

The CES-D Scale: A Self-Report Depression Scale for Research in the General Population

Lenore Sawyer Radloff

Center for Epidemiologic Studies
National Institute of Mental Health

The *CES-D* scale is a short self-report scale designed to measure depressive symptomatology in the general population. The items of the scale are symptoms associated with depression which have been used in previously validated longer scales. The new scale was tested in household interview surveys and in psychiatric settings. It was found to have very high internal consistency and adequate test-retest repeatability. Validity was established by patterns of correlations with other self-report measures, by correlations with clinical ratings of depression, and by relationships with other variables which support its construct validity. Reliability, validity, and factor structure were similar across a wide variety of demographic characteristics in the general population samples tested. The scale should be a useful tool for epidemiologic studies of depression.

The Center for Epidemiologic Studies Depression Scale (*CES-D* Scale) was developed for use in studies of the epidemiology of depressive symptomatology in the general population. Its purpose differs from previous depression scales which have been used chiefly for diagnosis at clinical intake and/or evaluation of severity of illness over the course of treatment. The *CES-D*

was designed to measure current level of depressive symptomatology, with emphasis on the affective component, depressed mood. The symptoms are among those on which a diagnosis of clinical depression is based but which may also accompany other diagnoses (including "normal") to some degree.

This definition of the variable being measured determines the appropriate criteria of validity and reliability (*Standards for Educational and Psychological Tests*, 1974). Content validity will be based on the clinical relevance of the symptoms which comprise the items of the scale. Criterion-oriented validity will include correlations with other valid self-report depression scales, correlations with clinical ratings of severity of depression, and discrimination between psychiatric patients and general population samples. Construct validity will be based on what is known about the theory and epidemiology of depressive symptoms. Evidence that the scale is reliable but is also sensitive to current levels of symptomatology will be based on predictability of test-retest changes in scores (e.g., scores of patients before and after treatment, or scores of household respondents before and after "Life Events Losses"). Since several comparable samples (essentially replications) were tested, consistency of results across the

samples will also be shown as indirect evidence of reliability.

The *CES-D* was designed for use in general population surveys, and is therefore a short, structured self-report measure. It is usable by lay interviewers, acceptable to the respondent, and not substantially influenced by the normal range of conditions during a household interview. The scale was designed for use in studies of the relationships between depression and other variables across population subgroups. To compare results from one subgroup to another, the scale must be shown to measure the same thing in both groups. Therefore, it will be shown that properties of the scale (validity, reliability, factor structure) are similar for the various population subgroups to be studied.

Development of the Scale

The *CES-D* items were selected from a pool of items from previously validated depression scales (e.g. Beck, Ward, Mendelson, Mock, & Erbaugh, 1961; Dahlstrom & Welsh, 1960; Gardner, 1968; Raskin, Schulerbrandt, Reatig, & McKeon, 1969; Zung, 1965). The major components of depressive symptomatology were identified from the clinical literature and factor analytic studies. These components included: depressed mood, feelings of guilt and worthlessness, feelings of helplessness and hopelessness, psychomotor retardation, loss of appetite, and sleep disturbance. Only a few items were selected to represent each component. Four items were worded in the positive direction to break tendencies toward response set as well as to assess positive affect (or its absence). To emphasize current state, the directions read: "*How often this past week did you . . .*" Each response was scored from zero to three on a scale of frequency of occurrence of the symptom.

Pretests on small "samples of convenience" indicated appropriate performance of the scale and guided minor revisions for clarity and acceptability. The 20-item scale used in the studies reported here is shown in Table 1. The

possible range of scores is zero to 60, with the higher scores indicating more symptoms, weighted by frequency of occurrence during the past week.

Field Tests: Methods

First Questionnaire Survey (Q1 Survey)

The *CES-D* scale was included in a structured interview containing over 300 items, including other scales designed to measure depression or depressed mood (Bradburn Negative Affect, 1969; Lubin, 1967), psychological symptoms (Langner, 1962), well-being (Bradburn Positive Affect, 1969; Cantril Ladder, 1965) and Social Desirability (Crowne & Marlowe, 1960). It also included standard sociodemographic items (age, sex, education, occupation, marital status) and measures of life events, alcohol problems, social functioning, physical illness and use of medications. The interview, which took about an hour, was conducted by an experienced lay interviewer in the home of the respondent.

Probability samples of households designed to be representative of two communities (Kansas City, Missouri, and Washington County, Maryland) were selected. An individual (aged 18 and over) was randomly selected for interview from each household in the sample. Independent samples of households were designated for each week of the study. Strong efforts were made to complete interviews in the assigned week, but up to three weeks (and unlimited numbers of callbacks) were allowed to maximize response rate. Interviewing was done from October 1971 through January 1973 in Kansas City and from December 1971 through July 1973 in Washington County. The response rate in Kansas City was about 75%, with a total of 1173 completed interviews; in Washington County the response rate was about 80%, with 1673 completed interviews. Informed consent was obtained from all respondents. Both sites had a refusal rate of about 17%, plus a small percentage of not-at-home and other reasons for nonresponse.

Demographic distributions of the samples are

Table 1. CES-D Scale

INSTRUCTIONS FOR QUESTIONS: Below is a list of the ways you might have felt or behaved. Please tell me how often you have felt this way during the past week. HAND CARD A.

