A Depressive Symptom Scale for the California Psychological Inventory: Construct Validation of the CPI-D

Meg Jay and Oliver P. John
University of California, Berkeley

To facilitate life span research on depressive symptomatology, a depressive symptom scale for the California Psychological Inventory (CPI) is needed. The authors constructed such a scale (the CPI-D) and compared its psychometric properties with 2 widely used self-report depression scales: the Beck Depression Inventory and the Center for Epidemiological Studies Depression Scale. Construct validity of the CPI-D was examined in 3 studies. Study 1 established content validity, classifying CPI-D items into *Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition* depressive symptoms. Study 2 used 3 large samples to gather evidence for reliability and validity: Correlational analyses demonstrated alpha reliability and convergent and discriminant validity; factor analysis provided evidence for discriminant validity with anxiety; and regression analyses demonstrated comparative validity with existing standard CPI scales. Study 3 used clinician ratings of depression and anxiety as criteria for external validity.

The California Psychological Inventory (CPI; Gough, 1957, 1987; Gough & Bradley, 1996), a multivariate self-report inventory assessing life-enhancing attributes of personality, is one of the most widely used measures in psychological research (Gough, 2000; Groth-Marnat, 2003). Most notably, CPI data have been collected in several studies on life span development, some spanning nearly 50 years of adult development (Block, 1971; Cartwright & Wink, 1994; Helson, Jones, & Kwan, 2002; Helson, Kwan, John, & Jones, 2002; Helson, Stewart, & Ostrove, 1995; Stewart & Vandewater, 1999; Twisk, Snel, Kempor, & Van Mechelen, 1998). Such longitudinal studies have greatly illuminated the stability and change in personality across adulthood, but similar questions about the development of depressive symptomatology have yet to be addressed. Commonly used depression scales, such as the Beck Depression Inventory (BDI; Beck, 1978; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) and the Center for Epidemiological Studies Depression Scale (CES–D; Radloff, 1977), were developed and validated only within the past 30 years.

Thus, many longitudinal studies of adult development, especially longer term ones begun before the 1970s, did not include these scales. The majority of research on depressive symptomatology is cross-sectional or, if longitudinal, spans only a few years; as a result, little is currently known about the development of depressive symptoms across adulthood (Kraemer, Yesavage, Taylor, & Kupfer, 2000). This lack of data hampers the study of depressive symptoms from a developmental perspective, despite the fact that theorists are calling for such research (e.g., Hammen, 2000).

The open system axiom of the CPI (Gough & Bradley, 1996) supports the construction of new CPI scales when an important criterion is not already well predicted by an existing CPI scale. Consistent with this, new scales have been developed to assess specific aspects of psychopathology and maladjustment, such as narcissism (Wink & Gough, 1990) and hostility (Adams & John, 1997). Our new scale, the CPI Depressive Symptom Scale or CPI-D, would do the same for depressive symptomatology. Despite a small literature about how depression impacts CPI scores, a depressive symptom scale is needed for research using the CPI. A few existing studies indicate how CPI profiles might reflect depression, but none of these studies point clearly toward how such profile changes can be effectively used for research on the development of depressive symptomatology (Holliman & Guthrie, 1989; Holliman & Montross, 1984). For example, Holliman and Montross (1984) found that the majority of the CPI scale scores were negatively correlated with depressive symptom scores; however, the CPI scales that best predicted clinical depression were different for men and women and varied depending on the assessment of depressive symptoms that was used. Thus, a depressive symptomatology scale for the CPI could assess these symptoms as a consistent construct, going beyond suggesting which individuals might suffer from depression. Because the CPI has been used in much longitudinal and archival research, such a scale would allow researchers to tap this rich vein of accumulated data and to address questions about the development of depressive symptomatology and its relationship to personality immediately, rather than waiting

for several more decades to collect new data. In addition, because CPI data are most often gathered from the general population, these data would allow researchers to understand the emergence, development, and role of depressive symptoms in nonpatient populations.

In Study 1, we begin with an analysis of content validity of the CPI-D by comparing the items of the scale with the construct definition of a major depressive episode offered by the *Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition* (*DSM–IV*; American Psychiatric Association, 1994). In Study 2, we use three large samples of undergraduate students to document various aspects of reliability and validity of the CPI-D, including alpha reliability and convergent and discriminant validity of the CPI-D, using standard self-report measurements of depression and anxiety. In Study 3, we use clinician ratings of both depression and anxiety in adult women to examine the external validity of the CPI-D. Taken together, the three studies were designed to establish the construct validity of the CPI-D to make it available for future work.

Development of a Depressive Symptom Scale for the CPI

Item Selection and DSM–IV

As the CPI-D is necessarily a scale embedded within the CPI, the items of the CPI were our original item pool. Before selecting candidate items from the CPI for the Depressive Symptom scale, we studied various definitions of depression (e.g., Abramson, Metalsky, & Alloy, 1989; Beck, 1967; Blatt, 1974; Gold, 1990), *DSM–IV* diagnostic criteria for depression and dysthymia, and a number of existing depression scales (e.g., BDI, CES–D, Raskin's Depression Scale [RDS], Minnesota Multiphasic Personality Inventory Depression Scale [MMPI-D], Hamilton Rating Scale [HRSD]). To make our scale applicable to both archival and future research using the CPI, we used as our item pool the 480-item version of the CPI, which has been in use since 1957; slightly shorter versions of the CPI have appeared subsequently (e.g., Gough, 1987; Gough & Bradley, 1996).[1] Using this CPI version, 41 items were initially chosen from the CPI as reflective of depressive symptomatology as it is assessed by other widely used depression scales or as it is described in the *DSM–IV*.

Refining Item Selection With the Tripartite Model

Discriminant validity is traditionally addressed or evaluated with the use of correlational analyses after item selection has been completed. Clark and Watson's (1991; Watson et al., 1995; see also Tellegen, 1985) tripartite model suggests a way in which researchers may address anxiety as a discriminant validity concern during item selection (but see Marshall, Sherbourne, Meredith, Camp, & Hays, 2003, for an argument for further research on the utility of this model). The tripartite model suggests that three symptom domains—negative affect, positive affect, and anxious arousal—underlie anxiety and depression symptoms. According to this model, because negative affect or general distress (e.g., crying, irritability) is characteristic of both depression and anxiety, researchers should not expect such symptoms to discriminate between depressed and anxious individuals. Rather, symptoms reflecting anhedonia or low positive affect (i.e., lack of interest or

lack of positive emotional experiences) are relatively specific to depression, whereas symptoms reflecting anxious somatic arousal are relatively specific to anxiety. From this, Watson et al. (1995) recommended that depression scales avoid or limit anxiety symptoms that are generally more characteristic of anxious arousal (e.g., fear of losing control, pounding heart) as well as limit nonspecific symptoms of general distress that are most indicative of anxious distress (e.g., feeling tense, feeling nervous, upset stomach). It is important to note, however, that Watson et al. did not suggest that depression scales be composed solely of anhedonia symptoms, as this would result in overly narrow content validity.

To evaluate our candidate items from the perspective of the tripartite model, we independently categorized the 41 CPI candidate items into three symptom groups: general distress, anhedonia, and anxious arousal. The general distress items were further divided, according to Watson et al.'s (1995) recommendations, into three subcategories—general distress-mixed type; general distress-depressed type; and general distress-anxious type—resulting in a five-category classification. The interrater agreement for these 41 items was 89% ($\kappa = .86$). Of the 41 candidate items, 4 items were categorized as general distress-anxious type (e.g., "have a lump in the throat"). Furthermore, 4 of the 41 items were categorized as being more reflective of anxious arousal (e.g., "about to go to pieces") than of general distress or anhedonia. Following Watson et al.'s recommendations, we dropped these 8 items from our item list. The resulting scale, the CPI-D, consists of 33 items and is shown in Table 1. Like all CPI items, the CPI-D items are administered in a true–false format, thus indicating the presence or absence of each symptom; 8 of the 33 items are reverse-keyed.

