] and the CDI 2:T both taking 5 minutes. The test can be

administered in a paper-and-pencil format or a computerized format. The tool kit can be

hand-scored, entered into the computer for online scoring, or by scoring software.

Both formats can be purchased online from the publisher, Multi-Health Systems. There

are three CDI 2 kits sold online ranging from $319 - $469. It includes items such as the test

manual and each of the four forms in packs of 25. Each hand-scored form can also be bought

separately in a pack of 25 forms for $67, and individual online forms can be purchased for $3.75 each. The CDI 2 manual costs $97, and the scoring program alone costs $151. Compared to the original CDI cost, it appears that the forms have increased in price, such that when purchased in bulk, the cost is about $2.70 per form for the CDI 2 and each additional online form is $3.75 per form.

The full-length CDI 2:SR is scored using two scales: Emotional Problems and Functional Problems and four subscales: Negative Mood/Physical Symptoms, Negative Self-Esteem, Interpersonal Problems, and Ineffectiveness. Only the full-length CDI 2:SR covers symptom characteristics of both MDD and DD. The Emotional Problem scale measures dysphoric affect (e.g., sadness) and neurovegetative symptoms (e.g., insomnia and decreased appetite). The Functional Problem scale measures problems with peers, school, and family interactions (e.g., number of friends). The four subscales further measure depressed mood, functional consequences of depressive symptoms, negative self-worth, and interpersonal relationship problems. The CDI 2: SR[S] has no scale scores or subscale scores but assesses affective, cognitive, and neurovegetative characteristics of depression. The CDI 2: SR[S] is ideal for screening purposes (Kovacs, 2011). The CDI 2:SR and CDI 2:SR[S] have item sets containing three items each and are organized in a 3-point ordinal scale that reflects the severity of depressive symptoms (0 = *none* to 2 = *definite*). Similarly to the CDI 2:SR, the CDI 2:P and CDI 2:T include Total Scores, Emotional Problems scale scores, and Functional Problems scale scores, but does not include subscale scores. The CDI 2:P and CDI 2:T items are centered on depressive symptoms that a depressed child is more likely to show or express. Participants respond to a 4-point Likert scale (0 = *not at all* to 3 = *much or most of the time*) on the CDI 2:P and CDI 2:T.

Each test form is separate. At the top of each form in all capital letters is the name of the form, and each form is a different color, which helps the test administrator and interpreter to keep the forms and responses separate from each other. There are designated spaces for the respondents' demographic information in a box at the top of each form. This separate box makes it easier for administrators to check when the respondents are finished that all the information is filled out. Also, a space for the child's name is included on all forms, making it easier to keep the separate forms together. For the self-report forms (CDI 2:SR and CDI 2: SR[S]), each item is separated into its own outlined box with three answer choices. The respondent selects which option describes them best within the past two weeks by marking one box with an "X." Both self-report forms include an example item for students to practice before beginning. Parents and teachers are instructed to circle their answers and put an "X" through any mistakes and circle their corrected answers for the CDI 2:P and CDI 2:T. Furthermore, the CDI 2:P and CDI 2:T items are distinguished by alternating in color, which visually helps a respondent differentiate each item. For example, each item on the CDI 2:T alternates between a shaded dark blue and a shaded light blue. Parents and teachers are asked to circle the number that corresponds best with the description of their observations of their child or student within the past two weeks. The respondents are asked to rate their experience with or observations of depressive symptoms over the past two weeks. The author mentions that a more extended period would "strain the respondents' memory and their reliability as informants" (Kovacs, 2011, p. 12). However, is it possible that the child's severity of depressive symptoms fluctuates within two weeks? Therefore, just asking about their past two weeks' experience may not represent the whole picture of the child's depression symptoms.

The items on the inventory are not cognitively demanding for the respondents. Each item

has only three to four short choices in simple language and does not require a high reading level. A strength of the CDI 2 is the low reading level, which ranges from first-grade to second-grade reading levels. The highest reading level required is 2.2 reading grade level for the CDI 2:T. If there is a concern that the respondent has poor reading skills, it is suggested for a valid self-report that the assessor read the items aloud while the respondent reads along simultaneously. To ensure honest and accurate responses, the author does not recommend that the CDI 2:SR is used if children and adolescents are reluctant to respond to a questionnaire, disoriented, severely impaired, or not proficient in English or the language the CDI 2 is translated in. It is unclear what languages the CDI 2 has been translated in other than Spanish.