Rarely or None of the Time (Less than 1 Day)

Some or a Little of the Time (1-2 Days)

Occasionally or a Moderate Amount of Time (3-4 Days)

Most or All of the Time (5-7 Days)

During the past week:

1. I was bothered by things that usually don't bother me.
 2. I did not feel like eating; my appetite was poor.
 3. I felt that I could not shake off the blues even with help from my family or friends.
 4. I felt that I was just as good as other people.
 5. I had trouble keeping my mind on what I was doing.
 6. I felt depressed.
 7. I felt that everything I did was an effort.
 8. I felt hopeful about the future.
 9. I thought my life had been a failure.
 10. I felt fearful.
 11. My sleep was restless.
 12. I was happy.
 13. I talked less than usual.
 14. I felt lonely.
 15. People were unfriendly.
 16. I enjoyed life.
 17. I had crying spells.
 18. I felt sad.
 19. I felt that people dislike me.
 20. I could not get "going."
-

reported elsewhere (Comstock & Helsing, in press), as are analyses of characteristics of those who refused to be interviewed (Comstock & Helsing, 1973; Klassen & Roth, 1974). Refusals were significantly more likely to have lower education and come from smaller households than respondents. Analyses have been made of respondents interviewed in the assigned week ("on time") versus the harder to find respondents interviewed in the following three weeks ("late") (Mebane, 1973). Males and working people were slightly overrepresented among the "late" respondents, but the "late" did not differ from the "on time" on the psychological measures in the

interview, including the *CES-D* scale. The samples probably have some underrepresentation of males and the poorly educated. However, they include respondents with a wide range of demographic characteristics, in numbers adequate for analyses of relationships among variables.

Second Questionnaire Survey (Q2 Survey)

The *CES-D* scale was also included in a slightly revised (mainly shortened) version of the questionnaire (Q2) used in Washington County only, from March 1973 through July 1974 (for three months Q1 and Q2 were used alternately). Sam-

ples were drawn for four-week periods. The revision was not expected to affect the *CES-D*, since the scale was placed very early in both interviews, with identical preceding sections. The major differences between the Q1 and Q2 surveys were: length of interview (60 vs. 30 minutes); the time-basis of the sampling frame (weekly vs. four-week); and the site (Kansas City and Washington County vs. Washington County *only*). The response rate for the Q2 survey was about 75%, with 1089 completed interviews, and about 22% refusals. Therefore, the obtained sample for Q2 may be slightly less representative than that of the Washington County Q1 survey.

Mail-backs

From May 1973 through March 1974, each respondent to Q2 was asked to fill out and mail back one retest on the *CES-D* scale either two, four, six, or eight weeks after the original interview. A total of 419 mail-backs was received (about 56% response rate).

Reinterview Survey (Q3 Survey)

The *CES-D* was also included in a reinterview (Q3) of samples of the original respondents to Q1 or Q2. In Kansas City, from July 1973 through December 1973, 343 respondents (78% of those attempted) were reinterviewed about 12 months after the original interview. From August 1973 through April 1974 in Washington County, 1209 respondents (about 79% of those attempted) were reinterviewed once—either three, six, or twelve months after the original interview.

Psychiatric Patient Samples

Two clinical validation studies have been done in coordination with the survey program: one in Washington County, Maryland (Craig & Van Natta, in press) and one in New Haven, Connec-

ticut (Weissman, Prusoff & Newberry, 1975). In the Washington County study, seventy patients residing in a private psychiatric facility were selected on the basis of willingness and ability to participate. Each patient was rated on the Rockliff Depression Rating Scale (Rockliff, 1971) by the nurse-clinician who was most familiar with the patient's current status. Immediately following this, the patient was interviewed by one of the interviewers from the Washington County general population survey, using the original interview form (Q1). In the New Haven Study, thirty-five people admitted to outpatient treatment for severe depression and scoring seven or higher on the Raskin Depression Rating Scale (Raskin et al., 1969) participated in the study. They were given the *CES-D* scale and the SCL-90 (Derogatis, Lipman, & Covi, 1973) as self-reports and rated by clinicians on the Hamilton Rating Scale (Hamilton, 1960) as well as the Raskin. The measures were taken upon admission for treatment, after one week, and after four weeks of treatment, using psychotropic medication and supportive psychotherapy.

Summary of Field Tests

In the present paper, results will be reported for the following populations: All Q1 Whites, All Q2 Whites, and All Q3 Whites (these analyses were confined to Whites to make the samples comparable, since the Q2 sample contained less than 3% nonwhites). These results are treated as replications to demonstrate repeatability of the properties of the scale across two samples considered equivalent (Q1 vs. Q2) and across two tests on essentially the same sample (Q1/Q2 vs. Q3). For test-retest reliability, scores of the same people at different times will be compared (Q2 vs. mail-backs and Q1/Q2 vs. Q3). To demonstrate generalizability across different groups and thereby justify the epidemiologic uses of the scale, results will be compared across age, sex, race, and educational subgroups of the combined Q1 and Q2 general populations and with the Washington County patient group.