The Present Studies: Comparative Design Across Four Samples

Here we report three studies that examine the construct validity of the CPI-D. In two studies, we used a comparative design allowing us to examine the psychometric properties of the CPI-D along with those of the BDI and CES–D in the same samples. The BDI and CES–D were used as comparison scales because they are among the most widely used depression scales of the past 30 years (Tennen, Hall, & Affleck, 1995).

Four nonclinical samples were used. In Study 2, three samples were taken from the university population, providing unique access to both large numbers of participants and to participants who can complete multiple measures. Two very large samples of college students completed the CPI-D, BDI, and CES–D, so that we could obtain stable estimates of reliability and convergent and discriminant validity. An additional sample completed the full CPI as well as the depressive symptom scales. To estimate comparability and enhance generalizability (Watson et al., 1995), our fourth sample was an older sample of women. We agree with Coyne (1994) that studies of depressive symptoms in nonclinical

---

[1] All 33 of the CPI-D items can be scored from the 1957 and 1987 editions of the CPI, and 31 items can be scored from the 1996 abbreviated version of the CPI (see Table 1). Similarly, because the CPI and Minnesota Multiphasic Personality Inventory share items, the CPI-D and MMPI-D have items in common but are not interchangeable scales. Only 6 of the 60 MMPI-D items are found verbatim on the CPI-D.

Table 1
Content Validity: The 33 CPI-D Items Classified Into the DSM–IV Depression Symptom Categories in Study 1

Symptom category	No.	CPI-D item
Sad or empty mood	245	Feels happy most of the time (R)
	416	Not quite as happy as others
	419	Life often has no meaning
	353	Not understood by others
	133	Feels as good as ever (R)
	398	Handed a raw deal in life
Diminished interest or pleasure in things	021	Life is full of interesting things (R)
	156	Hardly ever excited or thrilled
	280	Enjoys different kinds of play (R)
	124	Not likely to speak to others first
	070	Crosses the street to avoid meeting others
	161	Sometimes cross without good reason
Feelings of worthlessness or guilt	339	Thinks is no good at all
	279	Often gets disgusted with self
	147	Feels useless at times
	390	Has not lived the right kind of life
	311	Cannot do anything well
	257	Feels has done something wrong or wicked
	050	As capable and smart as most others (R)
	369	Has more regrets than others do
Suicidality or hopelessness about life	365	Future seems hopeless
	259	Life is worthwhile (R)
	015	Feels as if something bad is about to happen
	299	Does not care what happens
Diminished concentration	054	Hard to keep mind on task or job
	456	Has trouble concentrating
	099	Cannot keep mind on one thing
	013	Slow in making up mind
Fatigue	238	Sometimes can't get going
	426	Gets tired easily
Changes in appetite	400	Has a good appetite (R)
Changes in sleeping	459	Sleep is fitful and disturbed
	135	Wakes up fresh and rested (R)

Note. California Psychological Inventory (CPI) items are abbreviated and paraphrased and are included to illustrate the item content of the CPI-D. The number to the left of each item is the item number on the Form 462 version of the CPI (Gough, 1987). Item numbers on the Form 434 version of the CPI (Gough & Bradley, 1996) are the same except for Items 135 and 419, which are not included; Item 456, which is numbered Item 362; and Item 459, which is numbered Item 402. CPI-D = California Psychological Inventory Depressive Symptom scale; *DSM–IV = Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition*; R = reverse keyed.

populations should "not be interpreted as analog studies of depression" (p. 40), and here we intended neither to conduct studies of clinical depression nor to create a measure for clinical diagnosis. Rather, we intended to create a scale of depressive symptomatology that demonstrated psychometric properties similar to other widely used scales, such as the BDI and CES–D. The development of such a scale for a nonclinical inventory such as the CPI should facilitate research on depressive symptomatology, understood as a continuous variable and present in general populations.

Our CPI-D scale was developed and validated in nonclinical samples for three reasons. Most important, we emphasize the use of the CPI-D with general populations because the CPI-D is a scale from the CPI, a nonclinical inventory. Second, many researchers argue that depressive symptomatology warrants study in individuals who report so-called subclinical levels of depression (i.e., not meeting *DSM–IV* criteria for clinical depression) because these individuals experience difficulties in psychosocial functioning, morbidity, and occupational functioning equal to or greater than

those reported by individuals who are clinically depressed (Broadhead, Blazer, George, & Tse, 1990; Costello, 1992; Gotlib, Lewinsohn, & Seeley, 1995; Johnson, Weissman, & Klerman, 1992; Judd, Rapaport, Paulus, & Brown, 1994). Third, from a developmental perspective, elevated depressive symptomatology has been seen as a risk factor for later clinical depression (Compas, Ey, & Grant, 1993; Gotlib et al., 1995; Wells, Burnam, Rogers, Hays, & Camp, 1992). Thus, we concluded, as have similar studies (e.g., Dozois, Dobson, & Ahnberg, 1998), that large nonclinical samples were appropriate for our goals.

Study 1: Content Validity

Efforts to establish the content validity of depression scales are surprisingly few. For example, reviews of the BDI and CES–D tend to comment informally on the number of *Diagnostic and Statistical Manual of Mental Disorders* symptoms that are represented in each scale. Reviewing the CES–D, Rabkin and Klein

(1987) stated that "a few items" assessed each of six depressive symptoms (p. 76), yet the empirical basis and exact meaning of such statements are unclear. Similarly, Moran and Lambert (1983) reviewed the item content of the BDI and indicated which items, in their view, reflected various *Diagnostic and Statistical Manual of Mental Disorders—Third Edition* (American Psychiatric Association, 1980) depressive symptoms; this review, however, did not account for all of the items on the scale. In addition, subjective impressions of content differ from one set of reviewers to another (e.g., see Dozois et al., 1998; McDowell & Newell, 1996, for two different versions of BDI symptom coverage).

Under the criteria for a major depressive episode, the *DSM–IV* lists nine symptoms: (a) sad or empty mood, (b) diminished interest or pleasure in activities, (c) changes in appetite or weight, (d) changes in sleep, (e) psychomotor agitation or retardation, (f) fatigue, (g) beliefs of worthlessness or guilt, (h) diminished ability to concentrate, and (i) recurrent thoughts of death. Not all nine of the symptoms are equally important or central to depression; at least one must be either (a) sad or empty mood or (b) diminished interest or pleasure in activities. Thus, according to the *DSM–IV*, negative affect and anhedonia (low positive affect) are most central to depressive symptomatology; individuals who only report difficulties with sleeping, eating, energy, self-worth, concentration, or suicidality would not meet the *DSM–IV* criteria for depression because the core symptoms of depressed mood and anhedonia are not present. This is consistent with the fact that some of these less central symptoms of depression (e.g., fatigue, poor concentration, and changes in sleeping) are also symptoms of other *DSM–IV* disorders, such as generalized anxiety disorder.

Thus, to prioritize content validity in the development of the CPI-D and to respond to some of the gaps in the literature in reviews of other depressive symptom scales, we conducted an empirical study, relying on multiple and independent judges that could consensually assign the CPI-D, BDI, and CES–D items to *DSM–IV* depressive symptom categories for a major depressive episode. We used *DSM–IV* depressive symptomatology as our rubric, not because it guaranteed construct validity but because it would provide a common language for a discussion of content validity (Tennen et al., 1995). We expected that the CPI-D would demonstrate relatively broad content validity and would emphasize central depressive symptoms, such as depressed mood and anhedonia. Similarly, consistent with the fact that physiological symptoms are nonspecific depressive symptoms and that fewer such items are included on the CPI, we expected that physiological items would be less represented.