The flexibility of testing settings is a strength of the CDI 2 tool kit. All four forms can be administered in person or remotely. The self-report forms can be completed in an individual or group setting. The parent-rated and teacher-rated forms are less likely to be given in a group setting compared to an individual setting. However, each form must be answered entirely in one sitting and in a quiet place free of distractions. It is recommended that the administrator specify a return date and include necessary supplies (e.g., a stamped and addressed envelope) if the test will be completed remotely and returned at a later date. Also, parents or legal guardians should be contacted by the administrator to receive informed consent. This flexibility of testing settings helps a test administrator efficiently collect data.

The CDI 2 took kit is not intended to diagnose depression in children and adolescents; however, it can be used in combination with other assessments for decisions regarding diagnosis, treatment, and intervention. CDI 2 users can range from school psychologists to physicians. An administrator without clinical psychology or psychometrics training may use it as long as they closely conduct the procedures outlined in the manual. Only a qualified professional is allowed

to interpret the results. Interpreters of the CDI 2 are required to have "MHS b-level qualifications." Their qualifications include that they at least have taken measurement and testing courses at the graduate level, they are familiar with the test manual and testing standards, and they belong to professional associations or are licensed professionals in fields such as psychology, education, and social work.

The manual is set up in an organized, clear, and user-friendly format. The test manual begins with a table of contents page where one could easily search for the topic and page number to skip to the section of the manual they need. Six chapters cover the background of the CDI 2, research and history of the CDI, administration and scoring, interpretation, psychometric properties, and concluding comments. The title of each chapter and section heading has large and bolded font to help guide readers. There are also tables within each chapter that summarizes information related to that chapter and examples for scoring help. The manual is written in straight forward language; however, it may be hard to interpret for readers without some knowledge in statistics and psychometrics. There is also an index at the back of the manual to refer to for help.

Standardization Sample and Norms

The CDI 2:SR and CDI 2:SR[S] standardized sample included 1,100 7- to 17-year-olds from 28 different states in the U.S. representing the four major geographic regions of the U.S. (Northeast, Midwest, West, and South). The number of participants from each state and region is not specified. Age and sex were standardized by quotas for each age group. The sample size consisted of 50 females and 50 males for each age group. Although a sample size of 1,100 is sufficient, having 200 respondents total from each age group would have been more nationally representative. Each group was stratified to match U.S. Census data (within 5%) regarding race

and ethnicity. Each group of 50 males or females had about 2 Asians, 8 African Americans, 7 Hispanics, 31 Whites, and 2 children of other races or ethnicities. The author did not specify how many of the 1,100 participants completed the full-length or short version of the CDI 2 Self-Report. For the CDI 2:SR and CDI 2:SR[S], a clinical sample of 319 children aged 7- to 17-years-old with a variety of DSM-IV diagnoses (Major Depressive Disorder [MDD], Attention-Deficit/Hyperactivity Disorder [ADHD], Conduct Disorder/Oppositional Defiant Disorder [CD/ODD], and Generalized Anxiety Disorder [GAD]) were also assessed. Within the clinical sample, the groups of youth diagnosed with ADHD, CD/ODD, and GAD all had less than 100 participants, limiting the generalizability. For some analyses (e.g., internal consistency), the standardized sample and the clinical sample were combined, which the test developers referred to as the *total sample*.

The standardization sample for the CDI 2:P included 800 adults and 600 adults for the CDI 2:T. In these evaluations from parents and teachers, the characteristics of the rated youth had equal total amounts of males and females in the 7- to 17-year-old age range, and the race and ethnicity are similar (within 5%) to that of the U.S. population defined by the U.S. Census. Although there was an equal number of boys and girls within the 7- to- 17-year-old age range, there was not an equal number of students represented in each age group. For example, for the CDI 2:P and CDI 2:T, age was separated into groups (e.g., 7-9), and within those age groups, males and females were not equally represented (e.g., 75 males and 60 females). Also, there are less than 100 children and adolescents represented in each age group. For the 600 teachers, they all knew their rated student for at least one month. The majority (86.8%) of the standardization sample of parent-rated forms were completed by biological mothers, which introduces bias to the ratings.