Table 2. Parameters of the Distributions
of the CES-D scale

Sample	N	\bar{X}	SD	SKEW	% at & above score of 16	t-test: Q1 sample vs Q2, Q3 or patient sample
All Q1 Whites	2514	9.25	8.58	1.50	19	
All Q2 Whites	1060	8.17	8.23	1.69	15	3.48 ^a
All Q3 Whites	1422	7.94	7.53	1.51	15	5.04 ^a
WC Patients	70	24.42	13.51	.20	70	9.36 ^a

^asignificant, $p < .01$

Table 3. Internal Consistency of CES-D Scale

Sample	Inter-item Correlations		Item-Scale Correlations		Estimates of Reliability		
	Range of r's		Range of r's		Coefficient alpha	Split-Halves	Spearman-Brown
	Percent ^a Distribution < .10	.10-.30 > .30	Percent ^b Distribution < .40	.40-.60 > .60			
All Q1 Whites	12%	64%	24%	15%	.85	.77	.87
All Q2 Whites	10	65	25	15	.85	.76	.86
All Q3 Whites	13	65	22	20	.84	.77	.87
W.C. Patients	11	34	55	5	.90	.85	.92

^aout of 190 correlations

^bout of 20 correlations

Suitability for Use in Household Surveys

Acceptability

The *CES-D* scale proved acceptable to both general and clinical populations. In the Q1 and Q2 survey data, the average nonresponse to single items was less than 0.2%. Two items were answered "don't know" or "not applicable" somewhat more often than average: "I felt I was just as good as other people" (1.6%) and "I felt hopeful about the future" (2.2%). For comparison, the Q1 item with the highest nonresponse was household income (8%). The entire *CES-D* scale was considered missing if more than four items were missing. This occurred only twice in the Q1 and Q2 data combined.

Distributions

Parameters of the distributions of *CES-D* scale scores in the general population samples and the Washington County patients are shown in Table 2. The distribution of scores in the patient group was symmetrical, with a large standard deviation, while the general population distributions were very skewed, with a smaller standard deviation and a much larger proportion of low scores. This pattern is consistent with an interpretation of the scale as related to a pathological condition more typical of a patient population than a household sample. However, there was a wide enough range of scores in the general population to allow meaningful identification of relationships between depressive symptomatology and other variables.

The very skewed distributions and the fact that groups with higher means also tended to have higher variances should be noted. Standard parametric significance tests on these data will not be exact. However, several basic analyses (e.g., analysis of variance of *CES-D* scores by sex and marital status) have been replicated using normalizing transformations and nonparametric tests. In no case was the decision (accept vs. reject H_0) reversed. Nevertheless, probability

levels reported should be considered approximate, and borderline levels should be interpreted with caution.

Conditions of Interview

Controlling for age, race, sex, marital status, income and occupational role, analyses of conditions during the interview revealed no significant differences in the *CES-D* scores associated with the time of day, day of week, or month of year in which the interview took place (tables available on request). There were some differences in the average *CES-D* scores obtained by the various interviewers, significant at borderline levels, which warrant further analyses (Choi & Comstock, 1975; Handlin et al., 1974). The Washington County Q2 survey obtained significantly lower *CES-D* scores than did the Washington County Q1 survey. The Q3 reinterview survey (Kansas City and Washington County combined) also had significantly lower average *CES-D* scores than the initial (Q1 or Q2) survey of the same respondents. Further analyses will be required to assess the relative contribution of several factors (including possible real time trends, response bias in test-retest effects, interview forms, possible nonresponse bias, and sampling procedures).

These possible differences (among interviewers and among the three surveys) are small in magnitude and of minor practical importance. If further analyses confirm true differences, then suitable controls could be introduced in the sampling or analytic procedures. It is more important for present purposes that the properties of the scale were consistent across interviewers and questionnaire forms. Analysis of the interviewers (excluding those who completed fewer than 20 interviews) revealed very similar levels of reliability and patterns of relationships to other variables across interviewers (tables available on request). Results from the Q1, Q2 and Q3 surveys are reported separately in subsequent sections of this paper to demonstrate that they are also very similar on these properties.

Reliability

Internal Consistency

The scale contains 20 symptoms, any of which may be experienced occasionally by healthy people; a seriously depressed person would be expected to experience many but not necessarily all of these symptoms. In a healthy population, positive and negative affect are expected to co-exist, with a low (negative) correlation. However, it has been suggested (Klein, 1974) that severely depressed patients are characterized by absence of positive as well as presence of negative affect, so that positive and negative affect would be more highly (negatively) correlated. There is also evidence that different kinds of people may manifest different types of symptoms; e.g., lower socioeconomic status people report more physical symptoms while higher socioeconomic status people report more affective symptoms (Crandell & Dohrenwend, 1967). In summary, in a general population sample, we would expect a great deal of heterogeneity, with many people experiencing a few symptoms and a few experiencing many. Therefore, some inter-item correlations may be quite low, but the direction of correlations should be consistent enough to produce reasonably high measures of internal consistency. In a patient group, we would expect higher item means, higher inter-item correlations, and very high internal consistency.

The results support these expectations (see Table 3). Both inter-item and item-scale correlations were higher in the patient sample than in the general population samples (even when the small N and, therefore, greater sampling error of the patient sample is taken into account). Expectations were also confirmed by measures of internal consistency (coefficient alpha and the Spearman-Brown, split-halves method; Nunnally, 1967). They were high in the general population (about .85) and even higher in the patient sample (about .90).

This high internal consistency may include some component of response bias, i.e., the tendency of an individual to answer all questions in the same (positive or negative) direction. In the

extreme, this would result in a bimodal distribution of scores, which was not observed in the present data. Evidence of discriminant validity and of validity based on clinician's ratings, independent of self-report, also suggests that response bias is not the major contributor to the reliability of the scale.