Method

Judges. Judges were six advanced doctoral students in clinical psychology. These judges represented three ethnic groups (European American, African American, and Asian American) and ranged in age from 24 to 30 years. All six judges had received clinical training and supervision, as well as research training. They were naive to the purpose of the study.

Scales measuring depressive symptomatology. The BDI (Beck, 1978; Beck et al., 1961) is a 21-item scale developed to assess depressive severity in individuals already diagnosed with clinical depression, yet it is regularly used with both clinical and nonclinical samples.[2] In the most commonly used 1978 version, each BDI item consists of four statements indicating increasing severity of a symptom. Each choice is given a weight of 0–3 points with no reverse-keyed items to break response set. Respondents are instructed to describe the way they have been feeling during the past week. The BDI has demonstrated high internal consistency, with alpha reliabilities between .73 and .95, and an average alpha of .81 in nonpatient samples (Beck, Steer, & Garbin, 1988). Correlations between the BDI and the CES–D were conspicuously absent in reviews of these two scales, but, in two studies, the correlations were in the .80s (Santor, Zuroff, Ramsay, Cervantes, & Palacios, 1995; Weissman, Prusoff, & Newberry, 1975). Correlations between the BDI and anxiety scales tend to be in the .60s and .70s, showing moderate discriminant validity with anxiety (Baker & Jessup, 1980; Meites, Lovallo, & Pishkin, 1980; Tanaka-Matsumi & Kameoka, 1986).

The CES–D (Radloff, 1977) was developed for research on the levels of depressive symptomatology in the general population. The 20 items were adapted from existing depression scales including the BDI, the MMPI-D (Dahlstrom & Welsh, 1960), and the RDS (Raskin, Schulterbrandt, Rearig, & McKeon, 1969). Sixteen items were intended to represent cognitive, affective, and behavioral components of depressive symptomatology (e.g., "I felt depressed"); 4 reverse-scored items were included to break response set and to assess the absence of positive affect (e.g., "I was happy;" Devins & Orme, 1985). The CES–D assesses the frequency and duration of depressive symptoms during the past week. Respondents indicate how often during the week they have experienced each item on a scale of 0–3. The CES–D has high alphas, generally in the .80s and .90s (Devins & Orme, 1985; Nezu, Ronan, Meadows, & McClure, 2000; Radloff, 1977). Like those for the BDI, correlations between the CES–D and self-report anxiety scales range from .45 to .80 (Orme, Reis, & Herz, 1986; Weissman, Sholomskas, Pottenger, Prusoff, & Locke, 1977).

Procedure. All items from the CPI-D, BDI, and CES–D were randomly arranged in one list; individual scales were not identified, and scale directions and answer choices were omitted. The judges independently categorized the items into the nine depressive symptoms provided by the *DSM–IV*. In addition, judges were instructed to classify as anxiety items those that appeared to indicate anxiety (i.e., tension, worry) and anxious arousal (i.e., pounding heart, sweating) more so than depression. Also, judges were instructed to classify items with content reflecting hopelessness about the future under the *DSM–IV* symptom of thoughts of death or suicide. A similar approach of including cognitive indicators of extreme apathy and suicidality, as well as behavioral ones, was used in the HRS (Hamilton, 1960); in addition, hopelessness is included in the broader *DSM–IV* discussion of the features of a major depressive episode as an indication of depressed mood and as a motivation for suicide (American Psychiatric Association, 1994, pp. 320–322).

Results and Discussion

Interjudge agreement for this task was considerable and did not differ across the three instruments; specifically, agreement among the six judges for the categorization of the 74 items was 90% ($\kappa =$ .85). Obviously, not all classifications were made with perfect agreement, as some items addressed more than one symptom. For example, a BDI item reflecting inability to complete work was classified by four judges as indicative of difficulties with concen-

---

[2] More significant changes were made to the BDI in 1996 so that the BDI-II (Beck, Steer, & Brown, 1996) might be more compatible with the *DSM–IV* criteria for depression. Still, according to the PsycINFO database, the BDI-II was used in under 30 reported research studies between 1996 and 2001, whereas the BDI was used in more than 500. For this reason, and because psychometric characteristics of the BDI-II were only beginning to emerge at the time when the studies reported here were conducted (e.g., Dozois et al., 1998), the most commonly used version of the BDI was used in this research. By the end of 2003, the BDI had been used in almost 1,500 reported studies whereas the BDI-II had been used in just over 100.

tration and by two judges as indicative of fatigue. On items for which agreement was not perfect, the symptom identified by the majority of judges was used in our subsequent analyses; no items produced an even split among the judges. Two items were judged to be indicative of anxiety or anxious arousal: an item suggesting fearfulness on the CES–D and an item suggesting worry over physical problems on the BDI.

DSM–IV symptom categorizations for each of the 33 CPI-D items are shown in Table 1. These findings show that the CPI-D included at least one item for eight of the nine *DSM–IV* symptoms of depression, with only psychomotor agitation or retardation not represented. The percentages of CPI-D items that address each *DSM–IV* depressive symptom are shown in Table 2, as are those for the BDI and CES–D for comparison. The 33 CPI-D items were distributed fairly evenly across the symptoms of sad or empty mood (18%), diminished interest or pleasure in activities (18%), diminished ability to concentrate (12%), and apathy or hopelessness about the future (12%); slightly more items reflected worthlessness or guilt (24%). One or two items reflected each of the less central physiological symptoms of depression: fatigue (6%), difficulty sleeping (6%), and change in appetite (3%).

These symptom categorizations generally confirmed our expectations. The CPI-D achieved broad content validity, comparable with that of the BDI and the CES–D. The CPI-D equally emphasized the essential *DSM–IV* affective symptoms of sad and empty mood and loss of interest or pleasure in activities; symptoms reflecting beliefs of worthlessness or guilt were also emphasized. As is desirable, a smaller percentage of items addressed the relatively nonspecific somatic symptoms of depression. The CPI-D, BDI, and CES–D each achieved relatively broad coverage of the *DSM–IV* symptoms, but each scale addressed these symptoms to somewhat different extents.

Study 2: Reliability and Convergent and Discriminant Validity

The goal of Study 2 was to address the most central aspects in a program of construct validation (Messick, 1995). We recruited three large samples to address reliability and validity of the CPI-D

Table 2

Comparative Content Validity of the CPI-D: Percentage of Items From the CPI-D, BDI, and CES–D Representing DSM–IV Depressive Symptoms in Study 1

Symptom	CPI-D	BDI	CES–D
Sad or empty mood	18	14	40
Diminished interest or pleasure	18	14	10
Worthlessness or guilt	24	29	15
Suicidality or hopelessness	12	10	5
Diminished concentration	12	10	5
Fatigue	6	5	10
Change in appetite	3	10	5
Change in sleep	6	5	5
Psychomotor agitation or retardation	0	0	0

Note. CPI-D = California Psychological Inventory Depressive Symptom scale; BDI = Beck Depression Inventory; CES–D = Center for Epidemiological Studies Depression Scale; *DSM–IV = Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition.*

in comparison with established depressive symptom self-report scales (e.g., CES–D, BDI). In our largest sample, Sample A, we documented alpha reliability and convergent validity. In Sample B, we tested for the replication of findings from Sample A; also, using a measure of anxiety, we gathered evidence for discriminant validity of the CPI-D, both at the scale and item level. In Sample C, we administered the full CPI to test for the replication of reliability and validity evidence and to demonstrate the validity of the CPI-D compared with the existing standard scales of the CPI.