Overall, each test's norm samples are nationally representative in terms of age, geography, race/ethnicity, and gender. A shortcoming of the standardized samples was that the participants' education level and socioeconomic status were not specified. Also, other than how long the teachers knew the student for and identifying the parent's relationship to the child, characteristics of the adults of the standardization samples were not shared. Information including the adults' (parents and teachers) age, educational attainment, socioeconomic status, language, and race/ethnicity would ensure better generalizability of the results. Readers are only given the age, sex, and race/ethnic distribution of the children rated for the CDI 2:P and CDI 2:T evaluations. Given that the CDI 2 forms were standardized with U.S. Census data from 2000, it is likely that the data has changed now in 2020. Therefore, test administrators and interpreters should keep in mind that this data might not necessarily represent today's youth.

Scores and Interpretation

As previously mentioned, there are three different methods for scoring the CDI 2 tool kit. It can be hand-scored using QuikScore, scored online with CDI 2 Scoring Software, or scored online using the MHS Online Assessment Center. The manual includes a table in chapter 3 that outlines the maximum number of items that can be omitted or missed for the results to be valid and reliable. For missing or omitted items, the raw scores are prorated. This allows for the consideration of missing or omitted items and estimates the total raw score based on the other items the respondent selected for that specific scale.

The CDI 2 manual provides a step-by-step procedure with detailed instructions on scoring the CDI 2 tool kit. For all CDI 2 forms, a total raw score is calculated and converted into a standardized T-score. For the full-length CDI 2:SR, each item is scored ranging from 0 to 2, and each item corresponds to a specific letter. There is one letter for each of the four subscales

(Negative Mood/Physical Symptoms, Ineffectiveness, Negative Self Esteem, and Interpersonal problems). The Raw Scores for the subscales are obtained by totaling each item associated with the same letter and entering it into the box labeled for that letter. To find the Raw Score for the two scale scores (Emotional Problems and Functional Problems), values for two of the subscales are summed (i.e., Raw Score of Box A+ Raw Score Box B = Emotional Problems). The overall Total Raw Score is then obtained by adding the Raw Scores of the Functional Problems and Emotional Problems together. It is important to note that some of the items are listed in reverse order on the response sheet. Although the scoring sheet accounts for this by listing the raw score numbers in reverse order, the administrator should be cautious when summing the total raw score. Upon review, I incorrectly counted 2 instead of 0 in the example provided in the manual because of the reverse order, so the scorer should pay close attention as it is an easy mistake.

After obtaining the raw scores, a Profile Page is used to convert the Raw Scores into standardized T-scores. For scoring, the test administrator must ensure that the respondent filled out all of the items, especially the child's sex and age, in the demographic box at the top of each form. The administrator has to use the correct Profile Page for the child's sex (M/F). On the Profile Page, Raw Scores are circled in the appropriate scale column and matching the child's age group (7-12 or 13-17). There are two T-Score columns on the right and left of the table where administrators can identify which T-score matches each Raw Score. One column for T-score classifications (Average or Lower, High Average, Elevated, and Very Elevated) is also included. There are boxes at the bottom of each column to enter the T-scores. The circled values in each column are connected with straight lines to form a profile. This profile formation feature is not included on the CDI 2:SR[S]. Profiles are helpful for information on treatment progress for children who have been administered the CDI 2 multiple times.

Raw scores and T-Scores are obtained in the same way for the CDI 2:SR[S], CDI 2:T, and CDI 2:P. The only difference is that the CDI 2:SR[S] only includes one Total Raw Score. The CDI 2:T and CDI 2:P include Raw Scores for Emotional Problems, Functional Problems, and Total Raw Score. Additionally, for the CDI 2:T and CDI 2:P, the raw scores range from 0 to 3 instead of 0 to 2.