Test-retest Correlations

Predictions regarding test-retest correlations depend on several factors. The *CES-D* scale was explicitly designed to measure *current* ("this week") level of symptomatology, which is expected to vary over time. Changes over time may not be monotonic; they are more likely to be cyclic in at least some individuals, and the phase (length) of cycles may vary across individuals. The *CES-D* was designed to be sensitive to possible depressive reactions to events in a person's life; the timing of these events is unpredictable but presumably aperiodic. There are also methodological complications in test-retest measures. For example, there may be biases due to nonresponse, biasing effects of repeated testing, and asymmetric regression toward the mean due to the very skewed distribution of *CES-D* scores. Furthermore, in the present data, the test-retest time interval was confounded with differences in style of data collection: all initial scores were based on interviews; the short-interval (weeks) retests were different (i.e., self-administered mail-backs); the long-interval (months) retests were the same (i.e., interviews).

In light of these properties of the variable being measured, we would expect only moderate levels of test-retest correlations in the overall samples. Shorter test-retest time intervals should produce somewhat higher correlations than longer intervals. However, if people were selected by the information we have about what happened during the time interval, the correlations should be better differentiated. Specifically, life events are expected to introduce variability (i.e., some individuals may react more than others) and thus lower the test-retest correla-

tions. Tables 4 and 5 show that the results were consistent with these predictions.

Table 4 shows the test-retest correlations for those who responded to the request to fill out and mail in a retest of the *CES-D* (mail-backs) and those who were reinterviewed (Q3). All respondents were retested only once; each time interval represents a different group of people. The correlations were in the moderate range (all but one were between .45 and .70) and were, on the average, larger for the shorter time intervals.

In Table 5, all Q3 respondents (test-retest time interval ranging from three to twelve months) were classified by whether any one of 14 negative life events had occurred in the year prior to the first interview and in the interval between interviews. Those with no life events at either time had the highest test-retest correlation; those with life events at both times had the lowest correlation. Those with events at one time

but not the other had intermediate correlations. The correlation for those with no events ($r = .54$) might be considered the fairest estimate of test-retest reliability, in the sense of repeatability with conditions replicated, for the three- to twelve-month time interval. In the New Haven patient group, the correlation of *CES-D* scores at admission with scores obtained after four weeks of treatment was .53 (compared with $r = .58$ for the SCL-90). In this group, "events" had certainly occurred, but the effect of treatment may be assumed to be in the same direction for all or most patients. Therefore, it is reasonable that the correlation was about the same as that in the "no events" group.

Validity

Although not designed for clinical diagnosis, the *CES-D* scale is based on symptoms of depression as seen in clinical cases. Therefore, it

Table 4. Test-retest Correlations by Time Interval Between Test and Retest

Mail-backs (WC Only)		
Time Interval in Weeks	n	Test-Retest r
2	139	.51
4	105	.67
6	97	.59
8	78	.59
TOTAL	419	.57

Q3 Reinterview		
Time Interval in Months	n	Test-Retest r
WC 3	378	.48
WC 6	349	.54
WC 12	472	.49
KC 12	342	.32

Table 5. Test-retest Correlations by Life Events Losses Before Each Test

Life Events Loss Reported at:		n	Test- Retest r
Test	Retest		
No	No	617	.54
Yes	No	362	.48
No	Yes	270	.47
Yes	Yes	302	.31

should discriminate strongly between patient and general population groups, be sensitive to levels of severity of depressive symptomatology, and reflect improvements after psychiatric treatment. In addition, it should correlate well with other scales designed to measure depression and less well with scales which measure related but different variables; be related to a felt need for psychiatric services; and be sensitive to possible reactive depression in the face of certain life events.

Clinical Criteria

The *CES-D* scores discriminated well between psychiatric inpatient and general population samples and discriminated moderately among levels of severity within patient groups. Table 2 shows that the average *CES-D* score for the group of 70 Washington County psychiatric inpatients was substantially and significantly higher than the average for the general population samples. Seventy percent of the patients but only 21% of the general population scored at and above an arbitrary cutoff score of 16. In the patient group, the correlation between the *CES-D* scale and ratings of severity of depression by the nurse-clinician was .56 (Craig & Van Natta, in press). In the New Haven patient group, the average *CES-D* score at admission was 39.11, with no score below 16 (note that this group was screened to include only those above 6 on the Raskin scale). The correlations of the *CES-D* with the Hamilton Clinician's Rating scale and with the Raskin Rating scale were moderate (.44 to .54) at admission. After four weeks of treat-

ment, the correlations were substantially higher (.69 to .75). These correlations were almost as high as those obtained for the 90-item SCL 90 (Weissman et al., 1975).

Self-report Criteria

Table 6 shows correlations of the *CES-D* scale with other self-report scales in the several samples. (Note that Q2 and Q3 did not include all scales.) In all the samples, the pattern of correlations of the *CES-D* with other scales gives reasonable evidence of discriminant validity. The highest *r*'s were with scales designed to measure symptoms of depression (*i.e.*, Lubin, Bradburn Negative Affect and Bradburn Balance) or general psychopathology (Langner) and the Cantril Ladder. The correlation of the *CES-D* with the Bradburn Positive Affect scale was negative and was low positive with scales designed to measure different variables (medications, disability days, social functioning, aggression). The *CES-D* correlated moderately with interviewer ratings of depression but low negative to zero with interviewer ratings of cooperation and understanding of the question.