Method

Participants. Three samples, Samples A, B, and C, took part in a research study in exchange for course credits in introductory psychology courses. These samples are described in Table 3.

Measures of depressive symptomatology. The CPI-D, BDI, and CES–D are described in Study 1.

Measure of anxious symptomatology. The A-State Anxiety scale of the State–Trait Anxiety Inventory (STAI Anxiety; Spielberger, Gorsuch, & Lushene, 1970) consists of 20 statements that ask individuals to indicate how they feel at a given moment. Example items include "I feel anxious" and "I feel calm" (reverse-keyed). Respondents endorse one of four choices for each statement (i.e., 1 = *not at all*, 2 = *somewhat*, 3 = *moderately so*, or 4 = *very much so*). Alpha reliability is generally high, from about .83 to .92.

CPI. To determine the internal consistency of the CPI-D when scored from the full CPI and to examine the comparative validity of the CPI-D, we administered the full CPI (Gough & Bradley, 1996) to Sample C. Participants received detailed feedback by way of quantitative and narrative computer-scored protocols from Consulting Psychologists Press. We assumed this personalized feedback would motivate participants to complete the CPI faithfully, and the absence of invalid CPI protocols suggests this was the case.

Results and Discussion

Score distribution and demographic variables. The means and standard deviations for the CPI-D, BDI, CES–D, and STAI Anxiety in our three samples are shown in Table 4. Depressive symptomatology measured by the CPI-D, BDI, and CES–D was not related to ethnicity; sex differences were negligible, as the strongest correlation with sex (female keyed high) was .08 in Sample A.

Alpha reliability. For Samples A, B, and C, the alpha coefficients of the three depression scales are shown in parentheses on the diagonals of Table 4; they were all substantial and very similar to each other, ranging from .87 for the BDI in Sample A to .92 for the CES–D in Sample C. The alpha reliability estimates of the CPI-D were .88, .88, and .90, across Samples A, B, and C, respectively. Thus, despite the dichotomous response format of the CPI-D, the level of reliability achieved with the 33 CPI-D items was comparable to that of the 21-item BDI and the 20-item CES–D.

Convergent validity. Convergent validity correlations are shown in Table 4. The CPI-D showed substantial convergence with the BDI and CES–D. In Sample A, the CPI-D correlated .78 with the BDI and .69 with the CES–D, whereas the BDI and CES–D correlated .69 with each other. Similarly, in Sample C, the CPI-D correlated .81 with the BDI and .79 with the CES–D, whereas the BDI and the CES–D correlated .76 with each other. Across the three samples, the CPI-D had a mean convergent correlation of .76 (Table 4). Disattenuated correlations approached .90.

Table 3
Sample Characteristics and Measures Used in the Three Samples in Study 2

Characteristic or measure administered	Sample		
	A	B	C
Sample size	1,044	568	244
Mean age (years)	19.4	20.2	20.8
% women	59	66	73
% nonpsychology majors	60	64	59
% African American	4	4	3
% Asian American	44	34	49
% European American	35	29	30
% Latino	6	7	14
% other ethnicity	11	26	4
CPI-D	Yes	Yes	Yes
CES–D	Yes	Yes	Yes
BDI	Yes	No	Yes
STAI Anxiety	No	Yes	No
CPI (full length)	No	No	Yes

Note. These samples closely mirrored the ethnic composition of their university. Sample B did not complete the BDI because of time limitations. CPI-D = California Psychological Inventory Depressive Symptom scale; CES–D = Center for Epidemiological Studies Depression Scale; BDI = Beck Depression Inventory; STAI = State–Trait Anxiety Inventory; CPI = California Psychological Inventory.

Comparative validity. According to Gough and Bradley (1996), the CPI Well-Being scale (CPI-Wb) is the CPI scale that correlates most strongly with the BDI ($r = -.49$) and the MMPI-D ($r = -.54$). For this reason, and because the CPI-D shares its greatest number of items (i.e., seven items) with the CPI-Wb, we expected the CPI-D to do the same. At the same time, we also expected that the CPI-Wb would be too general a measure of adjustment to function specifically as a depressive symptoms scale, as the CPI-Wb contains 31 items whose content is not specific to depressive symptomatology (e.g., indicative of acid stomach or being treated like a child). Also, the CPI-Wb tends to correlate strongly with various aspects of adjustment (e.g., .81 with low anxiety, .77 with emotional stability; Gough & Bradley, 1996), making it "a rough estimate of a person's level of adjustment" (Groth-Marnat, 2003, p. 371). Overall, we expected that the CPI-D would have stronger and unique associations with measures of depressive symptomatology, such as the BDI and CES–D, than would the CPI-Wb. We tested these predictions in Sample C.

As expected, of all 20 CPI standard scales, the CPI-Wb had the strongest correlation with the CPI-D ($r = -.74$), followed by the Intellectual Efficiency scale ($r = -.61$).[3] Given the item overlap between the CPI-Wb and the CPI-D, we examined whether the CPI-Wb predicted depressive symptomatology (as measured by the BDI and CES–D) as well as the CPI-D did. The BDI correlated only $-.56$ ($p < .01$) with the CPI-Wb but .81 ($p < .01$) with the CPI-D. The difference between these correlations was significant, $t(241) = 9.22$, $p < .01$ (Cohen & Cohen, 1983). Similarly, the CES–D correlated $-.52$ with the CPI-Wb, compared with .79 for the CPI-D; again, this difference was significant, $t(241) = 9.59$, $p < .01$.

Finally, we tested whether the CPI-Wb captured unique variance related to depressive symptoms, above and beyond that accounted for by the CPI-D. We conducted multiple regression analyses, one predicting the BDI and the other predicting the CES–D. Table 5 summarizes the results. For both criterion variables, we entered the CPI-Wb in the first step and the CPI-D in the second step. When the CPI-D was entered in the second step, the CPI-Wb no longer predicted either BDI or CES–D scores. That is, once the effect of the CPI-D was controlled, the CPI-Wb no longer captured any depressive symptom variance. These results were the same even when the CPI Intellectual Efficiency scale, the next largest CPI-D correlate, was added as an additional predictor. These findings are consistent with the view that the CPI-Wb is a measure of global positive adjustment, whereas the CPI-D serves a unique function in the CPI as a specific measure of depressive symptomatology.

Discriminant validity at the scale level. Anxiety has been a persistent discriminant validity concern for measures of depressive symptoms. In Sample B, we computed convergent and discriminant correlations of the CPI-D and CES–D with STAI Anxiety to examine the discriminant validity of the CPI-D at the scale level and to compare it with that of the CES–D. Previous studies have found positive and fairly substantial correlations between depressive and anxious symptomatology, with correlations between depression and anxiety scales typically ranging from .45 to .75 (Clark & Watson, 1991; Meites et al., 1980; Tanaka-Matsumi & Kameoka, 1986; Weissman et al., 1977); Spielberger et al. (1970) reported that STAI Anxiety correlated .44 to .57 with a self-report measure of depression. To demonstrate adequate discriminant validity of the CPI-D, we expected the discriminant correlation between the CPI-D and the STAI Anxiety to fall within this range of values. More important, the discriminant correlation of the CPI-D with anxiety should be significantly lower than the convergent correlations of the CPI-D with the other depression scales.