A limitation of the Profile Page is that it is categorized by the nominal variable of gender, where the child's results are interpreted by either being male or female. With an increase in visibility of other gender categories, this may cause confusion for interpreting the results of students who identify as non-binary, for example. Interestingly, appendix C in the test manual includes tables of combined-sex norms to examine T-scores regardless of sex (Kovacs, 2011). The author does not specify what circumstance this table would be used, but this may be useful for cases where a child identifies as non-binary. For scoring, the risk of error seems minimal such that for the hand-scored option once the respondent marks an item on the response sheet is copied directly onto the scoring sheet. Similarly, this chance of error also seems low for respondents that complete the form online as the computer will automatically score their answers. However, in cases where the administrator manually inputs the respondent's answers onto the physical or digital score sheet, the chances of error increase. Additionally, the administrator does not need to calculate the T-scores themselves, which reduces the chance of error. Overall, the scoring seems straightforward, and no advanced calculations need to be made by the administrator. The administration instructions and procedure laid out in the manual make it easy for one without formal training to follow.

The Raw Scores gathered from each item of one of the CDI 2 forms indicate the level of depressive symptoms. For example, a score of 20 would indicate a higher level of depression

symptoms than a score of 10. However, scores cannot be compared between respondents because the Raw Scores include different numbers of items. For comparison, T-scores are used. Therefore, Raw Scores are converted into a standardized T-score using a normal distribution with a mean of 50 and a standard deviation of 10. The CDI 2 manual provides a table that breaks down the T-scores, percentile ranks, and their meaning. For example, a T-score of 70 and above is considered a "Very Elevated" score, and T-scores ranging from 40-50 are considered "Average" scores.

Percentile ranks can also be used to interpret the CDI 2 scores such that they exemplify the percentage of children in the normative sample who scored equal to or less than the respondent's raw score (Kovacs, 2011). Percentiles are used to compare the respondent's scores to peers of similar age and the same sex. A higher percentile score on the CDI 2 suggests that the respondent is experiencing more depression symptoms than their peer group. Problems with using percentile ranks are that they only tell us where a child stands in relation to other students, interpreters are unsure of the score differences, and parents may interpret them as percentages. In contrast, explaining results in terms of percentile ranks may be more digestible for parents.

The classifications suggested using a T-score or percentile rank help inform depression symptoms and areas of functional impairment. For example, a child with a T-score of 70 or above is considered to have "many more concerns than are typically reported" (Kovacs, 2011, p. 44). In the CDI 2 manual, the author explicitly reminds interpreters that the cutoff scores for the above T-score ranges and percentile ranks in the manual are not absolute rules, and interpretation of the results' significance should involve clinical judgment. For example, the difference between a T-score of 64 and 65 could be arbitrary, so taking extra caution when considering classification is necessary for interpreters. However, a strength of the scoring and interpretation

of the CDI 2 is that confidence intervals and standard error of measurement were considered.

The default setting on the scoring software is set to display 90% level of confidence intervals for each T-score, but it can be changed to show 95% confidence intervals (Kovacs, 2011). For the hand-scoring option, the manual includes the 90% and 95% confidence intervals for each T-score in appendix E.

The CDI 2 manual also outlines a step-by-step procedure to follow for the interpretation of scores. The first step is to interpret the Total Score, scale scores, or scale scores. It is important to explore the scales and subscale scores because the Total Score may be in the Average to High Average range, but one scale score or subscale score may reveal elevated T-scores. Elevated scale scores or subscales scores can lead to further investigation of areas of concern for the child, such as interpersonal problems. The second step is to assess the child's overall profile and the scores in reference to each other. The third step is to examine the individual items of the form concerning an elevated T-score. The fourth step is to integrate and compare the results from the individual CDI 2 assessments in addition to information from outside sources such as observations and interviews. Lastly, the practitioner should decide the most appropriate format to share the results with the parents, including a reminder about how scores are compared and what elevated or low scores could indicate for their child.

Overall, interpretation of the results seems more involved, and as the manual suggests, additional sources of information should be used to use next-step decisions. It makes sense that someone with more formal training should be responsible for interpreting the results as there is a lot to consider and to ensure the most accurate decision for the child's symptomatic status

Psychometric Properties

Reliability

The CDI 2 manual only mentions reliability and validity measured using the standardization sample and the total sample. Reliability refers to how consistent scores remain by the same person when they are readministered the same test or inventory on different occasions, under different conditions, or with different sets of comparable items (Anastasi, 1998, p. 102, as cited in Kovacs, 2011). The test developers analyzed internal consistency, test-retest reliability, and the standard error of measurement (SEM).