Table 6 also shows support for the concept of a "syndrome" of depression which is more consistent in the patient sample than in the general population samples. In the patient groups, the correlations with other depression scales were higher positive (in the New Haven patients, correlation with the SCL-90 was .83); with the Bradburn Positive Affect, higher negative; and with other scales, the same low positive.

The low negative correlations with the Marlowe-Crowne scale of "social desirability" suggest that there may be some general response set involved in the *CES-D* scores (see also Klassen, Hornstra, & Anderson, 1975). However, the pattern of correlations in Table 6 suggests that this bias is small and does not entirely mask meaningful relationships with other variables.

Need for Services

In the Q1 and Q2 surveys, the respondents were asked whether they had had an emotional

Table 6. Correlations of CES-D With Other Self-report Scales

Other Scales	Correlations with CES-D			
	All Q1a Whites	All Q2a Whites	All Q3a Whites	Washington Co. ^b Patients
Lubin	.51	.43	.37	.70
Bradburn Negative Affect	.60	.63		.55
Bradburn Balance ^a	.61	.62		.72
Langner	.54			.60
Cantril Ladder	.43	.44	.38	.74
Bradburn Positive Affect	-.21	-.25		-.55
Medications	.24	.20		.22
Disability Days	.28	.32		.30
Social Functioning	.19	.13	.15	.24
Aggression	.26	.29		.21
Interviewer Rating				
Depression	.49	.53		.46
Cooperation	-.03	-.01		.11
Understanding	-.13	-.05		-.03
Marlowe-Crowne Total	-.18			-.20
Assert Good	-.07			-.12
Deny Bad	-.23			-.21

^a $n > 1000$, $r = .06$ significant at $p < .05$
 $r = .08$ significant at $p < .01$

^b $n = 70$, $r = .23$ significant at $p < .05$
 $r = .30$ significant at $p < .01$

Table 7. CES-D by "Need Help"

Sample	N	\bar{X}	SD	Skew	>16	r with Bradburn		\bar{X} CES-D Yes vs. No t test ^a
						Negative affect	Positive affect	
All Q1 Whites								
Need Help: Yes	300	18.96	11.50	.38	57	.49	-.40	16.25
No	2194	7.89	7.11	1.53	13	.54	-.15	
All Q2 Whites								
Need Help: Yes	94	18.91	11.11	.44	56	.52	-.37	10.10
No	962	7.11	7.07	1.79	11	.59	-.21	
All Q3 Whites								
Need Help: Yes	247	12.92	9.51	.82	36	-	-	9.6
No	1171	6.87	6.55	1.62	10	-	-	
WC Patients ^b	70	24.42	13.51	.20	70	.55	-.55	

^asignificant, $p < .01$, all 3 groups

^bAll but 2 patients said "yes" to "Need Help"

problem in the past week for which they felt they needed help. The group who answered the "need help" question "yes" or "no, because it's no use to look for help" are considered "at risk" of becoming patients. Table 7 shows parameters of the *CES-D* scale by answers to this question. The Washington County patient population is included for comparison. The general population "need help" groups were more similar to the patients than were the "not need help" groups. The "need help" groups had high means (significantly higher than the "not need help" groups) and standard deviations, symmetrical distributions (low skew), high percentages of high scores (16 and above), and moderately high correlations with the Bradburn Positive Affect scale. The patterns of correlations with the Bradburn Positive and Negative Affect scales can also be considered in terms of discriminant validity.

Life Events

Past research has shown an association of illness, including mental illness, with certain significant life events (Dohrenwend & Dohrenwend, 1974). Table 8 shows the average *CES-D* scores for those who do and do not report certain events in the year (or during the retest interval for Q3) preceding the interview. The results were as predicted: the more negative the event, the higher the depression score of those who experienced it. Vacations were associated with low *CES-D* scores (possibly biased by socioeconomic status); marriage was ambiguous; separation was more strongly associated with depression than was divorce.

Table 9 shows the interview-reinterview (i.e., Q1/Q2 vs. Q3) *CES-D* scores by Life Events Losses (using the same criterion of Life Events Losses as was used for Table 5). The overall trend was for lower scores on Q3 than on the original interview, *except* in the group with no events before the first interview and at least one event in the retest interval. The four groups were significantly different in amount of change in *CES-D* score by several different methods of

Table 8. *CES-D* By Recent Life Events

Sample	Married		Vacation		Illness/Injury		Divorced		Separated		Widowed	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
All Q1 Whites: N	62	2440	1132	1369	526	1976	33	2469	47	2455	24	2478
\bar{X} <i>CES-D</i>	7.48	9.28	7.95	10.31	11.20	8.71	11.27	9.21	16.83	9.09	19.13	9.14
Yes vs. No: $p <$.04		.01		.01		.26		.01		.01
All Q2 Whites: N	29	1033	450	612	201	860	13	1049	16	1046	13	1049
\bar{X} <i>CES-D</i>	11.14	8.09	7.16	8.91	10.46	7.62	7.31	8.18	18.50	8.01	14.54	8.09
Yes vs. No: $p <$.12		.01		.01		.68		.01		.03
All Q3 Whites: N	27	1404	606	825	222	1209	11	1420	12	1419	7	1424
\bar{X} <i>CES-D</i>	7.19	7.73	7.08	8.19	9.30	7.43	11.64	7.69	13.58	7.67	17.29	7.67
Yes vs. No: $p <$.64		.02		.02		.21		.11		.08