Discriminant correlations are reported at the bottom of Table 4. The CPI-D correlated significantly more highly with the CES–D ($r = .68$) than with the STAI Anxiety ($r = .55$), as shown by a test of the significance of the difference between dependent correlations, $t(565) = 5.00$, $p < .01$ (Cohen & Cohen, 1983). For the CES–D, however, the difference between the convergent correlation with the CPI-D ($r = .68$) and the discriminant correlation with the STAI Anxiety ($r = .64$) was not significant, $t(565) = 1.47$, *ns*. Thus, the CPI-D performed somewhat better than the CES–D with respect to discriminant validity, and the .55 correlation between the CPI-D and the STAI Anxiety was well within the typical range of correlations between depression and anxiety as well as within the range of correlations between STAI Anxiety and depression.

Discriminant validity using item–scale correlations and confirmatory factor analysis (CFA). As described earlier, we used the tripartite model (Clark & Watson, 1991; Watson et al., 1995) to address content and discriminant validity during item selection and dropped candidate items that were conceptually related more to anxious arousal and anxious distress than to depression. To evaluate empirically whether the final CPI-D contained items that were associated more strongly with anxiety than with depression, we conducted a series of item analyses to compare the convergent and discriminant validity of the 33 individual CPI-D items. For each CPI-D item, we compared its corrected item–total correlation with its correlation with STAI Anxiety. Point-biserial correlations between individual CPI-D items and STAI Anxiety were generally

_____

[3] Full correlation tables are available from Meg Jay on request.

Table 4
Alpha Reliability (on the Diagonal), Convergent and Discriminant Validity Correlations, and Descriptive Statistics in the Three Samples in Study 2

Sample and measure	CPI-D	BDI	CES–D	*M*	*SD*	Skewness	Kurtosis
Sample A							
CPI-D	(.88)			8.5	6.2	0.91	0.44
BDI	.78	(.87)		6.1	6.2	1.60	3.00
CES–D	.69	.69	(.88)	14.6	9.4	0.92	0.63
Sample B							
CPI-D	(.88)			8.8	6.3	0.76	−0.10
CES–D	.68		(.88)	12.8	8.8	0.90	0.75
Sample C							
CPI-D	(.90)			8.4	5.7	0.65	−0.15
BDI	.81	(.90)		8.5	7.9	1.30	1.40
CES–D	.79	.76	(.92)	17.1	10.5	0.79	0.09
Mean convergent *r* [a]	.76	.76	.73				
STAI discriminant *r* [b]	.55		.64	16.6	10.3	0.52	−0.18

Note. All correlations were significant at $p < .01$. Alpha reliabilities are in parentheses on the diagonal. CPI-D = California Psychological Inventory Depressive Symptom scale; BDI = Beck Depression Inventory; CES–D = Center for Epidemiological Studies Depression Scale; STAI = State–Trait Anxiety Inventory.
[a] Mean of the convergent correlations across Samples A, B, and C using Fisher's *r*-to-*z* transformations.
[b] STAI correlations are discriminant correlations of CPI-D or CES–D with STAI Anxiety in Sample B.

low, ranging from .08 for the item (paraphrased here), 'don't care what happens,' to .35 for the item (paraphrased here), 'not as happy as others are.' Even more important, for every item, the corrected item–total correlation with the CPI-D was always higher than its point-biserial correlation with STAI Anxiety.

To examine the discriminant relations with anxiety more formally, we conducted a series of CFAs of the 33 CPI-D items, the 20 CES–D items, and the 20 STAI Anxiety items in Sample B ($N = 568$). We tested the fit of one-factor, two-factor, and three-factor models, each with ordered categorical and dichotomous indicators. To do so, we used Mplus, Version 2.1 (Muthén &

Table 5
Comparative Validity: Predicting BDI and CES–D From CPI-D and CPI-Wb in Study 2

Variable	Increase in R^2	β
	BDI	
Step 1	.31**	
CPI-Wb		−.56**
Step 2	.34**	
CPI-Wb		.08
CPI-D		.87**
	CES–D	
Step 1	.27**	
CPI-Wb		−.52**
Step 2	.35**	
CPI-Wb		.10
CPI-D		.85**

Note. The regression analyses were conducted in Sample C. BDI = Beck Depression Inventory; CPI-Wb = California Psychological Inventory Well-Being scale; CPI-D = California Psychological Inventory Depressive Symptom scale; CES–D = Center for Epidemiological Studies Depression Scale.
** $p < .01$.

Muthén, 2001), with WLSMV estimation (i.e., weighted least square parameter estimate using a diagonal weight matrix with robust standard errors and mean-and-variance-adjusted chi-square test statistic; Muthén & Muthén, 2001), generating Satorra–Bentler chi-square statistics (S-B χ^2). Most parsimonious is the one-factor model, which assumes that all the items measure a general distress dimension and no reliable discrimination between depression and anxiety items can be made. This one-factor model fit the least well, and the fit indices were as follows: S-B $\chi^2(268) = 1,554.00$, Tucker-Lewis Index (TLI) = .89, root-mean-square error of approximation (RMSEA) = .09, standarized root-mean-square residual (SRMR) = .10.[4] We then examined the fit of a two-factor model, with one factor representing depressive symptoms (i.e., all items from the CPI-D and the CES–D) and the other factor representing anxiety symptoms (i.e., all STAI items); the two factors were allowed to correlate.[5] As expected, in comparison to the one-factor model, the two-factor model achieved better fit, S-B $\chi^2(271) = 1,109.00$, TLI = .93, RMSEA = .07, SRMR = .09, with values closely approaching suggested cutoff scores for good models (Hu & Bentler, 1999; see also Hill, Neumann, & Rogers, 2004). The estimated correlation between the latent Depressive Symptom factor and the latent Anxiety factor in the two-factor model was .65.

For completeness, we also examined a three-factor model, with one factor representing depressive symptoms as measured by the CPI-D, another representing depressive symptoms as measured by

---

[4] In Mplus, for WLSMV estimation used with categorical indicators, degrees of freedom are calculated with the following formula: $df = (tr(U\Gamma))^2/(tr(U\Gamma)^2)$. (See Muthén & Muthén, 2001, Appendix 4, Formula 110.)

[5] In this model, five items were allowed to load on both the Depressive Symptom and the Anxiety factor: the CES–D item "I felt fearful" had been classified by our judges in Study 1 as more reflective of anxiety, whereas four STAI items ("I feel joyful," "I feel pleasant," "I feel content," and "I feel confident") seemed to reflect depression as much as anxiety.

the CES–D, and the third representing anxiety symptoms on the STAI. In our large sample, the added complexity of differentiating between CPI-D and CES–D variants of depressive symptoms led to little or no improvement in fit indices, S-B $\chi^2(271) = 1,030.00$, TLI = .93, RMSEA = .07, SRMR = .09. Also, the correlations among the three latent factors closely replicated the pattern of the simple convergent and discriminant correlations among the scales (see Table 4): The latent CPI-D Depressive Symptom factor correlated much more highly with the CES–D factor ($r = .78$) than with the STAI Anxiety factor ($r = .62$); the correlation of the CES–D and STAI factors fell in between ($r = .71$). In summary, the findings in scale-level analyses, item-level analyses, and CFAs showed the same consistent pattern, providing considerable discriminant validation evidence for the CPI-D scale and its items.

Study 3: External Validity

Study 3 used data from the Mills Longitudinal Study (Helson & Kwan, 2000; Helson et al., 2002; Helson, Pals, & Solomon, 1997), an ongoing study of women now in their early 60s. CPI data and clinician ratings were obtained when the women were 61 years of age. Thus, we were able to examine the psychometric properties of the CPI-D when completed by an older age group than the participants studied so far. Even more important, we addressed the issue of external validity of the CPI-D, using clinician ratings of depression and anxiety symptoms as convergent and discriminant external criteria. Clinician ratings of depression and anxiety symptoms were obtained under conditions suggested by Clark and Watson (1991): (a) Raters were similarly and adequately trained, (b) rating criteria were clearly specified, and (c) ratings were based on the same information.