Internal Consistency

Internal consistency refers to how reliably the inventory items measure the same construct (Kovacs, 2011). Internal consistency of the CDI 2:SR examined the total sample using Cronbach's alpha of the Total Score, two scales, and four subscales. The Total Score alpha score of 0.91 is adequate for education and clinical purposes. Less reliable but still adequate are the Emotional Problems scale consisting of 15 items (α =.85) and the Functional Problems scale consisting of 13 items (α = .83). The manual also includes the alpha scores for the four subscales, which are less reliable with α coefficients ranging from 0.73 to 0.77. However, for subscales with a limited number of items, a coefficient of 0.70 is sufficient. Interpreters should be cautious when using the scale scores and subscale scores for educational purposes because measurement error is increased.

Internal consistency of the CDI 2:SR[S], CDI 2:T, and CDI 2:P was also examined. Kovacs (2011) mentions that the CDI 2:SR[S] is useful as a screening measure; therefore, its Total Score Cronbach alpha coefficient of 0.82 is adequate. For the CDI 2:T, the Total Score consisted of 12 items (α =.89), the Emotional Problems scale consisted of 6 items (α =.82), and Functional Problems scale consisted of 6 items (α =.85). The CDI 2:P Total Score consisted of 26 Items (α =.88), the Emotional Problems scale consisted of 12 items (α =.86), and the

Functional Problems scale consisted of 14 items (α =.79). It is concerning that the internal consistency alphas for the Total Scores of the CDI 2:T and CDI 2:P were both below 0.90. Also, the alpha score of the Functional Problems scale of the CDI 2:P just missed the guideline of 0.80. Interpreters should note that these forms are less reliable than the full-length CDI 2:SR.

Cronbach alpha scores were also included for the individual age and sex groups. Although the test manual highlights that the Total Score alphas were high (ranging from 0.88 - 0.92) with the individual age and sex groups for the CDI 2:SR, reliability is questionable when taking a closer look at the alpha scores of the scales (Emotional Problems and Functional Problems) and the subscales (Kovacs, 2011). Generally, the alphas for the scales of the CDI 2:SR were at 0.80 or above, except for the Functional Problems scale for the female, 7- to 12-year-old group (α = 0.79). Generally, the alphas for the subscales were at 0.70 or above, except for the Negative Mood/Physical Symptoms subscale for the female, 7- to 12-year-old group (α = 0.69) and the Interpersonal Problems subscale for the female, 7- to 12-year-old group (α = 0.67). Internal consistency among individual age and sex groups for the CDI 2:SR[S] was adequate as the guideline for screeners is 0.80. However, the alpha scores for the female, 7- to 12-year-old age group (α = 0.77) and the 7- to 12-year-old age group (α = 0.79) were the least reliable.

For the CDI 2:T, the Total Score's internal consistency among individual age and sex groups ranged from 0.86 to 0.91, and for the CDI 2:P, it ranged from 0.85 to 0.91. Taking a closer look at the scale scores for both of these forms, most sex and age groups had alphas of 0.80 and above with four exceptions. In the Emotional Problems scale of CDI 2:T, the female, 7- to 12-year-old age group had an alpha of 0.79. In the Functional Problems scale of the CDI 2:P, for the 13- to 17-year-old age group, the alpha was 0.75; for the male, 13- to 17-year-old age group, the alpha was 0.79; and for the female, 13- to 17-year-old age group, the alpha was 0.69.

Interpreters should consider their clients' age and sex, especially if they match these sex or age groups with less internal consistency. They should be cautious about making educational decisions based on these groups since they contained the most measurement error.

Test-retest Reliability

Test-retest reliability, the stability of test scores across more than one test administration, was determined for the CDI 2:SR and CDI 2:SR[S] (Kovacs, 2011). A dependent t-test was conducted with 79 children from the standardization sample, which is a small sample size. The adequate sample size for test-retest reliability should be at least 100 participants. Participants completed the inventory twice within the time frame of two to four weeks. The mean period between when the first test was taken, and the retest was 16.1 days. Because the CDI 2 assesses a child's state, which is temporary compared to a trait, the test developers looked at the short-term stability of scores. The CDI 2:SR Total Score ($r = .89$) met appropriate guidelines of 0.70 - 0.80 and above as did the CDI 2:SR[S] ($r = .92$). For the scale and subscales, the correlations ranged from 0.76 – 0.92, indicating high temporal stability. However, the sample size reduces the adequacy of reliability. Unfortunately, test-retest reliability was not determined for the CDI 2:T and CDI 2:P. The test manual cites research by Kovacs on examining parent-rated CDIs for at-risk children, but this data is unpublished and did not utilize the same CDI 2 parent-rated form. While the test-retest reliability proves stable for the CDI 2:SR and CDI 2:SR[S], the sample size is small, and evidence of test-retest reliability for the whole CDI 2 tool kit is insufficient. More research would support the whole tool kit's temporal stability.