Table 9. Test-retest CES-D Average Scores
By Life Events Losses Before Each Test

Life Events Loss Reported at:	n	\bar{X}	SD	Average Change Score ^a (Retest-test)
Test: No	608	7.40	7.05	-.86
Retest: No		6.71	6.94	
Test: Yes	362	9.54	8.39	-2.14
Retest: No		7.55	6.85	
Test: No	270	8.25	7.88	+.45
Retest: Yes		8.92	7.61	
Test: Yes	302	11.66	10.01	-2.01
Retest: Yes		9.80	8.84	

^a Overall significance of difference between groups in change scores: $p < .01$ in one-way analysis of variance and in one-way analysis of covariance, with score at time 1 as covariate.

testing change scores. These relationships of the CES-D scores to life events are considered validation of its sensitivity to current mood state, which is a property desired for the scale.

Improvement After Treatment

Further evidence of response to change is furnished by the New Haven clinical study (Weissman, et al., 1975). The average CES-D score, along with the SCL 90, the Hamilton, and the Raskin, decreased significantly from the time of admission to one week and to four weeks of treatment (see Table 10). The mean for each of the 20 items was lower after four weeks of treatment than upon admission (tables available on request). The change was particularly large for patients rated "recovered" (by a Raskin score of less than 7) after four weeks. The average CES-D score went down 20 points in the recovered group and 12 points in the group rated "still ill."

Factor Analysis

Principal components factor analysis (with ones in the main diagonal) of the 20-item scale was done for the three general population groups (All Q1 Whites, All Q2 Whites, and All Q3 Whites). For each group, there were four eigenvalues greater than one, which together accounted for a total of 48% of the variance; therefore, the normal varimax rotation to four factors was examined (see Table 11). The pattern of factor loadings is quite consistent across the three groups. Including items with loadings above .40 in all three groups, the four factors are readily interpretable as follows:

- I. Depressed affect (*blues, depressed, lonely, cry sad*)
- II. Positive affect (*good, hopeful, happy, enjoy*)
- III. Somatic and retarded activity (*bothered, appetite, effort, sleep, get going*)
- IV. Interpersonal (*unfriendly, dislike*)

Table 10. Average Scores of New Haven Patients At Admission and After Treatment^a (N=35)

Time ^b	\bar{X} CES-D	\bar{X} SCL-90	\bar{X} Hamilton	\bar{X} Raskin
1 Admission	39.11	1.73	23.37	9.63
2 After 1 week treatment	29.29	1.09	14.89	7.29
3 After 4 week treatment	20.91	0.74	9.74	5.77

^aFrom Weissman et al (1975) Tables 3 & 4.

^bMatched t-test, $p < .001$ for all 4 measures, for 3 comparisons:
 Time 1 vs. Time 2
 Time 2 vs. Time 3
 Time 1 vs. Time 3

Including items with loadings of at least .35 in at least two groups would add the items *failure*, *fearful*, *happy*, and *enjoy* to the depressed affect factor and the items *blues*, *mind*, *depressed*, and *talk* to the somatic factor. In all three groups, the depressed affect factor shares the largest proportion of the variance (about 16%) and the interpersonal factor, the smallest proportion (about 8%).

Similarity of factor structure of the three samples was estimated by the Factorial Invariance Coefficient, r_{iv} (Derogatis, Kallmen, & Davis, 1971; Derogatis, Serio, & Cleary, 1972; Pinneau & Newhouse, 1964). The r_{iv} is a measure of the correlation of the loadings of all items on one factor in one group versus the loadings on one factor in another group. If the factor structure of two groups is similar, the r_{iv} will be very high when loadings on the *same* factor in both groups are correlated (the "diagonal" coefficients) and very low when *different* factors are correlated (the "off-diagonal" coefficients). Comparing Q1 with Q2 and Q1 with Q3, the diagonal coefficients were very high (.87 to .99). The off-diagonal coefficients (i.e., the similarity of *different* factors) were very low (the largest was .13).

This is very strong evidence that the *CES-D* has a similar factor structure in two samples from similar populations (Q1 vs. Q2) and across two tests on essentially the same sample (Q1 vs. Q3).

The factors found in the general population are consistent with the components of depression built into the scale. However, the high internal consistency of the scale found in all groups argues against undue emphasis on separate factors. The items are all symptoms related to depression. For epidemiologic research, a simple total score is recommended as an estimate of the degree of depressive symptomatology.

Generalizability Across Subgroups

To be useful for epidemiologic studies (e.g., distribution of depression across demographic subgroups), the *CES-D* scale must have adequate reliability and validity and a similar factor structure *within each subgroup of the population*. Therefore, the analyses of Tables 3, 4, 6, 7 and 11 were repeated on each of three age groups (under 25, 25-64, over 64), the two sexes, two races (Black and White), three levels of education (less than high school, high school, greater than high school), and the two "need

Table 11. Factor Loadings of CES Items for All Q1 Whites, All Q2 Whites and All Q3 Whites