Beck et al. (1988) reported that in nonpsychiatric samples, the mean correlation between clinician ratings of depressive symptoms and the BDI was .60; for the CES–D, correlations ranged from .46 to .53 (Radloff, 1977). We expected a similar correlation between the CPI-D and clinician ratings of depressive symptoms. Many external validity studies have not addressed the issue of discriminant validity on the criterion side. For the BDI, however, Beck et al. (1988) reported a .14 correlation between BDI scores and clinician ratings of anxiety. Similarly, we expected the CPI-D to correlate more highly with clinician ratings of depressive symptoms than with clinician ratings of anxiety symptoms.

Method

Participants. The participants were 110 women who are participants in the Mills Longitudinal Study (Helson, Jones, & Kwan, 2002; Helson & Kwan, 2000; Helson et al., 1995, 1997) and who graduated from college in either 1958 or 1960. According to their college grade point averages and Scholastic Aptitude Test scores, these women were representative of the Mills College population at the time. At age 61, the women participated in a 1-day assessment at the University of California, Berkeley.

CPI-D. The women completed the 1987 version of the CPI (Gough, 1987), and the CPI-D (described in Study 1) was scored from these CPI protocols.

Clinician Q-sorts. As part of the age 61 assessment, a 2.5-hr structured interview was administered individually to each participant. In the interview, participants were asked about their current involvement in and feelings about work, community activities, their relationships and friendships, childrearing, caretaking of aging parents, health, retirement, spirituality, and death.

Immediately following each interview, the clinicians conducting the interviews used the California Adult Q-Set (CAQ; Block, 1961) to quantify their observations of each participant interviewed. Block (1961) developed the CAQ to provide a comprehensive, generally applicable, and standardized language for describing a range of individual differences in experience, thought, and behavior. The CAQ is a general purpose instrument that originated from clinical and psychodynamic theory; it thus avoids the limitations of other instruments that are specifically focused on one or a few predetermined variables. The CAQ is a set of 100 cards with descriptive statements (e.g., "Feels a lack of personal meaning in life") to describe an individual; raters or interviewers divide and sort these cards into a quasi-normal distribution using nine piles of cards, with the piles scored from 1 (*least characteristic*) to 9 (*most characteristic*).

Interviewers were three practicing clinicians with doctoral degrees and three advanced graduate students working toward their doctorate in clinical psychology. Interviewers had received extensive clinical training and supervision and had worked with clients for at least 3 years. In a workshop that used transcripts and videotapes of the structured interview, clinician interviewers were trained to complete the interview in a uniform fashion. In a separate workshop, interviewers were trained in the CAQ method and were required to complete several practice CAQ sorts in which interrater reliability was at least .80.

CAQ depressive symptom index. To measure depressive symptoms from the clinician's CAQ of the participant, we used Block's (1989) expert-derived depression prototype. The nine items judged by a panel of experts as most characteristic of depression were aggregated to form the CAQ Depression Index (see Table 6). The CAQ Depression Index has good content validity, covering four central *DSM–IV* symptoms of depression: Three items measure worthlessness–guilt, and two items each measure sad–empty mood, lack of pleasure–interest, and diminished concentration. Alpha was .78.

CAQ anxiety symptom index. We also used the CAQ to derive an anxiety index. Four independent experts agreed on three CAQ items as clear indicators of anxiety: "Is basically anxious," "Anxiety and tension are manifested in bodily symptoms," and "Is calm, relaxed in manner" (reverse-keyed). The resulting CAQ Anxiety Index had an alpha of .70. The CAQ Depression and Anxiety Indices correlated .50 ($p < .01$), similar to the average correlation between clinical ratings of depressive symptomatology and clinical ratings of anxiety symptomatology (Clark & Watson, 1991).

Results and Discussion

Alpha reliability. The CPI-D had an alpha reliability of .82, suggesting that the scale is a reliable measure in non-college-age adults.

External validity. The CPI-D correlated .59 ($p < .01$) with the clinician-rated CAQ Depression Index, providing evidence of considerable validity against an independent, non-self-report criterion. This convergent correlation contrasts with the discriminant correlation of .33 ($p < .01$) between the CPI-D and the CAQ Anxiety Index.[6] Even in this relatively smaller sample, the difference between the convergent and discriminant correlations was again significant, $t(107) = 3.61$, $p < .01$ (Cohen & Cohen, 1983). Correlations between the CPI-D and the individual items on the CAQ indices are shown in Table 6; all correlations were positive, and eight of nine were significant.

---

[6] This substantial difference between convergent and discriminant correlations cannot be explained in terms of the slightly lower reliability of the CAQ Anxiety Index. Even when the correlations were corrected for attenuation due to unreliability, the shared variance percentages were 41% to 15% of the total variance, still almost a 3-to-1 ratio.

Table 6

Convergent and Discriminant Validity of the CPI-D: Correlations With External Depression and Anxiety Criteria in Study 3

Interviewer Q-sort	r with CPI-D
CAQ Depression Index	.59*
CAQ Anxiety Index	.33*
CAQ depression items	
Cheerful (R)	.45*
Feels a lack of personal meaning	.38*
Concerned with personal adequacy	.36*
Self-defeating	.34*
Productive (R)	.30*
Readiness to feel guilt	.29*
Has rapid personal tempo (R)	.29*
Ruminates	.24*
Initiates humor (R)	.06
CAQ anxiety items	
Basically anxious	.28*
Bodily manifestations of anxiety	.27*
Calm, relaxed manner (R)	.12

Note. Q-sort items are abbreviated and paraphrased. CPI-D = California Psychological Inventory Depressive Symptom scale; CAQ = California Adult Q-Set; R = reverse keyed.
* $p < .05$.

The substantial convergent correlation between CPI-D scores and clinician ratings of depression was similar in size to those reported in reviews of the BDI and the CES–D (Beck, 1967; Beck et al., 1988; Orme et al., 1986; Radloff, 1977). In addition, the significantly lower correlation between CPI-D scores and clinician ratings of anxiety is similar to findings for the BDI (Beck, 1967). This pattern of convergent and discriminant correlations with clinician ratings is considered strong evidence of the discriminant validity of the BDI (Beck et al., 1988). Overall, then, Study 3 provided promising evidence for both the convergent and discriminant validity of the CPI-D when compared with the external criteria of clinician ratings of depression and anxiety symptoms.

General Discussion

Construct Validity of the CPI-D: Convergent and Discriminant Evidence

Studies 1–3 provided substantial evidence for the construct validity of the CPI-D, a newly developed CPI scale designed to measure depressive symptomatology. As we demonstrated empirically in Study 1, clinically trained judges found that the CPI-D items demonstrated broad content validity in terms of *DSM–IV* depressive symptomatology. Nonetheless, it is important to note that the CPI-D is not intended to serve as an instrument for the clinical diagnosis of depression. Rather, the CPI-D is a scale that aims to assess depressive symptomatology as a continuous variable in nonclinical populations.

The psychometric properties of the CPI-D were investigated in Studies 2 and 3. The 33 items on the CPI-D assess the presence or absence of depressive symptoms, with total scores ranging from 0 to 33. In four samples, alpha reliability was substantial, ranging

from .82 to .90, and these findings held for both women and men and for both younger and older adults. Convergent validity was determined through comparison with the BDI and CES–D and with independent ratings by clinically trained interviewers. Across studies, on average, the CPI-D correlated .80 with the BDI and .72 with the CES–D. Consequently, the CPI-D correlated at least as highly with the BDI and the CES–D, as those two widely used scales correlated with each other. The CPI-D correlated much more strongly with these two depression scales than did any of the 20 standard CPI scales, and it accounted for more than twice the depressive symptom variance than the closest CPI scale, the CPI-Wb. Thus, the CPI-D makes a unique contribution to the existing set of CPI scales.