Overall, the literature on psychometric properties of the CDI 2 is limited. One study conducted by Kim et al. (2018) evaluated the psychometric properties of the CDI 2 in a community-based sample of youth from Korea. The CDI 2 tool kit was translated into Korean

with the publisher's consent. Test-retest reliability was evaluated using a dependent t-test. The sample included 96 students in grades 4 through 6 from one elementary school in Korea. They completed the CDI 2:SR twice within four weeks. The results indicated a Pearson correlation coefficient of 0.84, which meets the guideline of 0.70 and above. While this study offers further support of the reliability of the CDI 2, differences in the translated CDI 2, the small sample size, and the sample's demographics limit the generalizability of the results.

Standard Error of Measurement (SEM)

The test developers determined T-score-based standard error of measurement (SEM). SEM reflects how likely a child's given score mirrors what their "true" score would be (Kovacs, 2011). For a normative sample with a mean of 50 and a standard deviation of 10, a SEM of 3 is adequate. The SEM was assessed for individual age and sex groups. The T-score-based SEMs for all Total Scores, scales, and subscales of the CDI 2:SR, CDI 2:SR[S], CDI 2:T, and the CDI 2:P were above a 3. Therefore, they are considered to be reliable. For example, the lowest SEM was of the CDI 2:T form for respondents aged 7-12, which had a SEM of 3.33.

An area of improvement for the CDI 2 tool kit's reliability would have been to provide interrater reliability because the CDI 2 is a rating scale. Overall, the CDI 2 shows favorable evidence of its reliability, mainly internal consistency, but without more external research users should be cautious when interpreting results.

Validity

Validity refers to how well an inventory measures what it claims to measure (Kovacs, 2011). Evidence for discriminative validity and congruent validity is included in the CDI 2 manual.

Discriminative Validity

Discriminative validity is used to show that depression should not be related to measures they are not expected to be related to (Kovacs, 2011). Discriminate validity refers to how well the CDI 2 can discriminate children with MDD from children without an MDD diagnosis. To assess the CDI 2 self-report forms' discriminative validity, analyses determined how the CDI 2 scores differed within the diagnostic groups (MDD, GAD, CD/ODD, and ADHD) and the control group (from the standardization sample). Only the MDD and control groups had a sample size of over 100 participants and were matched in terms of sex, age, and race/ethnicity. This indicates that the validity is reduced for the other diagnostic groups. Planned comparisons ($p <$ .01) revealed that the MDD group scored significantly higher than the other groups for the Total Score, the Emotional Problems Scale, the Negative Mood/Physical Symptoms subscale, and the Negative Self-Esteem subscale in the CDI 2:SR. No significant differences were found between the MDD group and the GAD group in the Functional Problems scale, Ineffectiveness subscale, and Interpersonal Problems subscale; however, this is expected as functional problems anxiety disorders coexist with depression often. For the CDI 2:SR[S], the results indicated that the MDD group revealed having significantly more symptoms than the other groups. These results indicate that the CDI 2 self-report forms can differentiate children with a diagnosis of MDD well, yet it is unknown how well the CDI 2:T and CDI 2:P can.

The test developers also assessed sensitivity (the ability of the CDI 2 to identify MDD cases correctly) and specificity (the ability of the CDI 2 to identify cases of other diagnoses or control cases correctly) (Kovacs, 2011). Only the self-report forms of the CDI 2 were assessed. The previously mentioned sample was used, including children with one of four different DSM-IV diagnoses (from the clinical sample) and control cases (from the standardization sample). The guideline for sensitivity and specificity is 80%. When looking at the sensitivity

percentages of how well the CDI 2 identified MDD compared to the control group, the 80% guideline was met. For the Total Score, Emotional Problems scale, and the Functional Problems scale of the CDI 2:SR and the Total Score of the CDI 2:SR[S], the sensitivity percentages ranged from 83.2%-84.1%. However, specificity for these groups was below the guideline ranging from 73.3% - 77.4%. The sensitivity and specificity of comparing MDD versus other DSM-IV diagnoses did not meet the guideline either. Sensitivity ranged from 68.2% - 77.6% and specificity ranged from 68.5% - 74.2%. Practitioners should note the increased risk of false negative and false positive results when interpreting the CDI 2 self-report forms.