Item	Factor I Depressed			Factor II Positive			Factor III Somatic			Factor IV Interpersonal			Item Communalities		
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
1. Bothered	-.35	.24	.23	.01	-.11	.09	-.46	.43	-.51	-.01	-.02	-.10	.34	.26	.33
2. Appetite	-.16	.20	.12	.07	-.09	.00	-.54	.45	-.50	-.00	-.05	.13	.32	.25	.28
3. Blues	-.58	.61	.60	.19	-.11	.15	-.36	.44	-.41	.12	-.09	-.13	.52	.59	.57
4. Good	-.17	.11	.11	-.66	.63	-.68	.10	.01	.01	-.13	.28	.11	.49	.48	.48
5. Mind	-.21	.28	.24	.05	-.31	.10	-.58	.37	-.59	.17	-.21	-.11	.41	.36	.43
6. Depressed	-.63	.66	.64	.23	-.24	.18	-.43	.40	-.43	.16	-.12	-.15	.66	.66	.65
7. Effort	-.09	.06	.15	.05	.14	.07	-.67	.65	-.64	.07	-.12	-.06	.46	.45	.44
8. Hopeful	.09	-.12	-.10	-.70	.70	-.68	-.01	-.04	.06	.03	.05	-.01	.49	.51	.47
9. Failure	-.38	.29	.44	.22	-.18	.28	-.12	.30	-.07	.45	-.29	-.11	.41	.29	.29
10. Fearful	-.41	.55	.31	.03	-.03	.19	-.23	.25	-.26	.33	-.23	-.13	.33	.42	.22
11. Sleep	-.30	.08	.21	.05	-.13	-.01	-.47	.63	-.55	-.05	.06	.07	.31	.42	.35
12. Happy	.32	-.46	-.38	-.66	.52	-.62	.22	-.25	.25	-.06	-.13	.05	.59	.56	.59
13. Talk	-.12	.23	.00	.09	.01	.10	-.44	.39	-.54	.21	-.30	-.20	.25	.30	.35
14. Lonely	-.65	.67	.72	.09	-.11	.06	-.21	.20	-.18	.20	-.07	-.09	.51	.51	.57
15. Unfriendly	-.00	.04	.15	.01	-.10	.07	-.12	.14	-.07	.78	-.74	-.84	.62	.57	.74
16. Enjoy	.31	-.42	-.35	-.70	.55	-.68	.08	-.27	.14	-.09	-.12	-.02	.60	.56	.60
17. Cry	-.70	.76	.65	.03	.02	.01	-.09	.01	-.15	.04	-.04	.04	.50	.57	.45
18. Sad	-.73	.75	.78	.14	-.18	.09	-.28	.21	-.20	.18	-.19	-.15	.67	.68	.68
19. Dislike	-.26	.20	.15	.09	-.05	.04	-.08	.03	-.08	.72	-.79	-.83	.61	.66	.73
20. Get Going	-.07	.18	.14	.10	-.14	.11	-.69	.66	-.66	.19	-.16	-.07	.52	.51	.48
SS Cols.	3.116	3.504	3.164	2.036	1.784	1.997	2.757	2.682	2.921	1.701	1.651	1.607	9.610	9.622	9.688

help" groups ("need help," "not need help"). For these analyses, the data from Kansas City and Washington County Q1 and Q2 were combined to maximize numbers in the subgroups.

With few exceptions, the results for the total population were confirmed in all subgroups. (tables available on request). In all subgroups, coefficient alpha was .80 or above. Test-retest correlations were moderate (.40 or above) in all but three groups (Blacks, age under 25, and "need help"). The subgroup patterns of correlations with other scales (as in Table 6) and relationships to "need help" (as in Table 7) were very similar to those in the total population. The subgroups did not differ from each other or from the total population in factor structure. The "need help" group (which had been found to be similar to the Washington County patient group by various criteria above) was not like that patient group in factor structure but was very similar to the total general population.

Cautions and Conclusions

Some limitations in use of the scale should be noted. It is not intended as a clinical diagnostic tool, and interpretations of individual scores should not be made. Even group averages should be interpreted in terms of level of symptoms which accompany depression, not in terms of rates of illness. Appropriate cutoff scores for clinical screening are yet to be validated. There are some hints that understanding of the items may be a problem; there was a very small but consistent correlation between the *CES-D* score and the interviewer ratings of understanding of the questions, independent of education of the respondent. Analyses indicate that this was not simply due to respondents who did not notice the reversal of the positive items. Special caution is needed with bilingual respondents (Trieman, 1975). Further study of this issue is needed, with possible revision for simplicity of wording and removal of colloquial expressions. There is still some question as to the effect of the interviewer and the interview form on the mean level of scale scores. Until further study decides this issue, a

sampling design balanced by interviewer may be appropriate.

On the positive side, the results reported here are very favorable for the uses of the *CES-D* scale for which it was designed. The scale has high internal consistency, acceptable test-retest stability, excellent concurrent validity by clinical and self-report criteria, and substantial evidence of construct validity. These properties hold across the general population subgroups studied. The scale is suitable for use in Black and White English-speaking American populations of both sexes with a wide range of age and socioeconomic status for the epidemiologic study of the symptoms of depression. A group with a high average score may be interpreted to be "at risk" of depression or in need of treatment. The scale is a valuable tool to identify such high-risk groups and to study the relationships between depressive symptoms and many other variables.