The substantial convergent correlations among the CPI-D, BDI, and CES–D contrast with the much lower correlation between the CPI-D and self-reported anxiety ($r = .55$ in Study 2), demonstrating discriminant validity similar to that of commonly used depression scales. The same pattern of convergent and discriminant relations also emerged for the CPI-D in Study 3, in which clinician ratings were used as external criteria for depression and anxiety symptoms. These external validity correlations ($r = .59$ with clinician-rated depression symptoms, as compared with only $r = .33$ with clinician-rated anxiety symptoms) were comparable to those reported for the BDI and the CES–D (Beck, 1967; Beck et al., 1988; Radloff, 1977).

Also, CFA provided evidence of discriminant validity with anxiety. We compared one-, two-, and three-factor models of depression and anxiety items from the CPI-D, CES–D, and STAI. The two-factor solution, in which CPI-D and CES–D items comprised a depression factor and the STAI comprised an anxiety factor, provided the best fit for our data. Similarly, in correlational analyses, corrected item–total correlations for the individual CPI-D items were always higher than correlations of the CPI-D items with STAI Anxiety. Taken together, item-level analyses indicated that our item refinement procedure based on the work of Clark and Watson (Clark & Watson, 1991; Watson et al., 1995) had sufficiently eliminated anxiety items from our scale and that the discriminant validity of both our items and our scale were well within psychometric standards for depressive symptom scales.

Overall, Studies 1–3 were promising as evidence of the construct validity of the CPI-D. For future work, although the CPI-D was not intended to be a clinical measure, replication of the findings reported here with both nonclinical and clinical samples would be useful. In addition, because participants in Studies 1–3 were sampled from undergraduate populations and an older female population, it will be useful to document the reliability and validity of the CPI-D in other samples.

Future Applications of the CPI-D

One of the most important research applications for the CPI-D will be in longitudinal studies, as such research is increasingly recognized as invaluable for our understanding of developmental processes (Hammen, 2000; Kraemer et al., 2000). Depressive symptoms are the psychiatric symptoms most commonly found in nonclinical and community populations (Weissman & Boyd, 1981) and are increasingly recognized as a central mental health issue (Broadhead et al., 1990; Gotlib et al., 1995; Johnson et al., 1992; Judd et al., 1994). Unfortunately, little is known about the

development of depressive symptomatology in community samples, perhaps because a true understanding of developmental processes requires decades of research following the same individuals. Most longitudinal studies of such a length were begun before the impact of depressive symptoms on nonclinical populations was known and before modern depression scales were validated.

The CPI has been used widely in longitudinal research. With the availability of the CPI-D, decades of longitudinal data about depressive symptoms in adulthood become available immediately. Researchers can now address questions of stability and change, as well as begin to uncover the antecedents and sequelae of depressive symptomatology. Also, advances in statistical methods (see Collins & Sayer, 2001; McArdle & Nesselroade, 2002) will allow researchers to examine aspects of depressive symptomatology (e.g., growth and dynamic coupling) that so far could only be inferred from single time-point, short-term, or cross-sectional studies. However, to be immediately beneficial, pioneering statistical methods need to be accompanied by innovative assessment techniques. Not surprisingly, we found the construction of the CPI-D to be "a creative and fluid process, requiring as much inventiveness and resourcefulness as precision" (Kendall & Flannery-Schroeder, 1995, p. 892). We believe the CPI-D adds to the assessment potential of the CPI and offers a multitude of new research possibilities for both existing and future studies of depressive symptoms.

References

Abramson, L. Y., Metalsky, G. L., & Alloy, L. B. (1989). Hopelessness and depression: A theory-based subtype of depression. *Psychological Review, 96,* 358–372.

Adams, S. H., & John, O. P. (1997). A hostility scale for the California Psychological Inventory: MMPI, observer Q-sort, and Big Five correlates. *Journal of Personality Assessment, 69,* 408–424.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Baker, L. L., & Jessup, B. A. (1980). The psychophysiology of affective verbal and visual processing in dysphoria. *Cognitive Therapy and Research, 4,* 135–148.

Beck, A. T. (1967). *Depression: Causes and treatment.* Philadelphia: University of Pennsylvania Press.

Beck, A. T. (1978). *Beck Depression Inventory.* Unpublished manuscript. (Available from Center of Cognitive Therapy, Room 602, 133 South 36th Street, Philadelphia, PA 19104)

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory manual* (2nd ed.). San Antonio, TX: Psychological Corporation.

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77–110.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

Blatt, S. J. (1974). Levels of object representation in anaclitic and introjective depression. *Psychoanalytic Study of the Child, 29,* 107–157.

Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research.* Springfield, IL: Charles C Thomas.

Block, J. (1971). *Lives through time.* Berkeley, CA: Bancroft.

Block, J. (1989). *Prototypes for the California Adult Q-Set.* Unpublished manuscript. University of California, Berkeley, Department of Psychology.

Broadhead, W. E., Blazer, D. G., George, L. K., & Tse, C. K. (1990). Depression, disability days, and days lost from work in a prospective epidemiological survey. *Journal of the American Medical Association, 264,* 2524–2528.

Cartwright, L. K., & Wink, P. (1994). Personality change in women physicians from medical student years to mid-40s. *Psychology of Women Quarterly, 18,* 291–305.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100,* 316–336.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Collins, L., & Sayer, A. (Eds.). (2001). *New methods for the analysis of change.* Washington, DC: American Psychological Association.

Compas, B. E., Ey, S., & Grant, K. E. (1993). Taxonomy, assessment, and diagnosis of depression during adolescence. *Psychological Bulletin, 114,* 323–344.

Costello, C. G. (1992). Research on symptoms versus research on syndromes: Arguments in favor of allocating more research time to the study of symptoms. *British Journal of Psychiatry, 160,* 304–308.

Coyne, J. C. (1994). Self-reported distress: Analog or ersatz depression? *Psychological Bulletin, 116,* 29–45.

Dahlstrom, W. G., & Welsh, G. S. (1960). *An MMPI handbook.* Minneapolis: University of Minnesota Press.

Devins, G. M., & Orme, C. M. (1985). Center for Epidemiological Studies Depression Scale. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (Vol. 2, pp. 144–159). Kansas City, MO: Test Corporation of America.

Dozois, D. J., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment, 10,* 83–89.

Gold, J. R. (1990). Levels of depression. In B. B. Wolman & C. Stricker (Eds.), *Depressive disorders: Facts, theories, and treatment methods* (pp. 203–228). New York: Wiley.

Gotlib, I. H., Lewinsohn, P. M., & Seeley, J. R. (1995). Symptoms versus a diagnosis of depression: Differences in psychosocial functioning. *Journal of Consulting and Clinical Psychology, 63,* 90–100.

Gough, H. G. (1957). *Manual for the California Psychological Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G. (1987). *The California Psychological Inventory administrator's guide.* Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G. (2000). The California Psychological Inventory. In C. E. Watkins & V. L. Campbell (Eds.), *Testing and assessment in counseling practice* (2nd ed., 45–71). Mahwah, NJ: Erlbaum.

Gough, H. G., & Bradley, P. (1996). *The California Psychological Inventory manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

Groth-Marnat, G. (2003). *Handbook of psychological assessment* (4th ed.). Hoboken, NJ: Wiley.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 12,* 56–62.