Congruent Validity

Kovacs (2011) defined convergent validity as the extent of how well CDI 2 scores correlate with scores from other tests that also measure depression. However, it appears that the test developers have mislabeled convergent validity, and instead, it should be congruent validity. Congruent validity explores similar tests and how they correlate with each other, whereas convergent validity explores the association between related constructs (e.g., depression and anxiety). In this case, the construct of depression was assessed in each test, and the test developers were comparing the association between tests. The participants included 266 youth from the standardization sample and clinical sample. A majority of the participants were White (82%), and 142 were females. Future studies should consider having a more diverse sample for better generalizability. In addition to completing the CDI 2:SR, they were also assigned either the Beck Depression Inventory-Youth version (BDI-Y; Beck et al., 2001) or the Conners Comprehensive Behavior Rating Scales (Conners CBRS; Conners, 2008) to complete. For congruent validity, the correlation guideline is 0.7 at a minimum. The correlations between the CDI 2 and the BDI-Y ranged from 0.28 to 0.37, and the correlations between the CDI 2 and the

Conners CBRS (specifically its Major Depressive Episode scale) ranged from 0.38 to 0.59.

Interpreters should note that although these correlations are statistically significant, they did not

meet the correlation guideline of at least 0.7. It is concerning that the correlations between these

similar tests are not higher.

Confirmatory factor analyses were conducted for the CDI 2:SR, CDI: T, and CDI 2:P.

The new conceptual model grouped highly correlated items into their respective scales

(subscales, scales, and Total Score) and was found to "fit the data very well" (Kovacs, 2011, p.

60). The test manual briefly mentions that previous research established the validity of the

original CDI teacher and parent forms; however, updated validity studies should be conducted

for the new versions of the CDI 2:T and CDI 2:P. It is essential to assess the validity of the CDI

2:T and CDI 2:P forms to ensure that these forms measure what they claim to be measuring.

Lastly, the test developers should consider examining content validity and criterion validity as

they were not mentioned in the test manual.

Conclusions

As mentioned throughout the manual, this inventory appears to be useful as one part of a

comprehensive evaluation and as a screener for identifying at-risk youth. The test manual's

layout is organized, includes straightforward language, and provides easy to follow scoring

procedures. However, the CDI 2 has many limitations, which leads to less confidence in using

this as an effective measure of depression symptoms in children and adolescents. Areas of

improvement include updating the standardization sample to make it more representative of

today's U.S. population. While the standardization sample may have matched the U.S. Census

data, the data from some samples used for assessing the reliability and validity of the CDI 2 were

not representative. The data revealed that the reliability and validity of the CDI 2 are not as

strong as the manual claims it to be. The CDI 2 tool kit lacks sufficient evidence for reliability

and validity, and no new tests for the CDI 2:T and CDI 2:P were conducted. Additionally,

outside research on the CDI 2 tool kit is limited even though it has been published for almost a

decade. I would suggest that if used, it should be interpreted with its limitations in mind.

References

Beck, J. S., Beck, A. T., & Jolly, J. B. (2001). *Beck Depression Inventories—Youth (BDI-Y)* [Manual]. The Psychological Corporation.

Conners, C. K. (2008). *Conners Comprehensive Behavior Rating Scales (Conners CBRS)* [Manual]. Multi-Health Systems.

Kim, H. J., Lee, E. H., Hwang, S. T., Hong, S. H., & Kim, J. H. (2018). Psychometric properties of the Children's Depression Inventory-2 among a community-based sample of Korean children and adolescents. *Korean Journal of Clinical Psychology*, *37*(2), 177-186. https://doi.org/10.15842/kjcp.2018.37.2.005

Kovacs, M. (2011). *Children's Depression Inventory (2nd ed.) (CDI 2)* [Technical Manual]. Multi-Health Systems Inc.