References

- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. An inventory for measuring depression. *Archives of General Psychiatry*, 1961, 4, 561-571.
- Bradburn, N. M. *The structure of psychological well being*. Chicago: Aldine, 1969.
- Cantril, H. *The pattern of human concern*. New Brunswick: Rutgers University, 1965.
- Choi, I. C., & Comstock, G. W. Interviewer effect on responses to a questionnaire relating to mood. *American Journal of Epidemiology*, 1975, 101, 84-92.
- Comstock, G. W., & Helsing, K. Characteristics of respondents and nonrespondents to a questionnaire for estimating community mood. *American Journal of Epidemiology*, 1973, 97, 233-239.
- Comstock, G. W., & Helsing, K. Symptoms of depression in two communities. *Psychological Medicine* (in press).
- Craig, T. J., & Van Natta, P. Recognition of depressed affect in hospitalized psychiatric patients: Staff and patient perceptions. *Diseases of the Nervous System* (in press).
- Crandell, D. L., & Dohrenwend, B. P. Some relations among psychiatric symptoms, organic illness and social class. *American Journal of Psychiatry*, 1967, 12, 1527-1537.

- Crowne, D. P., & Marlowe, D. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 1960, 24, 349-354.
- Derogatis, L. R., Kallmen, C. H., & Davis, D. M. FMATCH: A program to evaluate the degree of equivalence of factors derived from analyses of different samples. *Behavioral Science*, 1971, 16, 271-273.
- Derogatis, L. R., Lipman, R. S., & Covi, L. SCL-90: An outpatient psychiatric scale: Preliminary report. *Psychopharmacology Bulletin*, 1973, 9, 13-27.
- Derogatis, L. R., Serio, J. C., & Cleary, P. A. An empirical comparison of three indices of factorial similarity. *Psychological Reports*, 1972, 30, 791-804.
- Dohrenwend, B. S., & Dohrenwend, B. P. (Eds.) *Stressful life events: Their nature and effects*. New York: Wiley-Interscience, 1974.
- Gardner, E. A. *Development of a symptom check list for the measurement of depression in a population*. Unpublished, 1968.
- Hamilton, M. A rating scale for depression. *Journal of Neurologic Neurosurgical Psychiatry*, 1960, 23, 56-62.
- Handlin, V., Klassen, D., Hornstra, R., & Roth, A. *Interviewer effects in a community mental health survey*. Technical Report, 1974, The Greater Kansas City Mental Health Foundation, Contract PH 43-66-1324, National Institute of Mental Health.
- Klassen, D., Hornstra, R., & Anderson, P. The influence of social desirability on symptom and mood reporting in a community survey. *Journal of Consulting & Clinical Psychology*, 1975, 43, 448-452.
- Klassen, D., & Roth, A. *Characteristics of non-respondents in the Community Mental Health Assessment survey*. Technical Report, 1974, The Greater Kansas City Mental Health Foundation, Contract PH 43-66-1324, National Institute of Mental Health.
- Klein, D. F. Endogenomorphic depression. *Archives of General Psychiatry*, 1974, 31, 447-454.
- Langner, T. S. A twenty-two item screening score of psychiatric symptoms indicating impairment. *Journal of Health and Human Behavior*, 1962, 3, 269-276.
- Lubin, B. *Manual for the depression adjective check lists*. San Diego: Educational and Industrial Testing Service, 1967.
- Mebane, I. *On time and late respondents*. Technical Report, 1973. Center for Epidemiologic Studies, National Institute of Mental Health.
- Nunnally, J. C. *Psychometric theory*. New York: McGraw Hill, 1967.
- Pinneau, S. R., & Newhouse, A. Measures of invariance and comparability in factor analysis for fixed variables. *Psychometrika*, 1964, 29, 271-281.
- Raskin, A., Schulterbrandt, J., Reatig, N., & McKeon, J. Replication of factors of psychopathology in interview, ward behavior, and self-report ratings of hospitalized depressives. *Journal of Nervous and Mental Disease*, 1969, 148, 87-96.
- Rockliff, B. W. A brief rating scale for anti-depressant drug trials. *Comprehensive Psychiatry*, 1971, 12, 122-135.
- Standards for educational and psychological tests*. Washington, D.C.: American Psychological Association, 1974.
- Trieman, B. *Depressive mood among middle class urban ethnic groups*. Technical Report, 1975, Contract HSM 42-73-238, National Institute of Mental Health.
- Weissman, M. M., Prusoff, B., & Newberry P. *Comparison of the CES-D with standardized depression rating scales at three points in time*. Technical Report, 1975, Yale University, Contract ASH-74-166, National Institute of Mental Health.
- Zung, W. W. K. A self-rating depression scale. *Archives of General Psychiatry*, 1965, 12, 63-70.

Acknowledgements

The CES-D Scale was originally developed by Mr. Ben Z. Locke, Chief, Center for Epidemiologic Studies (CES), National Institute of Mental Health and Dr. Peter Putnam, formerly at CES. The overall program was initiated by Dr. Robert Markush, former Chief, CES. The field studies were carried out by the Epidemiologic Field Station, Greater Kansas City Mental Health Foundation, Kansas City, Missouri (Dr. Robijn Hornstra, Director) and The Training Center for Public Health Research, Johns Hopkins University, Hagerstown, Maryland (Dr. George Comstock) under contract with CES. The clinical validation studies were carried out by Dr. Thomas Craig, formerly of Johns Hopkins and Dr. Myrna Weissman, Yale University, under contract with CES. Important advice on and review of this report was provided by those connected with the study, especially Dr. Thomas Craig and Dr. Evelyn Goldberg, Johns Hopkins; by Dr. Len Derogatis, Johns Hopkins; and by the reviewers and editor of this journal.

Author's Address

Lenore S. Radloff, Room 10C-09, Parklawn Building, 5600 Fishers Lane, Rockville, Maryland 20852.