Hammen, C. L. (2000). Interpersonal factors in an emerging developmental model of depression. In S. L. Johnson, A. M. Hayes, T. M. Field, N. Schneiderman, & P. M. McCabe (Eds.), *Stress, coping, and depression* (pp. 71–88). Mahwah, NJ: Erlbaum.

Helson, R., Jones, C., & Kwan, V. S. Y. (2002). Personality change over 40 years of adulthood: Hierarchical linear modeling analyses of two longitudinal samples. *Journal of Personality and Social Psychology, 83,* 752–766.

Helson, R., & Kwan, V. S. Y. (2000). Personality development in adulthood: The broad picture and processes in one longitudinal sample. In S. E. Hampson (Ed.), *Advances in personality psychology* (Vol. 1, pp. 77–106). Philadelphia: Psychologists Press.

Helson, R., Kwan, V. S. Y., John, O. P., & Jones, C. (2002). The growing evidence for personality change in adulthood: Findings from research with personality inventories. *Journal of Research in Personality, 36,* 287–306.

Helson, R., Pals, J., & Solomon, M. (1997). Is there adult development distinctive to women? In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 291–314). San Diego, CA: Academic Press.

Helson, R., Stewart, A. J., & Ostrove, J. (1995). Identity in three cohorts of midlife women. *Journal of Personality and Social Psychology, 69,* 544–557.

Hill, C. D., Neumann, C. S., & Rogers, R. (2004). Confirmatory factor analysis of the Psychopathy Checklist: Screening version in offenders with Axis I disorders. *Psychological Assessment, 16,* 90–95.

Holliman, N. B., & Guthrie, P. C. (1989). A comparison of the Millon Clinical Multiaxial Inventory and the California Psychological Inventory in assessment of a nonclinical population. *Journal of Clinical Psychology, 45,* 373–382.

Holliman, N. B., & Montross, J. (1984). The effects of depression upon responses to the California Psychological Inventory. *Journal of Clinical Psychology, 40,* 1373–1378.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structural analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Johnson, J., Weissman, M. M., & Klerman, G. L. (1992). Service utilization and social morbidity with depressive symptoms in the community. *Journal of the American Medical Association, 267,* 1478–1483.

Judd, L. L., Rapaport, M. H., Paulus, M. P., & Brown, J. L. (1994). Subsyndromal symptomatic depression: A new mood disorder? *Journal of Clinical Psychiatry, 55*(Suppl. 4), 18–28.

Kendall, P. C., & Flannery-Schroeder, E. C. (1995). Rigor, but not rigor mortis, in depression research. *Journal of Personality and Social Psychology, 68,* 892–894.

Kraemer, H. C., Yesavage, J. A., Taylor, J. L., & Kupfer, D. (2000). How can we learn about developmental processes from cross-sectional studies, or can we? *American Journal of Psychiatry, 157,* 162–171.

Marshall, G. N., Sherbourne, C. D., Meredith, L. S., Camp, P., & Hays, R. D. (2003). The tripartite model of anxiety and depression: Symptom structure in depressive and hypertensive patient groups. *Journal of Personality Assessment, 80,* 139–153.

McArdle, J. J., & Nesselroade, J. R. (2002). Growth curve analysis in contemporary psychological research. In J. Schinka & W. Velicer (Eds.), *Contemporary handbook of psychology, Volume 2: Research methods in psychology* (pp. 447–480). New York: Wiley.

McDowell, I., & Newell, C. (1996). Depression. In I. McDowell & C. Newell (Eds.), *Measuring health: A guide to rating scales and questionnaires* (pp. 238–286). New York: Oxford University Press.

Meites, K., Lovallo, W. R., & Pishkin, V. (1980). A comparison of four scales for anxiety, depression, and neuroticism. *Journal of Clinical Psychology, 36,* 427–432.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist, 50,* 741–749.

Moran, P. W., & Lambert, M. J. (1983). A review of current assessment tools for monitoring changes in depression. In M. J. Lambert, E. R. Christensen, & S. S. DeJulio (Eds.), *The assessment of psychotherapy outcome* (pp. 263–303). New York: Wiley.

Muthén, L. K., & Muthén, B. O. (2001). *Mplus user's guide* (2nd ed.). Los Angeles: Author.

Nezu, A. M., Ronan, G. R., Meadows, E. A., & McClure, K. S. (2000). *Practitioner's guide to empirically based measures of depression.* New York: Kluwer Academic/Plenum.

Orme, J. G., Reis, J., & Herz, E. J. (1986). Factorial and discriminant validity of the Center for Epidemiological Studies Depression (CES–D) Scale. *Journal of Clinical Psychology, 42,* 28–33.

Rabkin, J. G., & Klein, D. F. (1987). The clinical measurement of depressive disorders. In A. J. Marsella, R. M. Hirschfield, & M. M. Katz (Eds.), *The measurement of depression* (pp. 30–83). New York: Guilford Press.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385–401.

Raskin, A., Schulterbrandt, J., Rearig, N., & McKeon, J. (1969). Replication of factors of psychopathology in interview, ward behavior, and self-report ratings of hospitalized depressives. *Journal of Nervous and Mental Disease, 148,* 87–96.

Santor, D. A., Zuroff, D. C., Ramsay, J. O., Cervantes, P., & Palacios, J. (1995). Examining scale discriminability in the BDI and CES–D as a function of depressive severity. *Psychological Assessment, 7,* 131–139.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State–Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Stewart, A. J., & Vandewater, E. A. (1999). If I had it to do all over again: Midlife review, midcourse corrections, and women's well-being in midlife. *Journal of Personality and Social Psychology, 76,* 270–283.

Tanaka-Matsumi, J., & Kameoka, V. A. (1986). Reliabilities and concurrent validities of popular self-report measures of depression, anxiety, and social desirability. *Journal of Consulting and Clinical Psychology, 54,* 328–333.

Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681–706). Hillsdale, NJ: Erlbaum.

Tennen, H., Hall, J. A., & Affleck, G. (1995). Depression research methodologies in the *Journal of Personality and Social Psychology*: A review and critique. *Journal of Personality and Social Psychology, 68,* 870–884.

Twisk, J. W. R., Snel, J., Kempor, H. C. G., & Van Mechelen, W. (1998). Relation between the longitudinal development of personality characteristics and biological and lifestyle risk factors for coronary heart disease. *Psychosomatic Medicine, 60,* 372–377.

Watson, D., Clark, L. A., Weber, K., Assenheimer, J. S., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: II. Exploring the symptom structure of anxiety and depression in student, adult, and patient samples. *Journal of Abnormal Psychology, 104,* 15–25.

Weissman, M. M., & Boyd, J. H. (1981). Epidemiology of affective disorders: A reexamination and future directions. *Archives of General Psychiatry, 38,* 1039–1046.

Weissman, M. M., Prusoff, B., & Newberry, P. B. (1975). *Comparison of the CES–D, Zung, Beck self-report depression scales* (Tech. Rep. No. ADM 42–47–83). Rockville, MD: Center for Epidemiological Studies, National Institute of Mental Health.

Weissman, M. M., Sholomskas, D., Pottenger, M., Prusoff, B. A., & Locke, B. Z. (1977). Assessing depressive symptoms in five psychiatric populations: A validation study. *American Journal of Epidemiology, 106,* 203–214.

Wells, K. B., Burnam, M. A., Rogers, W., Hays, R. D., & Camp, P. (1992). The course of depression in adult outpatients: Results from the Medical Outcomes Study. *Archives of General Psychiatry, 49,* 788–794.

Wink, P., & Gough, H. G. (1990). New narcissism scale for the California Psychological Inventory and MMPI. *Journal of Personality Assessment, 54,* 446–462.